Removing the Field Size Loss from Duc et al.'s Conjectured Bound for Masked Encodings

Julien Béguinot[1], Wei Cheng[1,2], Sylvain Guilley[2,1], Yi Liu[1], Loïc Masure[3], Olivier Rioul[1], François-Xavier Standaert[3]

[1] LTCI, Télécom Paris, Institut Polytechnique de Paris, Palaiseau, France
[2] Secure-IC, Paris, France
[3] ICTEAM Institute, Université catholique de Louvain, Louvain-la-Neuve, Belgium

Abstract. At Eurocrypt 2015, Duc *et al.* conjectured that the success rate of a side-channel attack targeting an intermediate computation encoded in a linear secret-sharing, a.k.a. *masking* with $d+1$ shares, could be inferred by measuring the mutual information between the leakage and each share separately. This way, security bounds can be derived without having to mount the complete attack. So far, the best proven bounds for masked encodings were *nearly* tight with the conjecture, up to a constant factor overhead equal to the field size, which may still give loose security guarantees compared to actual attacks. In this paper, we improve upon the state-of-the-art bounds by removing the field size loss, in the cases of Boolean masking and arithmetic masking modulo a power of two. As an example, when masking in the AES field, our new bound outperforms the former ones by a factor 256. Moreover, we provide theoretical hints that similar results could hold for masking in other fields as well.

1 Introduction

If Side-Chanel Analysis (SCA) may be considered as a critical threat against the security of cryptography on embedded devices, it is no longer a fatality. Over the past decades, the *masking* counter-measure [4,13] has gained more and more success among designers and developers, both from an implementation and from a theoretical point of view. Masking can be seen as a linear secret sharing applied on each intermediate computation in the implementation of a cryptographic primitive that depends on some secret. In a nutshell, masking increases the attack complexity of any SCA adversary exponentially fast with the number of shares — provided that the leakages are sufficiently *noisy* and *independent* — while increasing the runtime and memory overhead at most quadratically [14]. This makes masking a theoretically *sound* counter-measure.

The Evaluation Challenge. Despite these achievements, the evaluation of a protected implementation remains cluttered by various technical and even conceptual difficulties. One way for evaluators to assess the security level of an implementation is to mount some known end-to-end attacks and to infer some

security level based on the outcomes of these attacks. Nevertheless, this relies on the assumption that the attacks mounted by the evaluators could depict well the optimal attacks that any adversary could realize. As an example, if for masking with two shares, end-to-end attacks using Deep Learning (DL) depict well optimal attacks [19,2], it is no longer true when masking uses more shares [3,21]. This could result in a false sense of security, and leaves the developers in an uncomfortable situation where implementations become increasingly hard to evaluate as their security level increases.

The Paradigm of Worst-Case Attacks. One way to circumvent this issue is to consider attacks in a so-called *worst-case* evaluation setting [1]. The core idea is to apply Kerckhoff's principles to side-channel security, by granting all the knowledge of the target to the adversary, *e.g.*, the random nonces used during the encryption, except the knowledge of the secret to guess. This way, the evaluator can efficiently profile the target implementation in order to (more) easily mount online attacks that approach the optimal ones. She can also analyze the leakage of the shares independently, in order to take advantage of masking security proofs to bound the security level under some assumptions.

Indeed, a series of theoretical works on masking allow to bound the amount of information leaked by a masked secret, depending on the amount of information leaked by each share separately, under the assumption that the shares' leakages are independent. Such bounds can be expressed, *e.g.*, in terms of the Mutual Information (MI), and then in turn be translated in terms of the Success Rate (SR) of any attack, as shown by Duc *et al.* at Eurocrypt 2015 [9]. Nevertheless, most of the current masking security proofs provide conservative bounds, possibly due to technical artifacts. In particular, they generally require more noise and more shares than expected by the best known attacks in order to reach a given security level [24,8,11,23].

Duc *et al.*'s Conjecture. Confronting this observation with empirical evidences, Duc *et al.* conjectured that the required number of queries to the target device needed to recover the target secret of a SCA is inversely proportional to the product of the MIs of each share [9]:

$$N_a(\mathsf{SR}) \approx \frac{f(\mathsf{SR})}{\prod_{i=0}^{d} \mathsf{MI}(\mathrm{Y}_i; \mathrm{L}_i)} \ ,$$

where d stands for the masking *order*,[1] and f is a "small constant depending on the target SR" [10, p. 1279]. Later at Ches 2019, Chérisey *et al.* bounded this constant based on the entropy of the target secret [7].

Due to its practical relevance, this conjecture recently gained attraction with two independent and simultaneous works by Ito *et al.* [16] at CCS 2022 and

[1] *i.e.*, the number of shares is $d + 1$ if the independence assumption is met.

by Masure *et al.* [22] at CARDIS 2022. Using different approaches, both works prove a *nearly*-tight version of Duc *et al.*'s conjecture for masked encodings, up to a constant factor equal to the field size M of the encoding.

This represents a significant improvement with respect to the previous proved bounds — *e.g.*, $\mathcal{O}\big(M^d\big)$ in Duc *et al.*'s proof [9], which additionally suffers from a reduced noise amplification rate. But it remains loose compared to empirical attacks performed against implementations of concrete ciphers like the Advanced Encryption Standard (AES). At a high level, Ito *et al.* and Masure *et al.*'s approaches used some back and forth between the MI and other metrics, such as the Total Variation (TV) [22] or the Euclidean Norm (EN) [16], in order to state noise amplification lemmata.[2] If these conversions taken separately are tight, their combination introduces an $\mathcal{O}(M)$ overhead, leading to the question whether tighter bounds could be proved, on which we focus.

Our Contribution. In this paper, we positively address the latter question, by removing this field size loss for masked encodings. At a high level, we do that by working directly with noisy leakages, without relying on reductions to more abstract (e.g., random probing) leakage models. Technically, our approach consists in stating the amplification lemma *directly* in terms of the MI, without any lossy conversion to other statistical distances. This idea is implemented using a result from Information Theory called Mrs. Gerber's Lemma (MGL) [5,17]. The MGL allows us to bound the MI between the secret and the whole leakage by a function of the MIs between each share and their corresponding leakage. Moreover, the bound given by the MGL is proved to be *tight*, in the sense that there exists some leakage distributions for which the inequality from the MGL is actually an equality. The only limitation compared to the previous works is that our bound only works for fields whose size is a power of 2. Thankfully, this limitation is not prohibitive, since our result covers, *e.g.*, Boolean masking or arithmetic masking modulo 2^n. Nevertheless, we argue at the end of this paper that similar results could also be obtained in different fields, whose size is not necessarily a power of 2. More generally, and since our results are for now specialized to masked encodings, it remains a natural question whether they generalize to computation, as also conjectured by Duc. et al. [9].

2 Statement of the Problem

We start the paper by stating the problem under consideration, before providing the solution in section 3, and discussing some perspectives in section 4.

2.1 Notations and Background

Side-Channel Attack. Let $(\mathcal{Y}, \oplus)$ be a group of finite order, denoted by M. Let $\mathrm{K} \in \mathcal{Y}$ be the secret key chunk to guess. To this end, we consider that

[2] *e.g.*, Young-Minkowski's convolution inequality for the TV [22] or Plancherel's formula combined with the convolution theorem for the EN [16].

the adversary knows a sequence of N_a plaintexts $\{\mathrm{P}\}_{N_a}$, and can observe the sequence of leakages $\{\mathbf{L}\}_{N_a}$ associated to the corresponding intermediate computations $\{\mathrm{Y} = \mathbf{C}(\mathrm{K}, \mathrm{P})\}_{N_a}$. Based on this side-channel information, the adversary returns a key guess $\widehat{\mathrm{K}}$. We define the Success Rate (SR) as $\mathsf{SR} = \Pr\left(\mathrm{K} = \widehat{\mathrm{K}}\right)$. Since the SR increases when the number of observed traces N_a increases as well, we next define the quantity $N_a(\mathsf{SR}, \mathcal{Y})$ as the minimal number of leakage traces required for any adversary to reach a success rate at least SR.

Masking. In order to protect cryptographic secrets against side-channel leakage, we consider the intermediate computation Y — assumed to be uniformly distributed — to be masked.[3] Let $\mathrm{Y}_0, \ldots, \mathrm{Y}_d$ be $d+1$ random variables out of which d are uniformly drawn from $\mathcal{Y}$, that we call the *shares*, and denote by $\mathrm{Y} = \mathrm{Y}_0 \oplus \ldots \oplus \mathrm{Y}_d$ the random variable to protect, that we call the *secret*. Concretely, for each trace $\mathbf{L} = (\mathrm{L}_0, \ldots, \mathrm{L}_d)$, the adversary observes a *leakage* L_i, whose distribution conditionally to Y_i is independent of all the other random variables. In our setting, we assume that an evaluator has been able to characterize the amount of uncertainty about Y_i that has been removed by observing L_i, measured in terms of the MI, whose definition is recalled hereafter.

Definition 1 (Mutual Information). *Let* p, m *be two Probability Mass Function (p.m.f.) over the finite set* $\mathcal{Y}$.[4] *We denote by* $\mathsf{D_{KL}}(\mathsf{p} \parallel \mathsf{m})$ *the Kullback - Leibler (KL) divergence between* p *and* m*:*

$$\mathsf{D_{KL}}(\mathsf{p} \parallel \mathsf{m}) = \sum_{y \in \mathcal{Y}} \mathsf{p}(y) \log_2 \left(\frac{\mathsf{p}(y)}{\mathsf{m}(y)} \right) \quad . \tag{1}$$

Then, we define the Mutual Information (MI) between a discrete random variable Y *and a continuous random vector* $\mathbf{L}$ *as follows:*

$$\mathsf{MI}(\mathrm{Y}; \mathbf{L}) = \mathop{\mathbb{E}}_{\mathbf{L}} \left[\mathsf{D_{KL}} \left(\mathsf{p}_{\mathrm{Y} \mid \mathbf{L}} \,\middle\|\, \mathsf{p}_{\mathrm{Y}} \right) \right] \quad , \tag{2}$$

where p_{Y} *and* $\mathsf{p}_{\mathrm{Y} \mid \mathbf{L}}$ *respectively denote the Probability Mass Function (p.m.f.) of* Y *and the p.m.f. of* Y *given a realization* $\mathbf{l}$ *of the random vector* $\mathbf{L}$*, with the expectation taken over* $\mathbf{L}$*.*

In the remaining of this paper, we will assume that for each share Y_i and its corresponding sub-leakage L_i, we have a bound $\mathsf{MI}(\mathrm{Y}_i; \mathrm{L}_i) \leq \delta_i$. Intuitively, the lower the δ_i, the less informative the leakages, and the lower the SR. Moreover, we do not focus on the potential implementation overhead of masking in this paper — that could grow quadratically in memory and runtime [15] — to only focus on the security aspect of the counter-measure.

[3] For cryptographic reasons, the vast majority of the intermediate computations are uniformly distributed, including the inputs and outputs of Sbox — provided that the plaintext and the key are uniformly distributed as well. The only non-uniform intermediate computations of a block cipher may be the potential intermediate calculations of an Sbox.

[4] We assume without loss of generality that m has full support over $\mathcal{Y}$.

2.2 Problem and Conjecture

The problem that we consider here is to obtain upper bounds of the shape:

$$N_a(\mathsf{SR}, \mathcal{Y}) \geq \frac{f(\mathsf{SR}, \mathcal{Y})}{\prod_{i=0}^{d} (\delta_i/\tau)^r} \ , \tag{3}$$

where $f(\mathsf{SR}, \mathcal{Y})$ is a constant, τ is the so-called *noise threshold*, *i.e.*, the maximum amount of leakage that can leak such that the masking counter-measure remains sound and r is the *amplification rate*. Duc *et al.* [9] conjectured that N_a satisfies an upper bound of the shape of Equation 3, where $\tau \approx 1$ and $r = 1$.

A Reduction to Mutual Information Maximization. At Ches 2019, Chérisey *et al.* have shown that N_a can be linked to the MI as follows:

$$N_a(\mathsf{SR}, \mathcal{Y}) \geq \frac{f(\mathsf{SR}, \mathcal{Y})}{\mathsf{MI}(\mathrm{Y}; \mathbf{L})} \ , \tag{4}$$

where f is a known, computable function of SR that can be bounded based on the entropy of Y so that $f(\mathsf{SR}, \mathcal{Y}) = \mathcal{O}(\log(M))$ [7]. In other words, it is possible to reduce the problem of bounding the security level of masked implementation to the problem of bounding the MI:

$$\begin{aligned} \max_{\Pr(\mathrm{L}_i \mid \mathrm{Y}_i), i \in [\![0,d]\!]} \quad & \mathsf{MI}(\mathrm{Y}; \mathbf{L}) \\ \text{s.t.} \quad & \mathsf{MI}(\mathrm{Y}_i; \mathrm{L}_i) \leq \delta_i \end{aligned} \ . \tag{5}$$

Following the previous conjecture, we expect that $\mathsf{MI}(\mathrm{Y}; \mathbf{L}) \approx \prod_{i=0}^{d} (\delta_i/\tau)^r$ is a valid upper bound for this problem, where $\tau \approx 1$, and $r = 1$, whereas it could so far only be proven that $\mathsf{MI}(\mathrm{Y}; \mathbf{L}) \approx M \prod_{i=0}^{d} (\delta_i/\tau)$.

We note that the optimization defined in Equation 5 is convex, with convex constraints, as stated hereafter.[5]

Proposition 1. *The optimization problem defined in Equation 5 is convex.*

Proof. Let $\mathbf{l}$ be fixed. The mapping

$$\Pr(\mathrm{Y}_0 \mid \mathrm{L}_0 = l_0), \ldots, \Pr(\mathrm{Y}_d \mid \mathrm{L}_d = l_d) \mapsto \Pr(\mathrm{Y} \mid \mathbf{L} = \mathbf{l})$$

is a convolution product [21, Prop. 1] so it is $(d+1)$-linear, and thereby convex. Hence, since the mapping $\Pr(\mathrm{Y} \mid \mathbf{L} = \mathbf{l}) \mapsto -\mathsf{H}(\mathrm{Y} \mid \mathbf{L} = \mathbf{l})$ is also convex, the composition of both mappings remains convex. Since $\Pr(\mathrm{Y} \mid \mathbf{L}) \mapsto -\mathsf{H}(\mathrm{Y} \mid \mathbf{L})$ is the expectation of the latter composed mappings, it remains convex. Adding $\mathsf{H}(\mathrm{Y}) = \log_2(M)$ keeps the convexity property unchanged. □

As a result of this convexity, the optimal solution to the optimization of Equation 5 is necessarily such that for each $i \in [\![0, d]\!]$, we have $\mathsf{MI}(\mathrm{Y}_i; \mathrm{L}_i) = \delta_i$.

[5] The interested reader may find a similar convexity result, stated in terms of statistical distance, in the works of Dziembowski *et al.* [12, Cor. 2].

Serial vs. Parallel Leakages. In this section, we have implicitly assumed that the leakages occured in serial, which mostly depicts what could happen in a software implementation. We stress that our results may also extend without loss of generality to leakages occuring in parallel, *e.g.*, leakages of the form $\mathbf{L} = \mathrm{L}_0 + \ldots + \mathrm{L}_d$, provided that the independence assumption remains verified. It suffices to reduce to the serial case, thanks to the Data Processing Inequality (DPI):

$$\mathsf{MI}(\mathrm{Y}; \mathrm{L}_0 + \ldots + \mathrm{L}_d) \leq \mathsf{MI}(\mathrm{Y}; \mathrm{L}_0, \ldots, \mathrm{L}_d) \quad .$$

3 A Proof without Field Size Loss

We now provide our main result, namely we give a solution to the optimization problem stated in Equation 5. Compared to previous works, we introduce a mild additional assumption on the group $\mathcal{Y}$, namely that its order is a power of two. Nevertheless, this assumption covers Boolean masking and arithmetic masking modulo 2^n. To this end, we need to introduce some definitions.

3.1 Introducing Mrs. Gerber's Lemma

We first recall the definition of the entropy for a binary random variable.

Definition 2 (Binary Entropy). *Let*

$$H_b : \left| \begin{array}{l} [0,1] \longrightarrow [0,1] \\ \quad p \longmapsto -p \log_2(p) - (1-p) \log_2(1-p) \end{array} \right.$$

be the binary entropy function.Let $H_b^{-1} : [0,1] \mapsto \left[0, \frac{1}{2}\right]$ *be the inverse of* H_b *restricted to* $\left[0, \frac{1}{2}\right]$.

Likewise, we introduce the convolution for a binary random variable.

Definition 3 (Binary Convolution $\star$). *Let*

$$\star : \left| \begin{array}{l} [0,1]^2 \longrightarrow [0,1] \\ \quad x, y \longmapsto (1-x)y + x(1-y). \end{array} \right.$$

Note that when $\star$ is iterated, it can be replaced by a product, as stated next.

Proposition 2 (Iterated Star for Bias). *For* $x_0, \ldots, x_d \in [0,1]$*, the* $\star$ *operations can be mapped into a product for operands in the form of a bias as follows*

$$\overset{d}{\underset{i=0}{\star}} \left(\frac{1}{2} - x_i \right) = \frac{1}{2} - 2^d \prod_{i=0}^{d} x_i.$$

Proof. This is proved by induction on d. □

Definition 4 (Mrs. Gerber's functions). *For any positive integers n, p, let $\mathrm{f}_{\mathsf{H},2^n} : [0,1]^{p+1} \to [0,1]$ be the function defined by*

$$\mathrm{f}_{\mathsf{H},2^n}(x_0, \ldots, x_p) = H_b\Big(\overset{p}{\underset{i=0}{\star}} H_b^{-1}(x_i)\Big) \quad .$$

Moreover, we also define the function $\mathrm{f}_{\mathsf{MI},2^n} : [0,1]^{p+1} \to [0,1]$ as

$$\mathrm{f}_{\mathsf{MI},2^n}(\delta_0, \ldots, \delta_p) = 1 - \mathrm{f}_{\mathsf{H},2^n}(1-\delta_0, \ldots, 1-\delta_p)\,.$$

Remark 1. The function f_{MI} is decreasing with respect to each of its inputs, and is equal to 0 when every $\delta_i = 0$.

We are now equipped to introduce the technical lemma that will set the ground for our result, namely the so-called MGL. MGL has been first established by Wyner and Ziv [30] for a two-element group $\mathcal{Y}$, but it has been extended to any Abelian group whose order is a power of two by Jog and Anantharam [17].

Theorem 1 (Mrs. Gerber's Lemma [17, Thm. V.1, Claim V.1]). *Let $(\mathcal{Y}, \oplus)$ be any Abelian group of order $M = 2^n$. Let $\mathrm{Y}_0, \ldots, \mathrm{Y}_d$ be $d+1$ independent $\mathcal{Y}$-valued random variables with side information $\mathrm{L}_0, \ldots, \mathrm{L}_d$. We assume that conditionally to Y_i, L_i is independent of any other random variable. Define $x_i = \mathsf{H}(\mathrm{Y}_i \mid \mathrm{L}_i)$, and without loss of generality assume that $x_0 \geq \ldots \geq x_d$. Let $k = \lfloor x_0 \rfloor$ and $p = \max\{i | \lfloor x_i \rfloor \geq k\}$, then*

$$k + \mathrm{f}_{\mathsf{H},2^n}(x_0 - k, \ldots, x_p - k) \leq \mathsf{H}(\mathrm{Y}_0 \oplus \ldots \oplus \mathrm{Y}_d \mid \mathrm{L}_0, \ldots, \mathrm{L}_d)\,. \tag{6}$$

Proof. Let us denote $\mathrm{Y} = \mathrm{Y}_0 \oplus \ldots \oplus \mathrm{Y}_d$ for short, and for any $i \in [\![0, d]\!]$, let $\mathrm{X}_i(l) = \mathsf{H}(\mathrm{Y}_i \mid \mathrm{L}_i = l)$, such that $\underset{\mathrm{L}_i}{\mathbb{E}}[\mathrm{X}_i(\mathrm{L}_i)] = x_i$. Moreover, notice that by assumption, all the $\mathrm{X}_i(\mathrm{L}_i)$ are mutually independent.

Jog and Anantharam claim [17, Thm. V.1] that for a fixed leakage $\mathbf{l} = (l_0, \ldots, l_d)$, we have

$$\varphi(\mathrm{X}_0(l_0) \ldots, \mathrm{X}_d(l_d)) \leq \mathsf{H}(\mathrm{Y} \mid \mathbf{L} = \mathbf{l}) \quad ,$$

for some function φ that is convex with respect to each variable, when the remaining are kept fixed [17, Cor. V.1]. Combining this property with the independence of the $\mathrm{X}_i(\mathrm{L}_i)$, we may apply Jensen's inequality $d+1$ times:

$$\varphi(x_0, \ldots, x_d) \leq \underset{\mathbf{l}}{\mathbb{E}}\left[\varphi(\mathrm{X}_0(\mathbf{l}_0) \ldots, \mathrm{X}_d(\mathbf{l}_d))\right] \leq \underset{\mathbf{l}}{\mathbb{E}}\left[\mathsf{H}(\mathrm{Y} \mid \mathbf{L} = \mathbf{l})\right] = \mathsf{H}(\mathrm{Y} \mid \mathbf{L}) \quad .$$

Finally, replacing φ by its expression from [17, Thm. 5.1] results in Equation 6. □

Remark 2. In this paper, for better readability, all the logarithms are taken in base 2, but all the results we rely on have been established with logarithms in natural base. Thankfully, the proof of the MGL for $M = 2$ can be straightforwardly extended to logarithms in any base [5, Thm. 1]. Likewise, all the technical results used in Jog and Anantharam's proof remain insensitive to the base, as they essentially involve computing ratios of logarithms [17, Sec. 2].

3.2 Application of Mrs. Gerber's Lemma to Masking

Using the MGL, we prove the following upper bound on the side-channel information leaked by a masked encoding.

Corollary 1 (Security of Masking). *Let $M = 2^n$ and d be a positive integer. Let $\mathrm{Y}_0, \ldots, \mathrm{Y}_d$ be a $(d+1)$-sharing of the uniform random variable Y and $\mathbf{L} = (\mathrm{L}_0, \ldots, \mathrm{L}_d)$ be such that, conditionally to Y_i, the variable L_i is independent of the others. For all $i \in [\![0, d]\!]$, define $\mathsf{MI}(\mathrm{Y}_i; \mathrm{L}_i) = \delta_i$, and assume without loss of generality that there is a positive integer p such that for all $i \leq p$, $\delta_i \leq 1$ and for all $i > p$, $\delta_i \geq 1$. Then*

$$\mathsf{MI}(\mathrm{Y}; \mathbf{L}) \leq \mathrm{f}_{\mathsf{MI},2^n}(\delta_0, \ldots, \delta_p) \quad . \tag{7}$$

Proof. We upper-bound $\mathsf{MI}(\mathrm{Y}; \mathbf{L}) = \mathsf{H}(\mathrm{Y}) - \mathsf{H}(\mathrm{Y} \mid \mathbf{L}) = n - \mathsf{H}(\mathrm{Y} \mid \mathbf{L})$ by lower-bounding $\mathsf{H}(\mathrm{Y} \mid \mathbf{L})$, using Theorem 1.

$$\begin{aligned}
\mathsf{H}(\mathrm{Y} \mid \mathbf{L}) &= \mathsf{H}(\mathrm{Y}_0 \oplus \ldots \oplus \mathrm{Y}_d \mid \mathbf{L}) \\
&\geq n - 1 + \mathrm{f}_{\mathsf{H},2^n}(\mathsf{H}(\mathrm{Y}_0 \mid \mathrm{L}_0) - (n-1), \ldots, \mathsf{H}(\mathrm{Y}_p \mid \mathrm{L}_p) - (n-1)) \\
&= n - 1 + \mathrm{f}_{\mathsf{H},2^n}(1 - \mathsf{MI}(\mathrm{Y}_0; \mathrm{L}_0), \ldots, 1 - \mathsf{MI}(\mathrm{Y}_p; \mathrm{L}_p)) \\
&= n - \mathrm{f}_{\mathsf{MI},2^n}(\mathsf{MI}(\mathrm{Y}_0; \mathrm{L}_0), \ldots, \mathsf{MI}(\mathrm{Y}_p; \mathrm{L}_p))
\end{aligned}$$

□

3.3 Comparison with Former Upper Bounds

The removal of the field size loss in Theorem 1 is illustrated by Figure 1. The graph depicts the upper bounds on $\mathsf{MI}(\mathrm{Y}; \mathrm{L})$ (in bits) with respect to the noise parameter δ — assuming that the δ_i are all equal. The dotted curves correspond to the bounds given by Masure *et al.* [22][6] for different masking orders, whereas the dashed curves are obtained with our new bound. It can for example be noticed that for $d = 1$ and $\delta_i = 2^{-7}$, the bound from Ito *et al.* and Masure *et al.* [22] is roughly equal to 2^{-5}, whereas our upper bound is less than 2^{-12}, meaning that the gain is roughly 2^{12-5} which corresponds to half the field size. A similar factor is observed for larger d values.

We also add the following proposition (proven in Appendix) that gives a more intuitive view of our results and makes the removal of the field size loss explicit.

[6] We have afterwards noticed that the proof of Masure *et al.*'s bound [22, Prop. 2, Thm. 3] was suboptimal, so the bound of Masure *et al.* is actually slightly better than Ito *et al.*'s one by a factor $\frac{1}{2}$. That is why in the remaining of this paper, we mainly compare against the bound of Masure *et al.*.

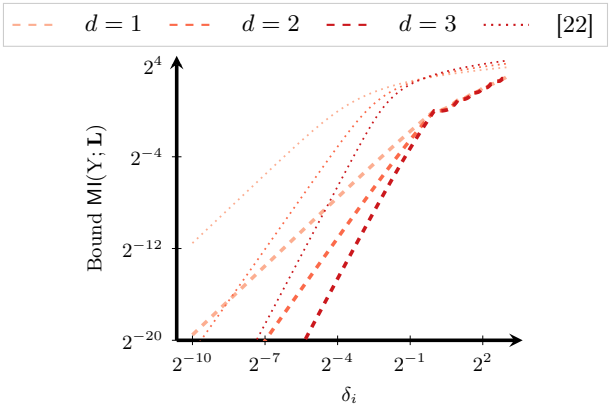


Fig. 1: Illustration of Equation 1 for $M = 256$ (e.g., the AES S-box).

Proposition 3 (Approximation in 0). *The Taylor expansion of the MGL function is the following:*

$$\mathrm{f}_{\mathsf{MI},2^n}(\delta_0, \ldots, \delta_d) = \eta \prod_{i=0}^{d} \frac{\delta_i}{\eta} + o\left(\prod_{i=0}^{d} \delta_i\right) \quad , \tag{8}$$

where $\eta = (2 \ln 2)^{-1} \approx 0.72$.

We note that the η parameter does not exactly correspond to the noise rate τ of subsection 2.2, since it depends on the noise level. But for high noise levels, where the first-order Taylor expansion is accurate, its value of 0.72 corresponds to the noise threshold in the CCS 2022 and the CARDIS 2022 papers.[7]

3.4 The MGL: Tighter or Tight?

We have shown in subsection 3.3 that our upper bound obtained from the MGL is tighter than the one achieved by Ito *et al.* [16] and Masure *et al.* [22]. We may therefore wonder to what extent the new MI upper bound is *tight*. In other words, are there some leakage models such that the MI between the secret and the leakage of all shares equals the MGL function. In this respect, Jog and Anantharam's results could be interpreted as the fact that the bound given by the MGL is at least *locally* tight, as stated hereafter.

Proposition 4 ([17, Thm. 5.1]). *For all* $(x_0, \ldots, x_d) \in [0, n]^{d+1}$*, there exists a leakage distribution* $(\mathbf{L} \mid \mathrm{Y})$ *such that:*

1. *For all* $i \in [\![0, d]\!]$*, we have* $\mathsf{H}(\mathrm{Y}_i \mid \mathrm{L}_i = l_i) = \delta_i$ *and*
2. $\mathsf{H}(\mathrm{Y} \mid \mathbf{L} = (l_0, \ldots, l_d)) = k + \mathrm{f}_{\mathsf{H},2^n}(x_0 - k, \ldots, x_p - k)$,

[7] For low noise levels, it gets gradually closer to one, but this gain has limited practical relevance since masking only provides high security with sufficient noise.

where k and p are the parameters defined in Theorem 1.

In other words, without further assumption on the leakage model, the bound given by the MGL is the best possible. We next investigate whether it is actually tight for practically-relevant leakage functions by confronting the bounds from the MGL to the direct computation of the MI for a shared secret. For this purpose, we assume that each share leaks a deterministic function of its value with an additive Gaussian noise, similarly to the experiments conducted by Ito *et al.* [16, Sec. 7.1] and Masure *et al.* [22, Sec. 3.1]. In particular, we consider two deterministic leakages, namely the Least Significant Bit (l.s.b.) of the share, or its Hamming weight. The MI is estimated with Monte-Carlo methods by sampling $N_v = 10,000$ leakages. Then, for each simulated leakage, the conditional p.m.f. can be exactly computed using a Soft-Analytical SCA (SASCA) [28].[8]

The results are depicted in Figure 2, for Boolean sharings with 2 shares and 3 shares, and for l.s.b. (Figs. 2a, 2b) and Hamming weight leakages (Figs. 2c, 2d). Each plot depicts the MI of the secret, depending on the variance σ^2 of the additive Gaussian noise. As one can observe on Figure 2a and Figure 2b, the bounds obtained by the MGL, depicted in dashed curves, are tight with the plain curves computed from the SASCA for the l.s.b. leakage model. However, for the Hamming weight leakage model, we observe a gap between our upper bound and the ground truth. Moreover, the gap between the dashed curve and the plain curve in Figure 2d seems wider than the one in Figure 2c. This shows that the Hamming weight leakage model does not verify Proposition 4. The combination of these observations confirms that no significant improvements of the bound can be obtained without making additional assumptions on the leakage function.

3.5 Linking the MI with the Success Rate

Having upper bounded the MI between the secret and *one* side-channel trace, we may then lower bound the required number of queries for any SCA adversary, by leveraging Chérisey *et al.*'s $f(\mathsf{SR}, \mathcal{Y})$ function, as stated hereafter.

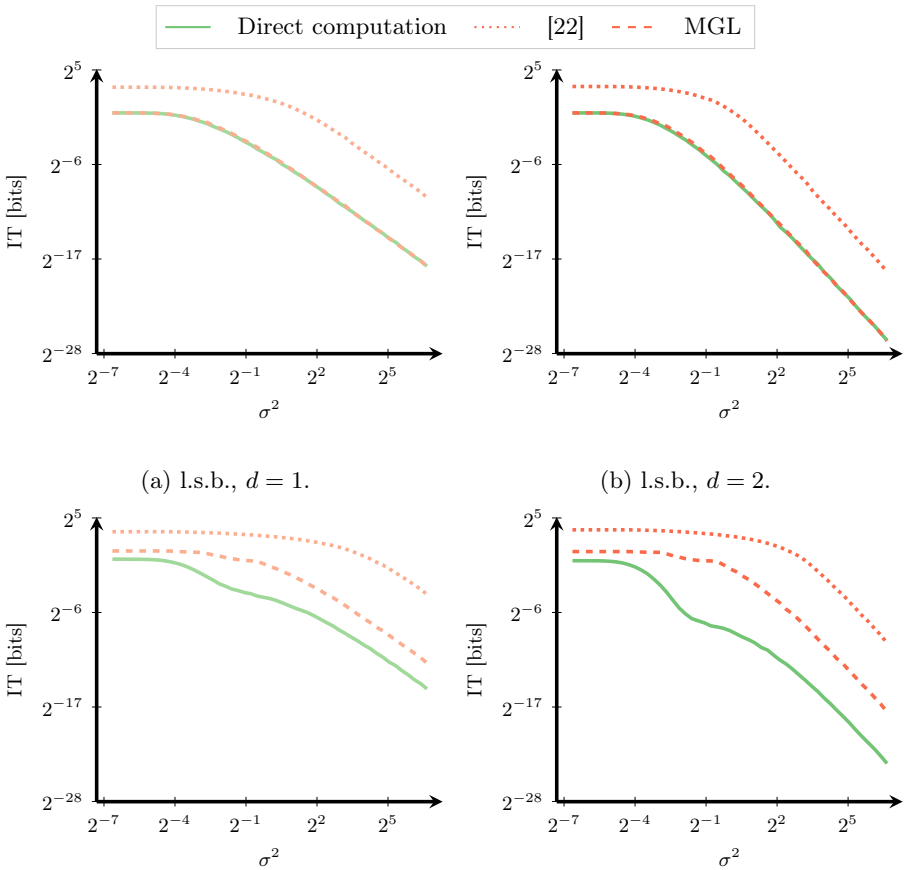
Corollary 2. *In the same setting as in Corollary 1,*

$$N_a(\mathsf{SR}) \geq \frac{f(\mathsf{SR}, \mathcal{Y})}{\mathrm{f}_{\mathsf{MI},2^n}(\delta_0, \ldots, \delta_p)} \cdot \tag{9}$$

Proof. Combining Corollary 1 with Equation 4. □

We compare this approach with a simulated SASCA attack on Figure 3, for the two leakage models investigated in subsection 3.4. The plain curves denote the attack complexity obtained from a key recovery. There, the success rate is estimated with *re-sampling* from a validation set of $N_v = 10,000$ traces. More precisely, the N_v validations traces are re-shuffled between 100 and $1,000$ times to emulate different attack sets. While this method is prone to be biased when N_a is close to N_v, the method remains sound if the success rate converged

[8] `https://scalib.readthedocs.io/en/latest/index.html`.



(a) l.s.b., $d = 1$.

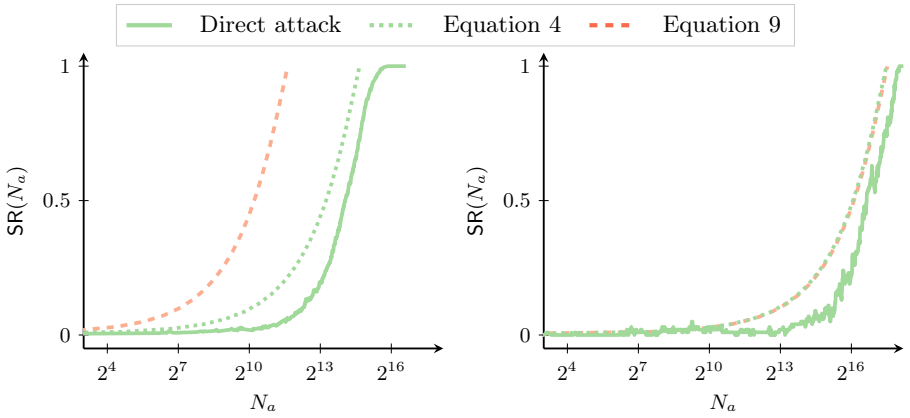
(b) l.s.b., $d = 2$.

(c) HW, $d = 1$.

(d) HW, $d = 2$.

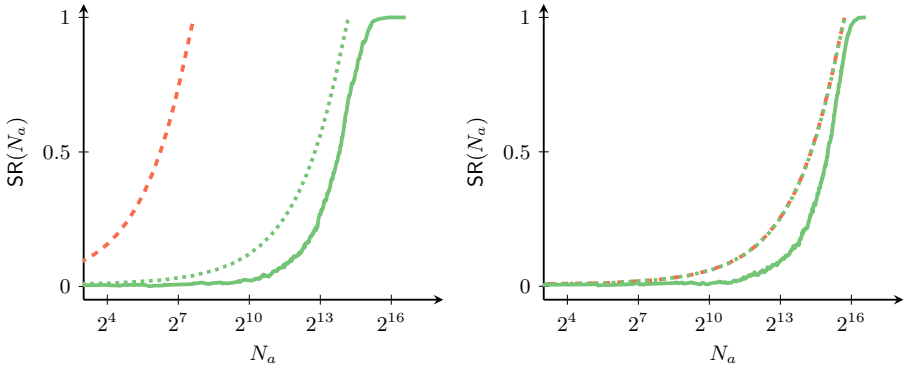
Fig. 2: MI in function of the Gaussian noise variance σ^2, for $n = 8$ bits.

towards 1 within N_v traces, as it cancels the bias.[9] The dotted green curves correspond to Equation 4 where the direct estimation of the MI between the shared secret and the leakage of the shares (from Figure 2) is used. The dashed red curves correspond to the bound given by Equation 9. One can notice that the



(a) HW. (b) l.s.b..

$\sigma^2 = 2^5, d = 1.$



(c) HW. (d) l.s.b..

$\sigma^2 = 2^2, d = 2.$

Fig. 3: Extending MI bounds to concrete security bounds.

plain curves and the dotted curves are always close to each other, meaning that Chérisey *et al.*'s function is reasonably tight in our context. Moreover, similarly

[9] This condition is verified retrospectively on Figure 3.

to what was noticed in subsection 3.4, the bound provided by Equation 9 is tight for the l.s.b. leakage model, but remains non-tight for the HW leakage model.

4 On the Dependence of the Group Structure

In our previous derivations, we assume that the field in which masking is applied is a power of two. Since this is the only limitation compared to the results of Ito *et al.* [16] and Masure *et al.* [22], we finally discuss whether this additional assumption is crucial. To this end, we show that Masure *et al.*'s approach using Pinsker [6, Lemma 11.6.1] and reverse Pinsker [26, Thm. 1] inequalities can be improved using the theory of *majorization* [20].

In a nutshell, majorization can be seen as a partial order relationship on p.m.f.'s quantifying "how spread out" a p.m.f. is, compared to another. The most spread out p.m.f. is the uniform distribution, so it can be used to assess how close to uniform a given p.m.f. is. Hereupon, Rioul recently characterized *optimal* Pinsker-like and reversed Pinsker-like inequalities [25]. While the optimal Pinsker inequality does not improve upon Pinsker's inequality, the optimal reverse Pinsker does improve it. Leveraging this improvement, the results of Masure *et al.* [22] are refined for arbitrary field size, as stated hereafter.

Theorem 2 (Informal). *Let $\mathcal{Y}$ be a group of order M, and $\mathrm{Y}, \mathbf{L}$ denote the joint distribution of a $d+1$-shared secret and its corresponding leakage. Let $\tau = (2\log(2))^{-1} \approx 0.72$, and let $P = \frac{1}{4}\prod_{i=0}^{d} \mathsf{MI}(\mathrm{Y}_i; \mathrm{L}_i)\,\tau^{-1}$. Then, for any $\alpha \in [0,1]$, there exists a constant $C_\alpha \in [\log_2(M), 2M]$ such that*

$$\mathsf{MI}(\mathrm{Y}; \mathbf{L}) \leq C_\alpha P^\alpha \ . \tag{10}$$

In particular, for $\alpha = \frac{1}{2}$, we have

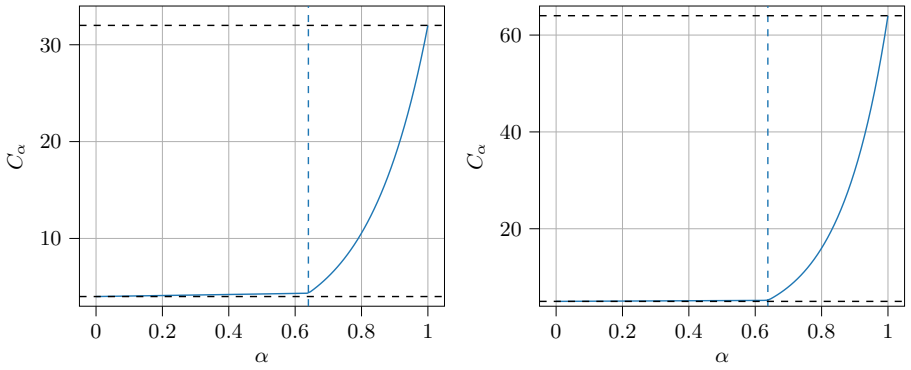
$$\mathsf{MI}(\mathrm{Y}; \mathbf{L}) \leq \log(M)\left(1 + \frac{1}{M}\right) P^{1/2} \ . \tag{11}$$

The bounds in Theorem 2 — whose formal statement is given and proven in Appendix — are not as tight as the ones from Corollary 1 but hold for any field size M, which makes them interesting when M is not a power of two.

Figure 4 depicts the range of C_α depending on α. It illustrates that there is a trade-off between the constant factor overhead C_α and the effective number of shares $\alpha \cdot (d+1)$. Overall, this provides good hints towards the conjectured absence of constant factor overhead in non-binary fields, and opens some perspectives towards a formal proof of the masked encoding bound in this context.

5 Conclusion and Perspectives

From a practical perspective, our work contributes to formalizing the soundness of so-called shortcut evaluations, where the security level of an implementation protected with higher-order masking is assessed based on the security of its



(a) $M = 16$. (b) $M = 32$.

Fig. 4: C_α for $M = 16, 32$. The two black dashed horizontal lines are at $C_\alpha = \log M$ and $C_\alpha = 2M$. The dashed vertical line is at $\alpha = \frac{\log(M \log M)}{2 \log(M-1)}$ and distinguishes two regimes for C_α, a logarithmic one and a polynomial one.

individual shares. By performing our proofs directly with noisy leakages, we show that such shortcuts are actually tight for masked encodings.

As mentioned in introduction, a natural extension of this work is to explore the tightness of bounds for masked computations (e.g., multiplications) and not only encodings. Besides, our results of subsection 3.4 show that while the bound we provide is locally tight (i.e., tight for some leakage functions), it is not tight for other practically-relevant leakage functions like the Hamming weight one (and, in general, for leakage functions having preimages of different sizes). It could therefore be interesting to study whether a mild characterization of the leakages could be used to improve the shortcut evaluation of masking for these functions. Another possible track of research is to study whether improved connections between the mutual information and the success rate can be obtained: despite the bounds of subsection 3.5 already give good evaluations, there remains a small gap that could possibly be removed (e.g., taking advantage of other information theoretic metrics like the Alpha-Information [18]). Eventually, yet another question is whether these bounds, for now studied in a known (random) plaintext context cover adaptive chosen-plaintext side-channel attacks [29]?

Acknowledgments

François-Xavier Standaert is a Senior Research Associate 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). This work has also partly benefited from the bilateral MESRI-BMBF project “APRIORI” from the ANR cybersecurity 2020 call. The authors also acknowledge financial support from the

French national Bank (BPI) under Securyzr-V grant (Contract DOS0144216/00),
a RISC-V centric platform integrating security co-processors.

A Proof of Proposition 3

$$\begin{aligned}
\mathsf{MI}(Y;\mathbf{L}) &\leq 1 - H_b\left(\frac{1}{2} - 2^d \prod_{i=0}^{d}\left(\frac{1}{2} - H_b^{-1}\big(1 - \mathsf{MI}(Y_i;\mathbf{L})\big)\right)\right) \\
&= \frac{2\log(e)}{4}\prod_{i=0}^{d} 4(\frac{1}{2} - H_b^{-1}(1-\mathsf{MI}(L_i;Y_i)))^2 \\
&\quad + o\Big(\prod_{i=0}^{d} 4(\frac{1}{2} - H_b^{-1}(1-\mathsf{MI}(L_i;Y_i)))^2\Big) \\
&= \frac{2\log(e)}{4}\prod_{i=0}^{d} 4\frac{\mathsf{MI}(L_i;Y_i)}{2\log e} + o\Big(\prod_{i=0}^{d} 4\frac{\mathsf{MI}(L_i;Y_i)}{2\log e}\Big)
\end{aligned}$$

That it under normalized form

$$\begin{aligned}
\mathsf{MI}(Y;\mathbf{L}) &\leq \frac{\log(e)}{2}\prod_{i=0}^{d}\mathsf{MI}(L_i;Y_i)\,\frac{2}{\log e} + o\Big(\prod_{i=0}^{d}\mathsf{MI}(L_i;Y_i)\Big) \\
&= \eta\prod_{i=0}^{d}\frac{\mathsf{MI}(L_i;Y_i)}{\eta} + o\Big(\prod_{i=0}^{d}\mathsf{MI}(L_i;Y_i)\Big).
\end{aligned}$$

B Technical Statements and Proofs from section 4

Proposition 5 (Optimal Reversed Pinsker). *Let f_{opt} be the optimal reverse Pinsker inequality, i.e.,*

$$\begin{aligned}
f_{\mathsf{opt}} : \Big| \begin{array}{rl} [0, \frac{1}{M}] & \longrightarrow \mathbb{R}_+ \\ \Delta & \longmapsto \frac{1}{M}((1+M\Delta)\log(1+M\Delta) \\ & \qquad +((1-\Delta)M - L)\log((1-\Delta)M - L)) \end{array} \quad (12)
\end{aligned}$$

where $L = \lfloor M(1-\Delta)\rfloor$. For all p.m.f. P we have $\mathsf{D}_{\mathsf{KL}}(P \parallel U) \leq f_{opt}(\Delta(P,U))$.

Proof. By applying the entropy which is Schur-concave to Eqn. 51 in [25]. We factor $-\log M$ in each term of the inequality to obtain Prop. 5. □

Theorem 3 (Formal version of Theorem 2). *Let $\mathcal{H}$ be the class of function that is lower bounded by f_{opt}, concave when composed with a square root and increasing. Let $P = \frac{1}{4}\prod_{i=0}^{d} C\,\mathsf{MI}(Y_i;L_i)$, we have*

$$\mathsf{MI}(Y;\mathbf{L}) \leq \inf_{f\in\mathcal{H}}(f \circ \sqrt{\ })(P). \quad (13)$$

Let $C_\alpha = \sup_{\Delta \in]0,1-\frac{1}{M}]} f^*(\Delta)\Delta^{-2\alpha} = \max_{\Delta = k/M, k\in\{1,\dots M-1\}} f^*(\Delta)\Delta^{-2\alpha}$. *We have*

$$\mathsf{MI}(Y;\mathbf{L}) \leq \min\left(\log(1+M^2(4^{\frac{1}{M}}-1)P), (\frac{1}{M}+\sqrt{P})\log(1+M\sqrt{P})\right) \tag{14}$$

$$\leq \inf_{\alpha\in[0,1]} C_\alpha \cdot P^\alpha \tag{15}$$

$$\leq \log(M)(1+\frac{1}{M})\cdot P^{\frac{1}{2}}. \tag{16}$$

Remark 3. The infinum in Eqn. 13 can be computed with the Legendre-Fenchel transform $f \mapsto f^*$ (i.e. $f^*(p) = \sup_x\{px - f(x)\}$). Indeed, it is given by $\Delta \mapsto -(-f_{\text{opt}} \circ \sqrt{\cdot})^{**}(\Delta^2)$ by applying Thm. 10 in [27].

The different inequalities are shown in Fig. 5. f_1 is the best for $\Delta \leq 1/M$ and else f_2 is the best. Eqn. 16 shows that if we reduce the security exponent to $\frac{1}{2}$ we can obtain a mild (logarithmic) dependency in the field size.

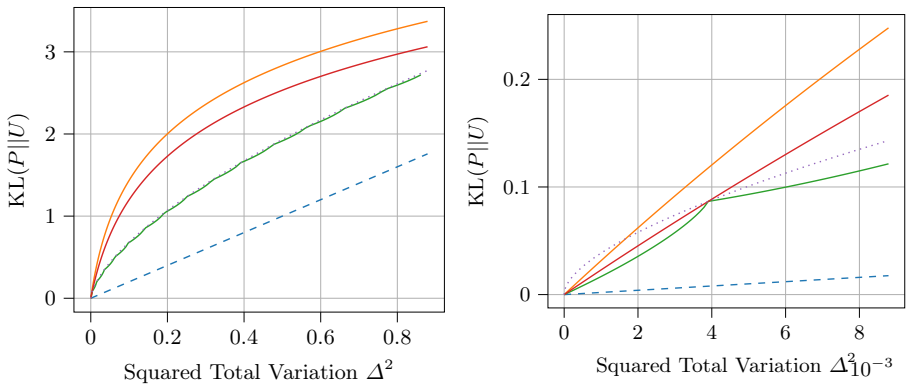


Fig. 5: Illustration of the inequalities for $M = 16$. Pinsker is the dashed blue line. The classical reverse Pinsker is the orange line. The optimal reverse Pinsker f_{opt} is the green curve. The dotted curve is f_2 and the red curve f_1.

Proof. All derivations of [22] hold for $f \in \mathcal{H}$ which shows Eqn. 13. Indeed,

$$\begin{aligned}
\mathrm{KL}(Y|\mathbf{L}||U) &\leq f_{\mathrm{opt}}(\Delta) && \textit{Prop. } 5\\
&\leq f(\Delta) && f_{\mathrm{opt}} \leq f\\
&\leq f(\frac{1}{2}\prod_{i=0}^{d} 2\Delta((Y_i|L_i);U)) && \text{XOR Lemma}\\
&= (f \circ \sqrt{\cdot})(\frac{1}{4}\prod_{i=0}^{d} 4\Delta((Y_i|L_i);U))^2\\
&\leq (f \circ \sqrt{\cdot})(\frac{1}{4}\prod_{i=0}^{d} C\,\mathrm{KL}(Y_i|L_i||U)) && \text{Pinsker}
\end{aligned}$$

Since $(f \circ \sqrt{\cdot})$ is concave, we apply Jensen inequality and take the expectation to obtain the desired inequality. The expectation of the product is simplified to the product of the expectations by independence of the terms. Let $f_1 : \Delta \mapsto \log(1 + M^2(4^{\frac{1}{M}} - 1)\Delta^2) \approx MC\Delta^2$ and $f_2 : \Delta \mapsto (\Delta + \frac{1}{M})\log(1 + M\Delta)$. We show that f_1 and f_2 are in $\mathcal{H}$. For f_2 it is clear since f_2 is f^* where we removed the negative $1/M$ periodic term. f_1 is clearly concave in Δ^2 and increasing. To ensure that $f_1 \geq f_{\mathrm{opt}}$ we consider the case of equality in $\frac{1}{M}$. This imposes $\log(1 + \beta_M M^{-2}) = \frac{2}{M}$ where $\beta_M = M^2(4^{\frac{1}{M}} - 1)$. For $\Delta \leq \frac{1}{M}$, $Mf_{\mathrm{opt}}(\Delta) = (1+M\Delta)\log(1+M\Delta)+(1-M\Delta)\log(1-M\Delta) \leq \frac{2}{M}\log(1+M^2\Delta^2)$ by Jensen in equality. This upper bound of f_{opt} is a lower bound of f_1. Since log is increasing the inequality holds if and only if $1+\beta_M\Delta^2 \geq (1+M^2\Delta^2)^{\frac{2}{M}}$. Equality holds in 0 and $1/M$ and we show that the derivative of the difference is increasing then decreasing. The derivative is given by $2\Delta(\beta_M - 2M(1 + M^2\Delta^2)^{\frac{2}{M}-1})$ and its sign is given by $\beta_M - 2M(1 + M^2\Delta^2)^{\frac{2}{M}-1}$. This quantity is positive in 0 and monotonically decreasing hence the result. It remains to prove the inequality for $\Delta \geq \frac{1}{M}$. To do so, we show that $f_1(\Delta) \geq \log(1 + M\Delta) \geq \frac{1+M\Delta}{M}\log(1 + M\Delta)$. Since log is increasing it is enough to have $1 + \beta_M\Delta^2 \geq 1 + M\ \Delta$ that is $\Delta \geq M/\beta_M$ i.e., $4^{\frac{1}{M}} - 1 \geq \frac{1}{M}$. This holds since $e^x - 1 \geq x$ by convexity of the exponential. This shows that $f_1 \in \mathcal{H}$ and Eqn. 14 is proved. Let $\mathcal{H}_{\mathrm{poly}} = \{f_\alpha : \Delta \mapsto C_\alpha\Delta^{2\alpha} | \alpha \in [0, 1]\}$, we show that $\mathcal{H}_{\mathrm{poly}} \subset \mathcal{H}$. Functions $\mathcal{H}_{\mathrm{poly}}$ are concave when composed with a square root since $\alpha \leq 1$, increasing since $\alpha \geq 0$ and lower bounded by f_{opt} by definition of C_α. This proves Eqn. 15. To prove Eqn. 16 we observe that $C_0 = \log(M)$, C_α is continuous and increasing in α. Consider the values of Δ for which the sup in the definition of C_α is reached. Since f_{opt} is square-root convex in the intervals $[k/M, (k + 1)/M]$ and $\Delta \mapsto C_\alpha\Delta^{2\alpha}$ is square-root concave we can only have equality in $\frac{k+1}{M}$. This shows that $C_\alpha = \max_{\Delta=k/M, k\in\{1,\dots M-1\}} f^*(\Delta)\Delta^{-2\alpha}$. We verify that the maximum is reached in $1 - 1/M$ when $\alpha = \frac{1}{2}$. The ratio of the sequence $f^*(k/M)(\frac{k}{M})^{-2\alpha}$ is larger than 1 which proves Eqn. 16. □

References

1. Azouaoui, M., Bellizia, D., Buhan, I., Debande, N., Duval, S., Giraud, C., Jaulmes, É., Koeune, F., Oswald, E., Standaert, F., Whitnall, C.: A systematic appraisal of side channel evaluation strategies. In: van der Merwe, T., Mitchell, C.J., Mehrnezhad, M. (eds.) Security Standardisation Research - 6th International Conference, SSR 2020, London, UK, November 30 - December 1, 2020, Proceedings. Lecture Notes in Computer Science, vol. 12529, pp. 46–66. Springer (2020). https://doi.org/10.1007/978-3-030-64357-7_3, https://doi.org/10.1007/978-3-030-64357-7_3
2. Benadjila, R., Prouff, E., Strullu, R., Cagli, E., Dumas, C.: Deep learning for side-channel analysis and introduction to ASCAD database. Journal of Cryptographic Engineering **10**(2), 163–188 (Jun 2020). https://doi.org/10.1007/s13389-019-00220-8
3. Bronchain, O., Standaert, F.X.: Side-channel countermeasures' dissection. IACR Transactions on Cryptographic Hardware and Embedded Systems **2020**(2), 1–25 (2020). https://doi.org/10.13154/tches.v2020.i2.1-25, https://tches.iacr.org/index.php/TCHES/article/view/8542
4. Chari, S., Jutla, C.S., Rao, J.R., Rohatgi, P.: Towards sound approaches to counteract power-analysis attacks. In: Wiener, M.J. (ed.) Advances in Cryptology – CRYPTO'99. Lecture Notes in Computer Science, vol. 1666, pp. 398–412. Springer, Heidelberg, Germany, Santa Barbara, CA, USA (Aug 15–19, 1999). https://doi.org/10.1007/3-540-48405-1_26
5. Cheng, F.: Generalization of Mrs. Gerber's lemma. Commun. Inf. Syst. **14**(2), 79–86 (2014). https://doi.org/10.4310/cis.2014.v14.n2.a1, https://doi.org/10.4310/cis.2014.v14.n2.a1
6. Cover, T.M., Thomas, J.A.: Elements of information theory (2. ed.). Wiley (2006)
7. 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
8. Duc, A., Dziembowski, S., Faust, S.: Unifying leakage models: From probing attacks to noisy leakage. In: Nguyen, P.Q., Oswald, E. (eds.) Advances in Cryptology – EUROCRYPT 2014. Lecture Notes in Computer Science, vol. 8441, pp. 423–440. Springer, Heidelberg, Germany, Copenhagen, Denmark (May 11–15, 2014). https://doi.org/10.1007/978-3-642-55220-5_24
9. 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
10. Duc, A., Faust, S., Standaert, F.X.: Making masking security proofs concrete (or how to evaluate the security of any leaking device), extended version. Journal of Cryptology **32**(4), 1263–1297 (Oct 2019). https://doi.org/10.1007/s00145-018-9277-0
11. 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

12. 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
13. 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
14. 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
15. Ishai, Y., Sahai, A., Wagner, D.A.: Private circuits: Securing hardware against probing attacks. In: Boneh, D. (ed.) Advances in Cryptology - CRYPTO 2003, 23rd Annual International Cryptology Conference, Santa Barbara, California, USA, August 17-21, 2003, Proceedings. Lecture Notes in Computer Science, vol. 2729, pp. 463–481. Springer (2003). https://doi.org/10.1007/978-3-540-45146-4_27, https://doi.org/10.1007/978-3-540-45146-4_27
16. 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.) Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security, CCS 2022, Los Angeles, CA, USA, November 7-11, 2022. pp. 1521–1535. ACM (2022). https://doi.org/10.1145/3548606.3560579, https://doi.org/10.1145/3548606.3560579
17. 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
18. Liu, Y., Cheng, W., Guilley, S., Rioul, O.: On conditional alpha-information and its application to side-channel analysis. In: ITW. pp. 1–6. IEEE (2021)
19. Maghrebi, H., Portigliatti, T., Prouff, E.: Breaking cryptographic implementations using deep learning techniques. In: Carlet, C., Hasan, M.A., Saraswat, V. (eds.) Security, Privacy, and Applied Cryptography Engineering - 6th International Conference, SPACE 2016, Hyderabad, India, December 14-18, 2016, Proceedings. Lecture Notes in Computer Science, vol. 10076, pp. 3–26. Springer (2016). https://doi.org/10.1007/978-3-319-49445-6_1, https://doi.org/10.1007/978-3-319-49445-6_1
20. Marshall, A.W., Olkin, I., Arnold, B.C.: Inequalities: Theory of Majorization and Its Applications. Springer Series in Statistics (1980)
21. Masure, L., Cristiani, V., Lecomte, M., Standaert, F.: Don't learn what you already know scheme-aware modeling for profiling side-channel analysis against masking. IACR Trans. Cryptogr. Hardw. Embed. Syst. **2023**(1), 32–59 (2023). https://doi.org/10.46586/tches.v2023.i1.32-59, https://doi.org/10.46586/tches.v2023.i1.32-59
22. Masure, L., Rioul, O., Standaert, F.X.: A nearly tight proof of duc et al.'s conjectured security bound for masked implementations. In: Buhan, I., Schnei-

der, T. (eds.) Smart Card Research and Advanced Applications. pp. 69–81. Springer International Publishing, Cham (2023). https://doi.org/10.1007/978-3-031-25319-5_4
23. 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
24. 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
25. Rioul, O.: What is randomness? the interplay between alpha entropies, total variation and guessing. Physical Sciences Forum **5**(1) (2022). https://doi.org/10.3390/psf2022005030, https://www.mdpi.com/2673-9984/5/1/30
26. Sason, I., Verdú, S.: Upper bounds on the relative entropy and Rényi divergence as a function of total variation distance for finite alphabets. In: 2015 IEEE Information Theory Workshop - Fall (ITW), Jeju Island, South Korea, October 11-15, 2015. pp. 214–218. IEEE (2015). https://doi.org/10.1109/ITWF.2015.7360766, https://doi.org/10.1109/ITWF.2015.7360766
27. Touchette, H.: Legendre-Fenchel transforms in a nutshell (2005), https://www.ise.ncsu.edu/fuzzy-neural/wp-content/uploads/sites/9/2019/01/or706-LF-transform-1.pdf
28. Veyrat-Charvillon, N., Gérard, B., Standaert, F.X.: Soft analytical side-channel attacks. In: Sarkar, P., Iwata, T. (eds.) Advances in Cryptology – ASIACRYPT 2014, Part I. Lecture Notes in Computer Science, vol. 8873, pp. 282–296. Springer, Heidelberg, Germany, Kaoshiung, Taiwan, R.O.C. (Dec 7–11, 2014). https://doi.org/10.1007/978-3-662-45611-8_15
29. Veyrat-Charvillon, N., Standaert, F.: Adaptive chosen-message side-channel attacks. In: ACNS. Lecture Notes in Computer Science, vol. 6123, pp. 186–199 (2010)
30. Wyner, A.D., Ziv, J.: A theorem on the entropy of certain binary sequences and applications-I. IEEE Transactions on Information Theory **19**, 769–772 (1973)