Formal Security Proofs via Doeblin Coefficients:
Optimal Side-Channel Factorization from Noisy Leakage to Random Probing

Julien Béguinot[1(✉)], Wei Cheng[1,2], Sylvain Guilley[1,2], and Olivier Rioul[1]

[1] LTCI, Télécom Paris, Institut Polytechnique de Paris, Palaiseau, France
{julien.beguinot,wei.cheng,sylvain.guilley,olivier.rioul}@telecom-paris.fr
[2] Secure-IC S.A.S., Paris, France

Abstract. Masking is one of the most popular countermeasures to side-channel attacks, because it can offer provable security. However, depending on the adversary's model, useful security guarantees can be hard to provide. At first, masking has been shown secure against *t-threshold probing adversaries* by Ishai *et al.* at Crypto'03. It has then been shown secure in the more generic *random probing model* by Duc *et al.* at Eurocrypt'14. Prouff and Rivain have introduced the *noisy leakage model* to capture more realistic leakage at Eurocrypt'13. Reduction from noisy leakage to random probing has been introduced by Duc *et al.* at Eurocrypt'14, and security guarantees were improved for both models by Prest *et al.* at Crypto'19, Duc *et al.* in Eurocrypt'15/J. Cryptol'19, and Masure and Standaert at Crypto'23. Unfortunately, as it turns out, we found that previous proofs in either random probing or noisy leakage models are flawed, and such flaws do not appear easy to fix.

In this work, we show that the *Doeblin coefficient* allows one to overcome these flaws. In fact, it yields optimal reductions from noisy leakage to random probing, thereby providing a correct and usable metric to properly ground security proofs. This shows the inherent inevitable cost of a reduction from the noisy leakages to the random probing model. We show that it can also be used to derive *direct* formal security proofs using the subsequence decomposition of Prouff and Rivain.

1 Introduction

1.1 Context

All cryptographic implementations leak some side information about the sensitive variables they manipulate through the so-called side-channels. These leakages can be of different natures: Timing [DKL+98], power consumption [KJJ99,KGG+18], electromagnetic [GMO01,AARR02]. The corresponding side-channel attacks can be very harmful if there is no countermeasure or if the countermeasure is not carefully implemented. One may classify countermeasures into three categories, that can be jointly implemented:

L. Reyzin and D. Stebila (Eds.): CRYPTO 2024, LNCS 14925, pp. 389–426, 2024.
https://doi.org/10.1007/978-3-031-68391-6_12

– **Key refreshing** regularly replaces the secret key by a new one, e.g., each time a given number of operations has been performed [AB00, UHIM24].
– **Hiding** equalizes the leakage, either by removing the variations caused by computation, or by creating artificial noise in the circuit. It can be achieved by physical means such as *shielding* [AARR02], *noise makers* [LBB19], *dual rail technology* [MSS09], *balancing* or adding *dummy operations* [LH20].
– **Noise amplification** leverages existing noise from the given side-channels to make their measurements harder. It can be achieved using *wire shuffling* [ISW03, CS21], *operation shuffling* [VCMKS12] or *masking* [ISW03].

Masking is one of the most effective countermeasures known so far. It is especially relevant because of its provable security [ISW03, RP10a, PR13, DDF14, DFS15, BBD+16, DFS19, BCG+23, MS23b]. Previously published security proofs for masking fall into two classes:

– **Simulation paradigm (indirect approach):** A black-box adversary is modeled by an algorithm that only accesses the public information, which corresponds to the usual cryptanalysis. By contrast, the side-channel adversary is given both public and side-channel information. If there always exists a black-box adversary whose output is indistinguishable from the side-channel adversary's output, then the implementation is considered secure.
– **Information Theoretic Paradigm (direct approach):** The implementation is considered secure if the mutual information (or some other informational metric) between the side-channel leakage and the corresponding sensitive variable is negligible, given the available public information. Equivalently, the required number of side-channel queries required to achieve a given success rate is prohibitively large for any attack.

Those two approaches have been respectively termed *indirect* and *direct* by Prest *et al.* [PGMP19]. Both approaches have their pros and cons. On the one hand, the security proof based on information theory is conceptually simpler and provides more realistic security parameters. On the other hand, the simulation-based approach is very generic and can be applied to a whole cipher at once. Indeed, as remarked in [DDF14, Footnote 4] a few pairs of plaintext/ciphertext completely reveal the key of an AES *in the information theoretic sense*. Hence block ciphers such as AES are not secure in this sense, a fortiori in the presence of side-channel information. The security of AES relies on a one-way computational assumption which cannot be taken into account in the information theoretic paradigm[1].

To prove the security of a cryptographic implementation in any of the two paradigms above, it is necessary to define the side-channel adversary's model. Some restrictions should be imposed on the adversary, since if she/he is allowed to observe all variables manipulated in the circuit, then the implementation would be trivially broken. Micali and Reyzin introduced *physically observable*

[1] Unless, for a given round in a divide-and-conquer attack, the round's output is assumed not disclosed to the attacker because it is hidden by the one-way computational assumption in the subsequent rounds.

cryptography [MR04], in which only computation can leak information. *Leakage resilient cryptography* [DP08,KR19] also considers memory leakage models. In this context, the simplest model is the *t-threshold probing model* [ISW03] in which the adversary is only allowed to probe the values of t wires within the circuit. In the more elaborated *region probing model* [GPRV22], the circuit is divided into small regions and the adversary can probe t values in each region. A more realistic model is the *random probing model* [DDF14] where the side-channels correspond to erasure channels. The most generic type of model is the *noisy leakage model* from Prouff and Rivain [PR13] where the $\mathcal{D}$*-noisy adversary* has minimal distortion $\mathcal{D}$ between the channel input and output. In this list of models, the security proof is all the more hard to establish as the model is more complex and realistic.

1.2 Contributions

In this paper, we aim at grounding security proofs of side-channel analysis countermeasures on solid mathematical foundations. This work has the following contributions.

1. We carry out a systematic mathematical study of the *complementary Doeblin coefficient* (CDC). This coefficient was originally used to study Markov chains [Doe37] and appeared in the side-channel literature as the value of ϵ in [DDF19, Eq. 9, Proof of Lemma 4]. We show that the CDC provides the optimal reduction from a noisy leakage model to the random probing model. Since the reduction is optimal it exhibits the unavoidable loss to pay to use a security proof based on the random probing model.
2. Bounds on the success rate (SR) and guessing entropy (GE) of a side-channel attack are derived using the CDC. Such bounds holds with equality for erasure side-channels, scale well with the number of channel queries, can be applied to adaptive adversaries and are amenable to practical evaluation, e.g., in Gaussian additive noise (Hamming weight or least significant bit model, etc.).
3. A new direct security proof is presented based on the CDC and on the Prouff-Rivain subsequence decomposition. As a supplementary material, some flaws in previous direct security proofs for masking in the noisy leakage model are identified (this does not necessarily mean that the corresponding results cannot hold) and some patches or bypasses are presented. Namely, this concerns Lemma 4 (hence Theorem 3) in [PR13], Lemma 8 (hence Theorem 6, Corollary 4) in [PGMP19], and Theorem 5 in [MS23b].
4. A new methodology providing indirect security proofs is presented, based on the optimal CDC reduction from the noisy leakage model to the random probing model. As a supplementary material, minor errors are also corrected from the original proof on a reduction to the t-threshold probing model of Lemma 4 in [DDF14]. As a result, the bounds derived in [DDF14,DFS15, DFS19,PGMP19] that leveraged this Lemma can be improved significantly.

1.3 Outline

The remainder of this paper is structured as follows. Preliminaries and mathematical results on channels, leakage measures (including CDC) and figures of merit are presented in Sect. 2. The key property of the CDC and the resulting bounds on figures of merit are presented in Sect. 3, along with some theoretical expressions for concrete evaluation. Direct security proofs leveraging CDC based on Prouff-Rivain subsequence decompositions are provided in Sect. 4. A new methodology for the derivation of indirect proofs is shown in Sect. 5. Section 6 concludes.

To fit page limitations most proofs and the descriptions of the existing flaws are deferred to the full version of the article [BCGR24] available at https://eprint.iacr.org/2024/199.

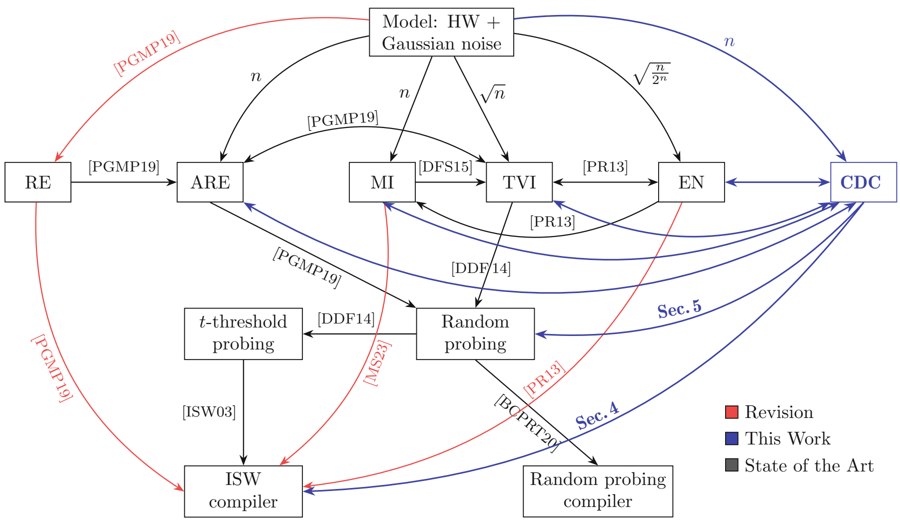


Fig. 1. Overview of formal security proofs, organized in four levels. Novelty is in blue, revisions of the state of the art is in red, and the state of the art is in black. (Color figure online)

Figure 1 updates Fig. 1 from [PGMP19] where the black arrows indicate the state of the art, the red arrows are flaws that we identified and revised and the blue arrows correspond to our new derivations using the CDC. The figure is organized in four levels: The top level corresponds to the Hamming weight leakage model. The second level contains the main leakage measures, corresponding to different noisy leakage adversaries. Each arrow label from the first to the second level indicates how the leakage measure scales with respect to the number n of bits, while each arrow label between two leakage models indicates the appropriate reference of the reduction from one model to another. The third layer contains the various adversarial models based on probing, and the bottom level contains the secure compilers that generate secure circuits against a given adversarial model. Each arrow label from the second or third to the fourth level indicates the appropriate reference of the corresponding security proof. The comparison with the other informational leakage measure is elaborated in Table 1 below.

1.4 Detailed Technical Overview

Theorem 1 provides the optimal factorization of a given channel into an erasure channel followed by another channel. As a consequence, any channel can be seen as a stochastically degraded erasure channel with the largest possible erasure probability. In particular, this implies that the optimal reduction of a side-channel adversary from the noisy leakage model (arbitrary channel) to the random probing model (erasure channel) is measured by the CDC.

The CDC equally applies to multivariate leakages. Lemma 6 shows that the CDC with multiple traces is bounded in terms of the CDC with one trace, even for an adaptive "chosen channel" adversary.

The main figures of merit (Definition 13) satisfy the data processing inequality (DPI) recalled in Lemma 8. In other words, the stochastically degraded adversary can only perform worse than the non-degraded one. As a consequence, the performance of any side-channel attacker can be bounded in terms of CDC as shown in Proposition 1. Intuitively, these bounds result from averaging two extreme cases: Either the leakage value is an erasure symbol and the figure of merit is that of a blind guess, or it is probed and the figure of merit is that of a disclosed value. This applies even for computationally bounded adversaries, allowing one to avoid complex simulation arguments [DDF14].

Lemma 7 shows that the CDC satisfies a strengthened data processing inequality which is useful in the derivation of the security proof.

Theorem 2 gives a direct security bound for ISW masked computations of an AES following the subsequence decomposition of Prouff and Rivain [PR13]. To achieve this, we derive a security lemma for each type of subsequence. For type 1 and 2 subsequences, we prove an analog of Mrs Gerber's Lemma (MGL) [WZ73] in Lemma 10 which shows that the CDC between the leakage and a masked value is upper bounded by the product of the CDCs share by share. This is expected since a sensitive value is probed if and only if all shares are probed. For type 3 subsequences, Lemma 11 provides a security bound for the cross-wise terms, in terms of the domination polynomial of the rook graph of Definition 17. The idea is that a value is probed if and only if all shares are probed at least once through cross-wise terms of one of the two shared inputs.

Theorem 4 explains how any formal security proof in the random probing model can be lifted to the noisy leakage model using the CDC. This is illustrated for the security proof of Duc et al. [DDF14].

The descriptions of the flaws appearing in previous derivations is deferred to the long version of the article [BCGR24]. To summarize, the derivations of [PGMP19, PR13] are invalidated because of an incorrect chain rule on probabilities. The flaw in [MS23b] is due to the fact that the bound appearing in the

MGL for mutual information is separately but not jointly convex in the variables. The CDC overcomes these difficulties by allowing a direct bound which is then degraded in terms of the corresponding leakage measure through Lemma 9.

2 Mathematical Framework

In this Section, we present the mathematical framework of side-channel analysis that we use in our analysis. The notations are given in Subsect. 2.1. The formal definition of a side-channel is given in Subsect. 2.2. The main informational leakages measures are recalled in Subsect. 2.3. Some useful properties of the complementary Doeblin coefficient (CDC) such as an adaptive single letterization and strengthened data processing inequality (DPI) are provided in Subsect. 2.4. The model for a side-channel attack is described in Subsect. 2.5. Finally, the figures of merit to evaluate the advantage of a side-channel adversary are introduced in Subsect. 2.6.

2.1 Notations

Random variables are denoted by uppercase letters like X, Y. The corresponding set of values taken by the random variables are denoted by the corresponding calligraphic letters like $\mathcal{X}, \mathcal{Y}$. Lowercase letters denote values taken by random variables, e.g., $x \in \mathcal{X}$, $y \in \mathcal{Y}$. Bold letters denote random vectors $\mathbf{X}$ taking vector values $\mathbf{x}$. The probability distribution of X is denoted P_X; we write $X \sim P_X$.

– When X is discrete, taking values in a discrete set $\mathcal{X}$ of cardinality $|\mathcal{X}|$, its probability mass function (pmf) is noted $p_X(x) = \mathbb{P}(X = x)$;
– When X is continuous, its probability density function (pdf) is also noted $p_X(x)$ where $\mathrm{d}P_X(x) = p_X(x)\,\mathrm{d}x$.

We use the unified notation $⨍$ which is a sum in the discrete case and an integral in the continuous case. Therefore, we write $\mathbb{P}(X \in E) = ⨍_{x \in E} p_X(x)$. Expectation is denoted by $\mathbb{E}_X[\cdot]$. The p-norm is noted $\|\cdot\|_p$.

– The uniform distribution on a set $\mathcal{X}$ is denoted by $\mathcal{U}(\mathcal{X})$;
– $\mathcal{B}(p)$ denotes the Bernoulli distribution with parameter p and $\mathcal{B}(n,p)$ denotes the Binomial distribution with parameters n, p. The survival function of $B \sim \mathcal{B}(n,p)$ is noted $Q_B(x,n,p) \triangleq \mathbb{P}(B > x) = \mathbb{P}(B \geqslant x+1)$.
– $\mathcal{N}(\mu, \sigma^2)$ denotes the Gaussian distribution of mean μ and variance σ^2. The survival function of the standard Gaussian $\mathcal{N}(0,1)$ is denoted by Q.

The joint probability distribution of (X,Y) is noted $P_{X,Y}$ with pmf or pdf $p_{X,Y}$. When X and Y are independent, $P_{X,Y} = P_X P_Y$, that is, $p_{X,Y}(x,y) = p_X(x) p_Y(y)$. The conditional probability distribution of Y given X is denoted $P_{Y|X}$ where $p_{Y|X}(y|x) = \frac{p_{X,Y}(x,y)}{p_X(x)}$.

Finally, the positive (resp. negative) part of x is $x^+ \triangleq \max(0,x)$ (resp. $x^- \triangleq \max(-x,0)$), and the complementary of $x \in [0,1]$ is $\overline{x} \triangleq 1 - x$.

2.2 Side-Channels

A *random transformation* with input X and output Y is defined by a transitional probability distribution $P_{Y|X}$, also known as a *Markov kernel* [PW23]. For example, when X is discrete, one has $p_Y(y) = \sum_x p_X(x) p_{Y|X}(y|x)$. When X is continuous, one has $p_Y(y) = \int p_X(x) p_{Y|X}(y|x)\, \mathrm{d}x$. This random transformation is noted $X \rightarrow \boxed{P_{Y|X}} \rightarrow Y$ or $X \rightarrow Y$ for short.

In the sequel, a *side-channel* is defined as a random transformation $X \rightarrow Y$, and we shall refer to any transformation $X \rightarrow Y$ as a "channel". In the side-channel literature, it is also commonly defined as a random function $Y = F(X)$ where $F = f$ is picked at random among a set of deterministic functions f according to some probability distribution P_F. It is true, but not obvious, that the two descriptions coincide. We say that the channel $X \rightarrow \boxed{P_{Y|X}} \rightarrow Y$ is *opaque* if Y is independent of its input X, that is, $p_{Y|X}(y|x) = p_Y(y)$ for all x and y.

Notice that any deterministic function $Y = f(X)$ can be seen as a "random" transformation where $p_{Y|X}(y|x) = \delta(y = f(x))$ (Dirac distribution). This functional channel will be denoted by $X \rightarrow \boxed{f} \rightarrow Y$. A functional channel with constant f is *opaque*. If f is the identity, the corresponding functional channel is named *identity channel*. When we write $X \rightarrow Y \rightarrow Z$ we always assume that it forms a Markov chain.

Additive masking of order $d \geqslant 0$ can be seen as a channel $X \rightarrow \boxed{\mathrm{Mask}_d} \rightarrow \mathbf{X}$, where $\mathbf{X} = (X_0, \ldots, X_d) \triangleq (R_0, \ldots, R_{d-1}, X - \sum_{i=0}^{d-1} R_i)$ where the R_i are independent and uniformly distributed $R_i \sim \mathcal{U}(\mathcal{X})$. The $d+1$ components of $\mathbf{X}$ are called *shares* of X. By the well-known *secret sharing property*, any subset of at most d shares of X is independent of X.

An important class of channels is as follows:

Definition 1 (Erasure Channel). *The channel*

$$X \rightarrow \boxed{\mathrm{EC}_{\mathcal{E}}^{\perp}} \rightarrow Y \tag{1}$$

is said to be an erasure channel with erasure probability $\mathcal{E} \in [0,1]$ and special erasure symbol $\perp$ if on input x, $\mathrm{EC}_{\mathcal{E}}^{\perp}$ outputs x with probability

$$\overline{\mathcal{E}} = 1 - \mathcal{E} \tag{2}$$

and the special erasure symbol $\perp$ otherwise (with probability $\mathcal{E}$). That is

$$\begin{cases} p_{Y|X}(\perp|x) = \mathcal{E} \\ p_{Y|X}(x|x) = \overline{\mathcal{E}} \end{cases} \qquad (\forall x \neq \perp) \tag{3}$$

When convenient, we also consider $\perp$ as input value, and let $p_{Y|X}(\perp|\perp) = 1$. Notice that an erasure channel with erasure probability $\mathcal{E} = 1$ is opaque.

Remark 1. The notation $\mathcal{E}$ for the erasure probability is classical in information theory. A few articles such as [DDF14, PGMP19] use the complementary $1-\mathcal{E}$ instead. Here we follow the standard information theoretic convention.

Erasure channels satisfy useful properties that are easy to check:

Lemma 1 (Commutative Property). *Let $P_{Y|X}$ be any channel from $\mathcal{X}$ to $\mathcal{Y}$ and $P^{\perp}_{Y|X}$ its extension to the input $\perp$ by setting $P^{\perp}_{Y|X}(\perp|\perp)=1$. Then*

$$\left(X \to \boxed{\mathrm{EC}^{\perp}_{\mathcal{E}}} \to \boxed{P^{\perp}_{Y|X}} \to Y'\right) = \left(X \to \boxed{P_{Y|X}} \to \boxed{\mathrm{EC}^{\perp}_{\mathcal{E}}} \to Y'\right). \tag{4}$$

Note that on the left-hand side the erasure channel is defined on $\mathcal{X}$ while on the right-hand side it is defined on $\mathcal{Y}$.

Lemma 2 (Composition of Erasure Channels). *Let $\mathcal{E}_0, \mathcal{E}_1 \in [0,1]$ and set $\overline{\mathcal{E}} = \overline{\mathcal{E}}_0\overline{\mathcal{E}}_1$. Then*

$$\left(X \to \boxed{\mathrm{EC}^{\perp}_{\mathcal{E}_1}} \to \boxed{\mathrm{EC}^{\perp}_{\mathcal{E}_0}} \to Y\right) = \left(X \to \boxed{\mathrm{EC}^{\perp}_{\mathcal{E}}} \to Y\right) \tag{5}$$

Proof. The output is not erased if and only if it is not erased in both channels, hence with probability $\overline{\mathcal{E}} = \overline{\mathcal{E}}_0\overline{\mathcal{E}}_1$. □

2.3 Informational Leakage Measures

There exist many noisiness metrics in the literature that quantify how noisy a channel $X \to Y$ can be. In this Subsection we list different leakage measures used in this paper.

The correlation coefficient is widely adopted in side-channel analysis for its simplicity in e.g., the associated correlation power analysis (CPA) [BCO04].

Definition 2 (Pearson's Correlation Coefficient)

$$\rho(X;Y) \triangleq \frac{\mathbb{E}_{XY}\left[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])\right]}{\sqrt{\mathbb{E}_X\left[(X-\mathbb{E}[X])^2\right]\mathbb{E}_Y\left[(Y-\mathbb{E}[Y])^2\right]}}. \tag{6}$$

The correlation coefficient is symmetric $\rho(X;Y)=\rho(Y;X)$. Note, however, that $\rho(X;Y)=0$ does not imply that X and Y are statistically independent.

Definition 3 (Kullback-Leibler Divergence and Total Variation Distance). *Let P, Q be two probability distributions with respective pdf or pmf p, q defined over $\mathcal{X}$. The Kullback-Leibler (KL) divergence between P and Q is*

$$D_{\mathrm{KL}}(P\|Q) \triangleq \sumint_{\mathcal{X}} p \log \frac{p}{q} \tag{7}$$

and the total variation distance (TV) between P and Q is

$$D_{\mathrm{TV}}(P\|Q) = \frac{1}{2}\sumint_{\mathcal{X}} |p-q| = \frac{1}{2}\|p-q\|_1. \tag{8}$$

KL divergence is not symmetric in general $D_{\mathrm{KL}}(P\|Q) \neq D_{\mathrm{KL}}(Q\|P)$ but the total variation is symmetric $D_{\mathrm{TV}}(P\|Q) = D_{\mathrm{TV}}(Q\|P)$.

Remark 2. Total variation is known to characterize indistinguishability in the sense that no statistical test can distinguish P and Q if $D_{\mathrm{TV}}(P\|Q)$ is negligible [PW23, § 7.3]. Both D_{TV} and D_{KL} are particular instances of *f-divergences*, that satisfy a data processing inequality (see [PW23, Def. 7.1]).

Definition 4 (Mutual Information). *The* mutual information *(*MI*) is the KL divergence between the joint distribution of* (X, Y) *and the product of its marginals:*

$$I(X;Y) \triangleq D_{\mathrm{KL}}(p_{XY}\|p_X p_Y) = \sum\!\!\!\!\!\!\int_{\mathcal{X}\times\mathcal{Y}} p_{XY}(x,y) \log \frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}. \tag{9}$$

MI is symmetric $I(X;Y) = I(Y;X)$. It is a measure of statistical dependence: if X and Y are statistically independent then $p_{XY} = p_X p_Y$ so that $I(X;Y) = 0$.

Definition 5 (Total Variation Information). *The* total variation information *(*TVI*) is the TV distance between the joint distribution of* (X, Y) *and the product of its marginals:*

$$\Delta(X;Y) \triangleq D_{\mathrm{TV}}(p_{XY}\|p_X p_Y) = \frac{1}{2}\|p_{XY} - p_X p_Y\|_1 \tag{10}$$

$$= \frac{1}{2}\sum\!\!\!\!\!\!\int_{\mathcal{X}\times\mathcal{Y}} |p_{XY}(x,y) - p_X(x)p_Y(y)|. \tag{11}$$

Note that TVI is symmetric, $\Delta(X;Y) = \Delta(Y;X)$. A negligible TVI implies that no test can distinguish p_{XY} from $p_X p_Y$, that is no test can exhibit a statistical dependence between X and Y. TVI can be seen as a particular f-information [PW23, Eq. 7.46]. In [PGMP19] TV is referred to as Statistical Distance (SD).

Definition 6 (Maximal Leakage [IWK20]). *The* maximal leakage *quantifies the maximal advantage in estimating X from the side-channel information Y:*

$$\mathcal{L}(X \to Y) = \log \sum\!\!\!\!\!\!\int_{\mathcal{Y}} \sup_{x\in\mathcal{X}} p_{Y|X}(y|x). \tag{12}$$

We use an arrow instead of a semicolon in the definition of $\mathcal{L}(X \to Y)$ because it depends only on the channel $X \to Y$ and not on the input probability distribution of X. Note that maximal leakage is not symmetric $\mathcal{L}(X \to Y) \neq \mathcal{L}(Y \to X)$.

Definition 7 (Euclidean Bias [PR13]). *The* Euclidean norm bias (EN) *is the expected Euclidean distance between the posterior distribution $p_{X|Y}$ and its prior p_X:*

$$\beta(X;Y) \triangleq \mathbb{E}_Y\|p_{X|Y}(\cdot|Y) - p_X\|_2. \tag{13}$$

Remark 3. $\beta(X;Y)$ is similar to $\Delta(X;Y)$ where $\|\cdot\|_1$ is replaced by $\|\cdot\|_2$ inside the expectation. However, $\beta(X;Y)$ is not equal to $\|p_{XY}-p_X p_Y\|_2$ because of the square root appearing in the Euclidean norm. In particular it is not symmetric $\beta(X;Y)\neq\beta(Y;X)$. A similar quantity $\Delta_2(X;Y)$ with squared norm, related to the Rényi 2-information, is used for side-channel leakage evaluation in [LBC+23].

Definition 8 ((Average) Relative Error [PGMP19]**).**

$$\mathrm{RE}(X;Y)\triangleq\sup_{x,y}\left|\frac{p_{X|Y}(x|y)}{p_X(x)}-1\right|=\sup_{x,y}\left|\frac{p_{XY}(x,y)}{p_X(x)p_Y(y)}-1\right|. \tag{14}$$

$$\mathrm{ARE}(X;Y)\triangleq\mathbb{E}_Y\left[\sup_x\left|\frac{p_{X|Y}(x|Y)}{p_X(x)}-1\right|\right]. \tag{15}$$

Remark 4. While relative error is symmetric $\mathrm{RE}(X;Y)=\mathrm{RE}(Y;X)$ the average relative error is not symmetric $\mathrm{ARE}(X;Y)\neq\mathrm{ARE}(Y;X)$.

This paper focuses on another important quantity:

Definition 9 (Complementary Doeblin Coefficient [Doe37, Dob56], [MS23a]**).**

$$\overline{\mathcal{E}}(X\to Y)=1-\sum\!\!\!\!\!\!\int_y\inf_x p_{Y|X}(y|x)=\sum\!\!\!\!\!\!\int_y\sup_x\big(p_Y(y)-p_{Y|X}(y|x)\big) \tag{16}$$

$$=\mathbb{E}_Y\left[\sup_x\left(1-\frac{p_{Y|X}(Y|x)}{p_Y(Y)}\right)\right]=\mathbb{E}_Y\left[\sup_x\left(1-\frac{p_{X|Y}(x|Y)}{p_X(x)}\right)\right]. \tag{17}$$

Remark 5. Doeblin's coefficient appeared implicitly in [Doe37, p. 1], later explicitly in [Sen73, Eq. 2.6] and has been recently known as the "Doeblin ergodicity coefficient", see e.g., [CL10, Eq. 10].

Remark 6. While the expression of CDC resembles both that of maximal leakage and ARE, it is fundamentally different. CDC is non-symmetric $\overline{\mathcal{E}}(X\to Y)\neq\overline{\mathcal{E}}(Y\to X)$. Like for maximal leakage, we use an arrow instead of a semicolon in $\overline{\mathcal{E}}(X\to Y)$ because it depends only on $X\to Y$ and not on P_X. The original Doeblin coefficient is $\mathcal{E}(X\to Y)=1-\overline{\mathcal{E}}(X\to Y)$.

Remark 7. Maximal leakage is a particular Sibson's α-information [Ver15, EVG22] of order $\alpha=+\infty$: $\mathcal{L}(X\to Y)=I_\infty(X;Y)$, while Mutual information can be seen as Sibson's α-information of order $\alpha=1$: $I(X;Y)=I_1(X;Y)$. The Doeblin coefficient is the exponential of minus Sibson's α-information of order $\alpha=-\infty$ also known as maximal cost leakage: $\mathcal{E}(X\to Y)=\exp(-I_{-\infty}(X;Y))$ [EVG22].

2.4 Properties of the Complementary Doeblin Coefficient

In Lemma 2 we have seen the composition property of erasure channels sharing the same erasure symbol. What happens now if we compose two erasure channels with different erasure symbols? The following Lemma shows that even though the resulting channel is not an erasure channel, its CDC is identical:

Lemma 3 (Erasures Composition). *For the channel* $X \to \boxed{\mathrm{EC}_{\mathcal{E}_1}^{\perp_1}} \to Y \to \boxed{\mathrm{EC}_{\mathcal{E}_0}^{\perp_0}} \to Z$,

$$\overline{\mathcal{E}}(X \to Z) = \overline{\mathcal{E}}(X \to Y)\overline{\mathcal{E}}(Y \to Z). \tag{18}$$

Consider a channel $X \to Y$ and suppose one has a post processing $Y \to Z$ (such that $X \to Y \to Z$ is a *Markov chain* [PW23]). Intuitively, Z does not contain more information than Y about X, and we have the following:

Lemma 4 (CDC Consistency). *For any* $X \to Y \to Z$, *one has*

$$\mathcal{E}(X \to (Y, Z)) = \mathcal{E}(X \to Y). \tag{19}$$

A sensitive variable X may leak several times in a side-channel attack. For instance, the adversary may access two side-channels $X \to Y_1$ and $X \to Y_2$. What can be said about the CDC of the combined leakages? The following Lemma provides an answer.

Lemma 5 (Single Letterization). *For the multi-channel* $X \begin{array}{l} \nearrow \boxed{P_{Y_1|X}} \to Y_1 \\ \searrow \boxed{P_{Y_2|X}} \to Y_2 \end{array}$ *denoted by* $X \to Y_1, Y_2$, *we have*

$$\mathcal{E}(X \to Y_1, Y_2) \geqslant \mathcal{E}(X \to Y_1)\mathcal{E}(X \to Y_2). \tag{20}$$

Generally with q channels in parallel i.e. $P_{Y_1,\ldots,Y_q|X} = \prod_{i=1}^{q} P_{Y_i|X}$ *we have*

$$\mathcal{E}(X \to Y_1, \cdots, Y_q) \geqslant \prod_{i=1}^{q} \mathcal{E}(X \to Y_i). \tag{21}$$

In terms of CDC, *Eq.* (21) *reformulates as*

$$\overline{\mathcal{E}}(X \to Y_1, \ldots, Y_q) \leqslant 1 - \prod_{i=1}^{q} \left(1 - \overline{\mathcal{E}}(X \to Y_i)\right) \leqslant \sum_{i=1}^{q} \overline{\mathcal{E}}(X \to Y_i). \tag{22}$$

In the adaptive setting, the adversary may observe Y_1 through the side-channel $X \to Y_1$ and then chose the channel $X \to Y_2$ based on his observation of Y_1. The following adaptive single letterization lemma extends Lemma 5 by showing how the CDC of the combined leakages can be derived even when the channels are chosen adaptively:

Lemma 6 (Adaptive Single Letterization). *In the adaptive setting where all channels satisfy $\overline{\mathcal{E}}(X \to Y_i) \leqslant \overline{\mathcal{E}}$, we still have*

$$\mathcal{E}(X \to Y_1, Y_2) \geqslant \mathcal{E}^2. \tag{23}$$

More generally we have $\mathcal{E}(X \to Y_1, \dots, Y_q) \geqslant \mathcal{E}^q$.

2.5 Side-Channel Attack Models

We use the following terminology from [PR13, PGMP19, MS23b].

Definition 10 (δ-Noisy Channel). *A channel $X \to \boxed{P_{Y|X}} \to Y$ is said to be* δ-noisy for input X with respect to some metric $\mathcal{D}$ *if $\mathcal{D}(X;Y) \leqslant \delta$. For short, it is said to be* δ-noisy *with respect to $\mathcal{D}$ (without reference to X) when X is taken uniformly distributed $X \sim \mathcal{U}(\mathcal{X})$. $\mathcal{D}$ should be understood as a distortion measure of the channel. For instance $\mathcal{D}$ can be $\rho, I, \Delta, \mathcal{L}, \beta, \mathrm{RE}, \mathrm{ARE}$ or $\overline{\mathcal{E}}$. The lower δ, the noisier the channel.*

Definition 11 (Side-Channel Exploitability). *Consider a set of l sensitive values $(X_1, \dots, X_l)$. A side-channel adversary obtains multiple side information $(Y_1, \dots, Y_l)$ through the channels $\varphi_i = (X_i \to Y_i)$, $i = 1, \dots, l$. The tuple of channels $\varphi = (\varphi_1, \dots, \varphi_l)$ is restricted so that the adversary's ability is limited. Typically, the adversary is said to be:*

- ***t-threshold probing*** *[ISW03]: if φ contains at most t identity channels and opaque channels on the remaining positions;*
- ***$\overline{\mathcal{E}}$-random probing*** *[DDF14]: if φ is made of $\mathcal{E}$-erasure channels;*
- ***δ-noisy*** *[PR13, PGMP19]: if φ contains only δ-noisy channels with respect to some metric $\mathcal{D}$;*
- ***(σ, f)-additive****: if φ is made of channels of the form $X \to Y \triangleq f(X) + \sigma N$ where f is a fixed deterministic leakage function and $N \sim \mathcal{N}(0,1)$ is an independent additive Gaussian noise. Typically, f can be the Hamming weight function or the least significant bit function. When X is a bit leaking as $X \to Y = X + \sigma N$ it specializes to the leakage model of Chari et al. [CJRR99].*

A cryptographic implementation is classically modeled as a circuit Γ. There are two main paradigms about the channel inputs that are both legitimate. Either it is assumed that every wire within the circuit leaks like [DDF14]. Or it is assumed that every gate within the circuit leaks like [PR13]. In this case, the channel takes as input the operands of the gate. For unary gates both models are exactly equivalent. The models differ whenever the gates process multiple operands. Assuming that the wires leak leads to tighter security bounds, while assuming that the gates leak seems to be closer to the physical nature of leakages. [DDF14, § 5.5] discusses more in depth the trade off between both models.

One concern in side-channel analysis is to cover adaptive adversaries $\mathcal{A}$. This term can be confusing as it is used with different meanings. To avoid any ambiguity, we make more precise the terminology with the following definitions:

Definition 12 (Adaptive Adversary Flavors). *We clarify the different notations of adaptivity in the context of side-channel attacks:*

1. *When $\mathcal{A}$ is allowed to choose sequentially the public information [MS23b] used by Γ then she/he is a* ***chosen public information adversary****. It corresponds to the usual setting of chosen plaintext or ciphertext adversary in cryptology.*
2. *When $\mathcal{A}$ is allowed to specify φ sequentially [DDF14] she/he is said to be a* ***chosen channel adversary****. This differs from chosen public information adversary; in this setting the adversary is allowed to move the position of the side-channel acquisition instruments (probes) from one query to the other.*
3. *If $\mathcal{A}$ can specify $\varphi_1, \dots, \varphi_l$ sequentially within a query [DDF14], she/he is said to be a* ***strong chosen channel adversary****. The adversary is even allowed to move the position of its probe within a query. This last type of adaptivity is, however, unrealistic in most of practical settings.*

The activity of a side-channel adversary $\mathcal{A}$ with q queries can be viewed as a game. This game unfolds differently depending on the side-channel adaptivity and depending on the gate/wire leakage model. After side-channel collection, the adversary exploits them to distinguish the correct key K and outputs $\mathrm{out}_{\mathcal{A}}(K)$.

The complete acquisition and attack led by $\mathcal{A}$ is formalized by Algorithm 1. In practice $\mathrm{out}_{\mathcal{A}}(K)$ can be a score vector sorting the key hypotheses (or parts of the key). If $\mathcal{A}$ is restricted to opaque channels she/he does not learn anything through them. In this case, $\mathcal{A}$ is said to be a black-box adversary.

Algorithm 1: Side-Channel Acquisition and Attack

Data: A number of queries q and a set of allowed side-channels.
Result: The output of the adversary $\mathcal{A}$.
Oracle $\mathcal{O}$ draws uniformly at random a secret key K.
for $i = 1, \dots, q$ **do** `/* Sequential Acquisition of` q `Traces. */`
  $\mathcal{A}$ specifies a public information $\mathbf{t}_i$ and send it to $\mathcal{O}$. `/* Sequential Choice of Public Information */`
  $\mathcal{O}$ draws uniformly at random the randomness $\mathbf{R}_i$ and computes the corresponding wire values $\mathbf{X}_i$.
  `/* This is the wire leaking model. In the gate leakage model the loop is over the gates instead. */`
  for $j = 1, \dots, l$ **do** `/* Sequential Choice of the Side-Channels */`
    $\mathcal{A}$ specifies φ_i to $\mathcal{O}$
    $\mathcal{O}$ sends back the corresponding leakage x_i from side-channel φ_i to $\mathcal{A}$ under the constraint that φ is an allowed tuple of channel.
  `/* Restriction on the type of allowed side-channels */`
return $\mathcal{A}$ outputs $\mathrm{out}_{\mathcal{A}}(K)$

2.6 Figures of Merit

When $\text{out}_{\mathcal{A}}(K)$ is a key ranking the performances of the adversary are measured via three classical figures of merits: The *success rate* (SR) $\mathbb{P}_s$, the *success rate of order o* (SR_o, success rate in o-trials) $\mathbb{P}_{s,o}$ [SMY09] and the *guessing entropy* GE [Mas94]. We follow Ito *et al.* [IUH22, § 2.3] and express these metrics in terms of the a posteriori rank of the key hypothesis given the side-information.

Definition 13 (Success Rate (SR) and Guessing Entropy (GE)). *Let K be a secret random variable taking values in a finite set $\mathcal{K}$. Let Y be an arbitrary random variable representing a side-information. The* success rate *(*SR*) is given by the Maximum a Posteriori (MAP) rule*

$$\mathbb{P}_s(K|Y) \triangleq \mathbb{P}(\text{rank}(K|Y) = 1) \tag{24}$$

The success rate of order *o,* SR_o *is*

$$\mathbb{P}_{s,o}(K|Y) \triangleq \mathbb{P}(\text{rank}(K|Y) \leqslant o) \tag{25}$$

The guessing entropy *(*GE*) is the minimum average number of guesses of an optimal guessing strategy*

$$G(K|Y) \triangleq \mathbb{E}\{\text{rank}(K|Y)\} \tag{26}$$

In general rank *is defined as a function such that for each y, $k \to \text{rank}(k|y)$ is a permutation of the key space $\mathcal{K}$. In an optimal guessing strategy, for each $y \in \mathcal{Y}$,* $\text{rank}(k|y) \in \{1, \dots, |\mathcal{K}|\}$ *is the rank of $p_{K|Y}(k|y)$ in the list $\{p_{K|Y}(k|y) | k \in \mathcal{K}\}$ sorted in decreasing order. In case of collisions, ties are resolved randomly which does not change the statistical quantities at stake.*

Remark 8. Since $\mathbb{P}_s(K|Y) = \mathbb{P}_{s,o=1}(K|Y)$ holds, results for SR will be derived from results in terms of SR_o.

Definition 14 (Blind Guess). *When no side-information is available, the adversary performs a* blind guess *whose figures of merits are constants depending only on the a priori key-distribution. Namely*

$$\mathbb{P}_{s,o}(K) \triangleq \mathbb{P}(\text{rank}(K) \leqslant o) \qquad \textit{and} \qquad G(K) \triangleq \mathbb{E}\{\text{rank}(K)\} \tag{27}$$

where $\text{rank}(k) \in \{1, \dots, |\mathcal{K}|\}$ *is the rank of $p_K(k)$ in the list $\{p_K(k) | k \in \mathcal{K}\}$ sorted in decreasing order.*

The *advantage* of the adversary is quantified by $\mathbb{P}_{s,o}(K|Y) - \mathbb{P}_{s,o}(K) \geqslant 0$, $G(K) - G(K|Y) \geqslant 0$ and for statistical tests $\Delta(K;Y) \geqslant 0$. If further K is uniformly distributed then $\mathbb{P}_{s,o}(K) = \frac{o}{|\mathcal{K}|}$ and $G(K) = \frac{|\mathcal{K}|+1}{2}$.

If an adversary $\mathcal{A}$ is computationally bounded then she/he may not be able to fully exploit the side-information Y. The corresponding figures of merit are denoted by $\mathbb{P}_s^{\mathcal{A}}(K|Y)$, $\mathbb{P}_{s,o}^{\mathcal{A}}(K|Y)$ and $G^{\mathcal{A}}(K|Y)$. Obviously $\mathbb{P}_s^{\mathcal{A}}(K|Y) \leqslant \mathbb{P}_s(K|Y)$, $\mathbb{P}_{s,o}^{\mathcal{A}}(K|Y) \leqslant \mathbb{P}_{s,o}(K|Y)$ and $G^{\mathcal{A}}(K|Y) \geqslant G(K|Y)$.

3 Mathematical Key Properties of the CDC

In this section we derive the key mathematical properties of the CDC that will be useful to derive security bounds. In Subsect. 3.1 we exhibit the optimal factorization of a side-channel into a stochastically degraded erasure channel. This shows that CDC is the optimal parameter in the reduction from noisy leakages to the random probing model. The word optimal refer to the reduction from noisy leakage to a random probing adversary, we do not claim however that the CDC yields an optimal bound on the success rate of a side-channel attack. In Subsect. 3.2 we show how the figures of merit of a side-channel attack can be bounded using CDC leveraging the Data Processing Inequality (DPI). We show that CDC is amenable to evaluation in Subsect. 3.3. Finally, we compare the CDC to the informational measure (introduced in Subsect. 2.3) in Subsect. 3.4.

3.1 Optimal Channel Degradation

It is known that security in the noisy leakage model can be reduced to security in the random probing model. In [DDF14, Lemma 3], security in the noisy leakage model measured by TVI is reduced to security in the random probing model. In [PGMP19, Lemma 3], security in the noisy leakage model measured by ARE is reduced to security in the random probing model. Finally, [DFS19, Theorem 3] proves security in the noisy leakage model measured by MI by upper bounding [DDF14, Lemma 3] using Pinsker's inequality [PW23, Thm 7.10].

The key property is that any channel can be seen as a stochastically degraded erasure channel. This stochastic degradation can be seen as a factorization like [DDF14, Lemma 4] of $\mathrm{Noise}(X)$ into $\mathrm{Noise}'(\varphi(X))$ where φ is an erasure channel (termed ϵ-identity function in their presentation). In this section, we derive the optimal parameter in this reduction and show it corresponds to the complementary Doeblin coefficient (CDC). This unifies previous results that can seen as a weakened version of our reduction by upper bounding the CDC.

Definition 15 (Degraded Channel). *The channel $X \to Z$ is said to be* stochastically degraded *with respect to the channel $X \to Y$ if there exists a channel $Y \to Z$ such that*

$$\Big(X \to \boxed{P_{Y|X}} \to Y \to \boxed{P_{Z|Y}} \to Z\Big) = \Big(X \to \boxed{P_{Z|X}} \to Z\Big). \tag{28}$$

Theorem 1. *Any channel $X \to \boxed{P_{Y|X}} \to Y$ is a stochastically degraded erasure channel:*

$$X \to \boxed{\mathrm{EC}_{\mathcal{E}}^{\perp}} \to X' \to \boxed{P_{Y|X'}} \to Y \tag{29}$$

with the maximum *erasure probability*

$$\mathcal{E}(X \to Y) = \sumint_y \inf_{x \in \mathcal{X}} p_{Y|X}(y|x). \tag{30}$$

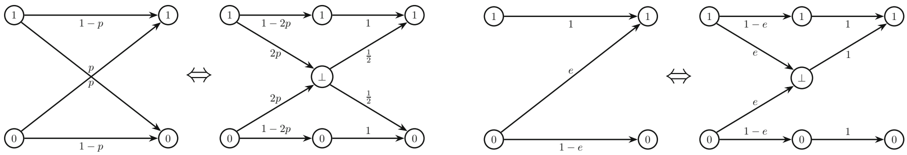
Remark 9. $\mathcal{E}(X \to Y)$ is known in the literature as the *Doeblin coefficient of ergodicity* of the channel [Doe37, Dob56, Mak20, MS23a]. In our context $\mathcal{E}(X \to Y)$ represents the erasure probability while $\overline{\mathcal{E}}(X \to Y)$ represents the probing probability. Thm 1 was proved for binary input channels in [BB11, Prop. 6.4] in the context of physical layer security and wiretap channels and in the general for network coding in [Mak20, Lemma 6] or key agreements in [GGK20, Lemma 5]. The CDC appears for the first time in the side-channel literature as the value of ϵ in [DDF19, Eq. 9, Proof of Lemma 4].

Remark 10. Maximum erasure probability in Theorem 1 means that there exists at least one channel degradation (in the form of Eq. (29)) achieving $\mathcal{E} = \mathcal{E}(X \to Y)$ and that there does not exist any channel degradation with $\mathcal{E} > \mathcal{E}(X \to Y)$. In this sense, the CDC is the optimal parameter in the reduction from noisy leakage to the random probing model.

Obviously, the erasure channel is optimally degraded into itself, that is, $\mathcal{E}(X \to \boxed{\mathrm{EC}^{\perp}_{\mathcal{E}}(X)} \to Y) = \mathcal{E}$.

Example 1 (Binary Symetric Channel (BSC)). If X is a binary random variable and $X \to Y$ a BSC with crossover probability $0 \leqslant p \leqslant \frac{1}{2}$ then $\mathcal{E}(X \to Y) = 2p$. The factorization given by Theorem 1 is shown in Fig. 2a.

Example 2 (Z-Channel). If X is binary and $X \to Y$ is a Z-channel with parameter $0 \leqslant e \leqslant 1$ then $\mathcal{E}(X \to Y) = e$. The factorization given by Theorem 1 is shown in Fig. 2b.



(a) BSC (b) Z-Channel

Fig. 2. Illustration of Theorem 1

Theorem 1 implies the following corollary in terms of simulatability similar to [DDF14, Lemma 3] in terms of CDC:

Corollary 1. *Any $\overline{\mathcal{E}}$-noisy adversary $\mathcal{A}$ with respect to* CDC *can be perfectly simulated by a $\overline{\mathcal{E}}$-random probing adversary $\mathcal{S}$.*

Strengthened Data Processing Inequality. By Lemma 2, the composition of several erasure channels is an erasure channel with a larger erasure probability. In general, the composition of several channels should leak less information. This is formalized by a strengthened data processing inequality (DPI):

Lemma 7 (Strengthened Data Processing Inequality (DPI)). *For any* $X \to Y \to Z$, CDC *satisfies the following strengthened-DPI*

$$\overline{\mathcal{E}}(X \to Z) \leqslant \overline{\mathcal{E}}(X \to Y)\overline{\mathcal{E}}(Y \to Z) \tag{31}$$

which implies a preprocessing-DPI (preprocessing $X \to Y$ *can only reduce leakage):* $\overline{\mathcal{E}}(X \to Z) \leqslant \overline{\mathcal{E}}(Y \to Z)$*, and a post-processing-DPI (post-processing* $Y \to Z$ *can only reduce leakage)* $\overline{\mathcal{E}}(X \to Z) \leqslant \overline{\mathcal{E}}(X \to Y)$*. In a nutshell, stochastic degradation reduces the value of the* CDC*.*

Proof. By Theorem 1 we can degrade the channels $X \to Y$ and $Y \to Z$ so that the channel $X \to Z$ rewrites as

$$X \to \boxed{\mathrm{EC}^{\perp_x}_{\mathcal{E}(X\to Y)}} \to X' \to \boxed{P_{Y|X'}} \to Y \to \boxed{\mathrm{EC}^{\perp_y}_{\mathcal{E}(Y\to Z)}} \to Y' \to \boxed{P_{Z|Y'}} \to Z. \tag{32}$$

Since by Lemma 1 an erasure channel commutes with any other channel, the channel is equivalent to

$$X \to \boxed{\mathrm{EC}^{\perp_x}_{\mathcal{E}(X\to Y)}} \to X' \to \boxed{\mathrm{EC}^{\perp_y}_{\mathcal{E}(Y\to Z)}} \to X'' \to \boxed{P^{\perp_y}_{Y'|X''}} \to Y \to \boxed{P_{Z|Y'}} \to Z. \tag{33}$$

We have two different erasures symbols but by Lemma 3, the concatenated channel is stochastically degraded with respect to an erasure channel with $\overline{\mathcal{E}} = \overline{\mathcal{E}}(X \to Y)\overline{\mathcal{E}}(Y \to Z)$. So we have

$$X \to \boxed{\mathrm{EC}^{\perp}_{\mathcal{E}}} \to \tilde{X} \to \boxed{P_{\tilde{X}|X''}} \to X'' \to \boxed{P^{\perp_y}_{Y'|X''}} \to Y \to \boxed{P_{Z|Y'}} \to Z. \tag{34}$$

Now let $p_{Z|\tilde{X}} = p_{Z|Y'} \to p^{\perp_y}_{Y'|X''} \to p_{\tilde{X}|X''}$ so that

$$X \to \boxed{\mathrm{EC}^{\perp}_{\mathcal{E}}} \to \tilde{X} \to \boxed{P_{Z|\tilde{X}}} \to Z. \tag{35}$$

Since $\overline{\mathcal{E}}(X \to Z)$ is the infimum such that this factorization holds, we have $\overline{\mathcal{E}}(X \to Z) \leqslant \overline{\mathcal{E}} = \overline{\mathcal{E}}(X \to Y)\overline{\mathcal{E}}(Y \to Z)$ which concludes the proof. □

Remark 11. Makur and Singh [MS23a], with a very different proof, established a similar property of CDC in the discrete setting and interpreted it as the submultiplicativity of CDC.

3.2 Bounds on the Figures of Merit

An adversary tries to recover the sensitive variable X with the help of the side-information Z through the side-channel $X \to Z$ which is stochastically degraded with respect to the channel $X \to Y$. Intuitively she/he can only perform worse than an adversary that accesses Y. This intuition is formalized by a DPI data processing inequality (DPI).

Lemma 8 (Data Processing Inequality (DPI)). *Consider the channel $U \to V \to W \to X$. Then*

$$I(V;W) \geqslant I(U;X) \quad \textit{and} \qquad \Delta(V;W) \geqslant \Delta(U;X). \tag{36}$$

Consider the channel $X \to Y \to Z$, where X is valued in the finite set $\mathcal{X}$, the SR_o *and* GE *verify a post-processing DPI,*

$$\mathbb{P}_{s,o}(X|Y) \geqslant \mathbb{P}_{s,o}(X|Z) \qquad \textit{and} \qquad G(X|Y) \leqslant G(X|Z). \tag{37}$$

Proposition 1. *Let $\lambda_{\mathrm{SR}_o} = (1 - \mathbb{P}_{s,o}(K)), \lambda_{\mathrm{GE}} = (G(K) - 1), \lambda_{\mathrm{TVI}} = (1 - \exp(-H_2(K)))$ be three constants that only depend on the a priori secret key distribution (where H_2 is the collision entropy). The adversary's advantage for* SR, GE *and* TVI *can be bounded as follows:*

$$\begin{array}{lcl} 0 \leqslant \mathbb{P}_{s,o}(K|Y) - \mathbb{P}_{s,o}(K) & \leqslant & \overline{\mathcal{E}}(K \to Y)\lambda_{\mathrm{SR}_o}, \\ 0 \leqslant \quad G(K) - G(K|Y) & \leqslant & \overline{\mathcal{E}}(K \to Y)\lambda_{\mathrm{GE}}, \\ 0 \leqslant \qquad\quad \Delta(K;Y) & \leqslant & \overline{\mathcal{E}}(K \to Y)\lambda_{\mathrm{TVI}}. \end{array} \tag{38}$$

We cannot directly deduce from Lemma 8 that $\mathbb{P}_s^{\mathcal{A}}(K|Y)$, $\mathbb{P}_{s,o}^{\mathcal{A}}(K|Y)$ and $G^{\mathcal{A}}(K|Y)$ verify a DPI. But as shown by Duc *et al.* [DDF14, Lemma 2] if the channel $K' \to Y$ can be efficiently sampled then $\mathcal{A}$ can efficiently reproduce $\tilde{Y}$ equal in distribution with Y from X'. As a consequence, under this hypothesis we can assume that $\mathbb{P}_s^{\mathcal{A}}(K|Y)$, $\mathbb{P}_{s,o}^{\mathcal{A}}(K|Y)$ and $G^{\mathcal{A}}(K|Y)$ also verify the bounds from Proposition 1. As shown by Brian *et al.* [BFO+22], a large class of noisy channels $K' \to Y$ can be indeed simulated almost for free. In fact, we do not need to have an efficient simulation of the channel noise $K' \to Y$. Indeed, since $K \to K'$ is an erasure channel we obtain

$$\mathbb{P}_s^{\mathcal{A}}(K|Y) \overset{(a)}{\leqslant} \mathbb{P}_s(K|Y) \overset{(b)}{\leqslant} \mathbb{P}_s(K|K') \overset{(c)}{=} \mathbb{P}_s^{\mathcal{A}}(K|K') \tag{39}$$

where (a) holds because a computationally bounded adversary can only perform worse than the optimal unbounded adversary, (b) is the usual DPI and (c) is due to the fact that for an erasure channel an optimal attack is efficiently computable. For example, the attack that outputs the key when it is not erased and a random ranking otherwise is both optimal and efficient. The same derivation holds for GE (with reversed inequalities).

Discussion on the Optimality of the CDC. The trade-off between a CDC-based bound and a MI based bound depends on the nature of the channel $K \to Y$ which is factorized optimally with the CDC into $K \to K' \to Y$.

- If $K \to Y$ is an erasure channel then the factorization is $K \to K' = Y$ so that using the DPI to consider K' instead of Y as a leakage can be done without any degradation of the final bound. In this case the bound using CDC is optimal and holds with equality.
- If the channel $K \to Y$ is far from an erasure channel then the channel $K' \to Y$ can be noisy. As a consequence, using the DPI to consider K' as the leakage instead of Y incurs an unavoidable loss. In this case the CDC based bound can be loose and another informational leakage measure such as MI or TVI may be more suitable to capture the noise in the side-channel. A very bad channel in this respect is the channel from $K \to Y$ where K and Y are both taking their values in $\{1, \ldots, |\mathcal{X}|\}$, $p_{Y|K}(y|k) = (|\mathcal{X}| - 1)^{-1}$ if $y \neq k$ and $p_{Y|K}(y|k) = 0$ otherwise. In this case, $\overline{\mathcal{E}}(K \to Y) = 1$ while $I(K;Y) = \log(|\mathcal{X}|/(|\mathcal{X}| - 1))$ is small.

Both of these two extreme cases are toy examples that do not occur in practice. Depending on the nature of the practical side-channel the bound based on CDC will be tight or not.

3.3 Theoretical Expressions for Concrete Evaluation

In this Subsection we show how CDC can be evaluated in a practical setting. We first show how we can derive a closed form expression for univariate functional channels perturbed by an additive Gaussian noise. (This corresponds to (σ, f)-additive adversaries.) Then we show how this allows to bound CDC in the multivariate case perturbed by an additive Gaussian noise with a given correlation matrix $\mathbf{\Sigma}$. This corresponds to the widely used setting from *template attack* (TA) [CRR02, LRP07, UKM+17]

We show that in this model CDC is suitable for concrete evaluation, even when the noise is multivariate and potentially high dimensional.

Univariate Case

Definition 16 (Radially Symmetric Decreasing). *The real-valued r.v. Z is said to be radially symmetric decreasing if $p_Z(z) = p_Z(|z|)$ and decreasing in $|z|$.*

We derive a closed-form expression for CDC when the channel is functional perturbed by a radially symmetric decreasing additive noise. This includes Gaussian, Laplacian or Cauchy distribution for example. As CDC only depends on the channel the result does not depend on the probability distribution of X. As expected the CDC tends to zero as the noise increases.

Proposition 2. *Let X be a random variable taking values in $\mathcal{X}$ and $Y = f(X) + Z$ where f is an arbitrary real-valued function and Z is a radially symmetric decreasing noise with survival function S. Let $m = \inf_{x \in \mathcal{X}} f(x)$ and $M = \sup_{x \in \mathcal{X}} f(x)$. Then*

$$\mathcal{E}(X \to Y) = 2\, S\left(\frac{M - m}{2}\right). \tag{40}$$

For the widely adopted linear leakage model, $\mathcal{X} = \mathbb{F}_2^n$ *and* $f(X) = \sum_{i=1}^n a_i X_i$ *where* $\mathbf{a} = (a_1, \ldots, a_n) \in \mathbb{R}^n$ *and* X_i *denotes the* i*-th bit in the binary representation of* X*, we have* $m = -\sum_i a_i^-$ *and* $M = \sum_i a_i^+$ *so that* $M - m = \sum_i |a_i| = \|\mathbf{a}\|_1$ *and the expression simplifies to*

$$\mathcal{E}(X \to Y) = 2\, S\left(\frac{\|\mathbf{a}\|_1}{2}\right). \tag{41}$$

If $Z \sim \sigma\mathcal{N}(0,1)$*, the survival function* S *is the Marcum function* Q*, and*

$$\overline{\mathcal{E}}(X \to Y) = 1 - 2Q\left(\frac{\|\mathbf{a}\|_1}{2\sigma}\right) \overset{\sigma\to\infty}{=} \frac{\|\mathbf{a}\|_1}{\sqrt{2\pi}}\frac{1}{\sigma} + O\left(\sigma^{-3}\right). \tag{42}$$

For the classical Hamming weight (HW) model, $\mathbf{a} = (1, \ldots, 1)$ *and* $\|\mathbf{a}\|_1 = n$*.*

Remark 12. When f is constant then $m = M$, and we obtain $\mathcal{E}(X \to Y) = 1$. This is expected as in this case the channel is opaque.

Multivariate Case. Let $f : x \in \mathbb{F} \mapsto f(x) \in \mathbb{R}^m$ be a multivariate leakage function and $\mathbf{Y} = f(X) + \mathcal{N}(0, \mathbf{\Sigma})$. Then $\tilde{\mathbf{Y}} \triangleq \mathbf{W}\mathbf{Y} = (\mathbf{W} \cdot f)(X) + \tilde{\mathbf{Z}}$ where $\mathbf{W}$ is a given whitening matrix (e.g. $\mathbf{W} = \mathbf{\Sigma}^{-\frac{1}{2}}$) so that $\tilde{\mathbf{Z}} = \mathbf{W}\mathbf{Z} \sim \mathcal{N}(0, \mathbf{I}_m)$. Given X, $\tilde{\mathbf{Y}}$ is a Gaussian vector whose covariance matrix is diagonal. Hence, by theorem on Gaussian vectors, the different components of $\tilde{\mathbf{Y}}$ are independent given X and Lemma 5 implies that

$$\mathcal{E}(X \to \mathbf{Y}) = \mathcal{E}(X \to \tilde{\mathbf{Y}}) \geqslant \prod_{i=1}^{m} \mathcal{E}(X \to \tilde{Y}_i). \tag{43}$$

Since every channel $X \to \tilde{Y}_i$ is univariate additive Gaussian noise, its expression is given by Proposition 2. This methodology yields a positively biased estimator of $\mathcal{E}(X \to \mathbf{Y})$ from the non-biased estimator of each $\mathcal{E}(X \to \tilde{Y}_i)$. This approach is more conservative but ensures that we do not overestimate the security parameter which would result in a false sentiment of security. This is by opposition with the perceived information (PI [RSV+11, Eq. 3]) which is a negatively biased estimator of MI as shown in [BHM+19,IUH22].

Example 3. As an example, consider the channel $\mathbf{Y} \triangleq (f(X), f(X))^T + \mathbf{Z}$ where f is a univariate leakage function and $\mathbf{Z} \sim \mathcal{N}(0, \mathbf{\Sigma})$ with a covariance matrix $\mathbf{\Sigma} = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ where σ is the noise standard deviation and $\rho \in [-1, 1]$ is the correlation coefficient of the noise components. Let $m = \inf_x f(x)$ and $M = \sup_x f(x)$. With a little of linear algebra we observe that $\mathbf{\Sigma} = \mathbf{P}\mathbf{D}\mathbf{P}^T$ where $\mathbf{P} = \mathbf{P}^T = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ is orthonormal and $\mathbf{D} = \sigma^2 \begin{pmatrix} 1+\rho & 0 \\ 0 & 1-\rho \end{pmatrix}$ is diagonal. We compare two noise whitening techniques.

Karhunen-Loève transform (PCA Whitening): Let $\mathbf{W} = \mathbf{D}^{-\frac{1}{2}}\mathbf{P}^T$ be the noise whitening of Karhunen-Loève transform and $\mathbf{WZ} \triangleq \tilde{\mathbf{Z}} \sim \mathcal{N}(0, \mathbf{I}_2)$,

$$\tilde{\mathbf{Y}} = \mathbf{WY} = \sqrt{\frac{2}{1+\rho}}\frac{1}{\sigma}\begin{pmatrix} f(X) \\ 0 \end{pmatrix} + \tilde{\mathbf{Z}}. \tag{44}$$

By Proposition 2, $\mathcal{E}(X \to \tilde{Y}_1) = 2Q\Big(\sqrt{\frac{2}{1+\rho}}\frac{1}{\sigma}\frac{M-m}{2}\Big)$. Furthermore, $\mathcal{E}(X \to \tilde{Y}_2) = 1$ so that Eq. (43) becomes an equality here

$$\mathcal{E}(X \to \mathbf{Y}) = 2Q\Big(\sqrt{\frac{2}{1+\rho}}\frac{1}{\sigma}\frac{M-m}{2}\Big) = 1 - \frac{1}{\sigma}\sqrt{\frac{2}{1+\rho}}\frac{M-m}{\sqrt{2\pi}} + O(\sigma^{-2}). \tag{45}$$

Equation (45) is coherent:

- When $\rho = 1$, the leakage is repeated and as expected from Lemma 4 the CDC remains unchanged.
- When $\rho = 0$, we have two independent noises, and it is optimal to average the samples such that the global noise variance is halved.
- As $\rho \to -1$, $\mathcal{E}(X \to \mathbf{Y}) \to 0$ which is expected since averaging both components completely cancels out the noise and $f(X)$ is revealed.

Mahalanobis transform (ZCA Whitening): Let $\mathbf{W} = \mathbf{\Sigma}^{-\frac{1}{2}} = \mathbf{P}\mathbf{D}^{-\frac{1}{2}}\mathbf{P}^T$ be the noise whitening of Mahalanobis transform and $\mathbf{WZ} \triangleq \tilde{\mathbf{Z}} \sim \mathcal{N}(0, \mathbf{I}_2)$,

$$\tilde{\mathbf{Y}} = \mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{Y} = \sqrt{\frac{1}{1+\rho}}\frac{1}{\sigma}\begin{pmatrix} f(X) \\ f(X) \end{pmatrix} + \tilde{\mathbf{Z}}. \tag{46}$$

By Proposition 2, $\mathcal{E}(X \to \tilde{Y}_i) = 2Q\Big(\sqrt{\frac{1}{1+\rho}}\frac{1}{\sigma}\frac{M-m}{2}\Big)$. Equation (43) becomes

$$\mathcal{E}(X \to \mathbf{Y}) \geqslant 4Q^2\Big(\sqrt{\frac{1}{1+\rho}}\frac{1}{\sigma}\frac{M-m}{2}\Big). \tag{47}$$

Interestingly, with Mahalanobis transform we have an inequality in (43) which shows that the choice of the whitening technique can affect the bound tightness.

3.4 Comparison with the Other Informational Leakage Measures

Let X be a random variable taking values in a finite set $\mathcal{X}$ and a channel $X \to Y$. The following inequalities hold:

Lemma 9

$$\left.\begin{array}{r} \frac{I(X;Y)}{\log|\mathcal{X}|} \leqslant \frac{I(X;Y)}{H(X)} \\ \left.\begin{array}{r} \frac{\mathrm{ARE}(X;Y)}{2\gamma_X\lambda_{\mathrm{TVI}}} \\ \\ \frac{\beta(X;Y)}{2\lambda_{\mathrm{TVI}}} \end{array}\right\} \leqslant \frac{\Delta(X;Y)}{\lambda_{\mathrm{TVI}}} \\ \\ \frac{\exp(\mathcal{L}(X\to Y))-1}{|\mathcal{X}|-1} \end{array}\right\} \leqslant \overline{\mathcal{E}}(X\to Y) \leqslant \left\{\begin{array}{l} \mathrm{ARE}(X;Y) \leqslant \mathrm{RE}(X;Y) \\ \\ \gamma_X\beta(X;Y) \\ \gamma_X\Delta(X;Y) \leqslant \gamma_X\left(\frac{I(X;Y)}{2\log e}\right)^{\frac{1}{2}} \\ \\ (|\mathcal{X}|-1)(\exp(\mathcal{L}(X\to Y))-1) \end{array}\right. \tag{48}$$

where H is Shannon entropy, H_2 is the collision entropy, $\lambda_{\mathrm{TVI}} = 1 - \exp(-H_2(X))$ and $\gamma_X \triangleq \left(\inf_{x\in\mathcal{X}} p_X(x)\right)^{-1}$. If $X \sim \mathcal{U}(\mathcal{X})$ then $\gamma_X = |\mathcal{X}|$ and $\lambda_{\mathrm{TVI}} = 1 - \frac{1}{|\mathcal{X}|}$.

Lemma 9 does not lower bound the CDC in terms of RE because it is impossible to obtain a meaningful bound. Indeed, as remarked by Masure & Standaert [MS23b] if $X \to Y$ is an erasure channel with an arbitrarily small parameter $\mathcal{E} > 0$ then $\mathrm{RE}(X;Y) = |\mathcal{X}| - 1$. As a consequence, RE cannot provide a smooth reduction from noisy leakages to the random probing model. We compare the different leakage measures in Table 1 via three criteria:

- the ratio of their lower bound by their upper bound in Lemma 9 when $X \sim \mathcal{U}(\mathcal{X})$ (as a measure of relative looseness);
- their maximal value (which measures their normalization);
- their asymptotic values in the Hamming weight leakage model when $X \sim \mathcal{U}(\mathbb{F}_2^n)$ hence $|\mathcal{X}| = 2^n$ (which measures their performance for a typical leakage model).

Table 1. **T** is the ratio of the lower bound by the upper bound in Lemma 9 when $X \sim \mathcal{U}(\mathcal{X})$. **M** indicates the maximal value of the leakage measures. Finally, **H** indicates the asymptotic values of the leakage measure in the Hamming weight leakage model. We used the values of [PGMP19, Prop. 3] for ARE, EN and TVI, for MI we used [BCPZ16], for CDC we used Proposition 2.

	$I(X;Y)$	$\Delta(X;Y)$	$\mathcal{L}(X \to Y)$	$\beta(X;Y)$	$\mathrm{RE}(X;Y)$	$\mathrm{ARE}(X;Y)$	$\overline{\mathcal{E}}(X \to Y)$
T	$\frac{\lvert\mathcal{X}\rvert \log \lvert\mathcal{X}\rvert}{\sqrt{2\log e I(X;Y)}}$	$\lvert\mathcal{X}\rvert - 1$	$(\lvert\mathcal{X}\rvert - 1)^2$	$2(\lvert\mathcal{X}\rvert - 1)$	$+\infty$	$2(\lvert\mathcal{X}\rvert - 1)$	1
M	$\log \lvert\mathcal{X}\rvert$	$1 - \frac{1}{\lvert\mathcal{X}\rvert}$	$\log \lvert\mathcal{X}\rvert$	$\sqrt{1 - \frac{1}{\lvert\mathcal{X}\rvert}}$	$\lvert\mathcal{X}\rvert - 1$	$\lvert\mathcal{X}\rvert - 1$	1
H	$\frac{n \log e}{8} \frac{1}{\sigma^2}$	$\frac{\sqrt{n}}{2\pi\sigma}$	$\frac{n \log e}{\sqrt{2\pi}\sigma}$	$\sqrt{\frac{n}{2\pi 2^n}} \frac{1}{\sigma}$	$2^n - 1$	$\frac{n}{\sqrt{2\pi}\sigma}$	$\frac{n}{\sqrt{2\pi}\sigma}$

This allows us to label the introductive Fig. 1. Table 1 shows that CDC and ARE have the same asymptotic expression in the Hamming weight leakage model with high noise. While ARE is suboptimal and do not verify the properties of the CDC, it provides a tight reduction to the random probing model in this scenario. However, the range of ARE is $[0, |\mathcal{X}| - 1]$ and its relative looseness is $2(|\mathcal{X}| - 1)$ which indicates that in a sense ARE contains the field size in its definition. In any case, it remains preferable to use the CDC which provides the optimal reduction from noisy leakage to random probing.

4 Direct Proofs *via* CDC and Prouff-Rivain Subsequences

In this section we revisit the direct security proof in the noisy leakage model based on Prouff and Rivain's subsequence decomposition [PR13] to obtain a new derivation in terms of CDC. The subsequence decomposition of Prouff and Rivain is recalled in Subsect. 4.1. We first prove security for subsequences of type 1 and 2 in Subsect. 4.2. The security of type 3 and type 4 subsequences is obtained in Subsect. 4.3 and Subsect. 4.4 respectively. Finally, we combine the security bounds obtained for each subsequence into a security bound for the whole circuit in Subsect. 4.5 and compare it with a MI based bound in Subsect. 4.6.

4.1 Subsequence Decomposition

For typical block ciphers like the AES, featuring substitution boxes (denoted by Sboxes), Prouff and Rivain [PR13, § 4.2] decompose the computations in four different types of subsequences:

Type 1 $(z_i \leftarrow g(x_i))_i$ where g is a linear function (of the block cipher)
Type 2 $(x_i \leftarrow g(y_i))_i$ where g is an affine function (of Sbox evaluation)
Type 3 $(v_{i,j} \leftarrow a_i b_j)_{i,j}$ (First step of non-linear secure multiplication)
Type 4 $(t_{i,j} \leftarrow t_{i,j-1} + v_{i,j})_{i,j}$ (Last step of non-linear secure multiplication)

This decomposition has become standard to derive security proofs [PGMP19, MS23b]. Note that in this model it is classically assumed that the gates leak.

The first type of subsequences considers linear operations on a shared uniform variable. The second type of subsequence considers linear operations on a shared polynomial expression of a uniform variable. This is typically the case of linear operations within Sboxes. The third type of subsequences deals with the first part of the ISW multiplications involving the cross-product of the input shares of two (non-necessarily independent) random variables. Finally, the type 4 subsequences correspond to the compression layer of the ISW multiplication.

Flaws for Type 3 Subsequences in the State of the Art. In the long version, we list some flaws in the preceding direct proofs in the noisy leakage model [PR13,PGMP19,MS23b]. While these three proofs are different in nature, the flaws appear at a similar step: proving security for type 3 subsequences. For [PR13,PGMP19] it is due to an incorrect derivation of the chain rule for conditional probabilities. For [MS23b] it is due to the fact that the function used in Mrs. Gerber's Lemma is convex in one variable when the others are fixed but not jointly convex. In this section, X is a sensitive random variable taking values in $\mathcal{X} = \mathbb{F}_2^n$ that can be expressed as a function of the secret K and a public information (e.g., plaintext or ciphertext).

4.2 Security of Type 1 and Type 2 Subsequences

In [BCG+23, Coro. 1] Béguinot *et al.* leveraged Mrs. Gerber's Lemma (MGL) to derive security bounds for encodings in terms of MI. Masure and Standaert [MS23b, Coro. 4 & 5] showed how to exploit such MGL to prove the security of type 1 and type 2 subsequences. We now show that CDC also verifies a sort of MGL that quantifies the security for both type 1 ($f = \mathrm{id}$) and type 2 subsequences (generic f).

Lemma 10 (Mrs. Gerber's Lemma for CDC**, Type 1 and Type 2 Subsequences).** *Let* $\mathbf{G} = (G_i)_{i=0}^d$ *be a* d*-th order encoding of* $G = g(X)$ *where* g *is a given function. Assume that each share leaks independently through the side-channels* $(G_i \to Y_i)_{i=0}^d$*. Let* $\mathbf{Y} \triangleq (Y_0, \ldots, Y_d)$ *then,* $\overline{\mathcal{E}}(X \to \mathbf{Y}) \leqslant \prod_i \overline{\mathcal{E}}(G_i \to Y_i)$*.*

4.3 Security of Type 3 Subsequences

Let $\mathbf{G} = (G_i)_{i=0}^d$ and $\mathbf{H} = (H_i)_{i=0}^d$ be d-th order encodings of $g(X)$ and $h(X)$ where g, h are given functions. This section proves security of type 3 subsequences involving the computations with the pairs (H_i, G_h). We need to introduce family of polynomials to express the security bound:

Definition 17 (Rook Domination Polynomial [Mer24]**).** *Let* $(E_{i,j})_{0 \leqslant i,j \leqslant d}$ *be a collection of independents events with respective probabilities* $((\overline{\mathcal{E}}_{i,j})_{0 \leqslant i,j \leqslant d})$*. Let*

$$\Upsilon((\overline{\mathcal{E}}_{i,j})_{0 \leqslant i,j \leqslant d}) \triangleq \mathbb{P}\left((\cap_{i=0}^d \cup_{j=0}^d E_{i,j}) \cup (\cap_{j=0}^d \cup_{i=0}^d E_{i,j}) \right). \tag{49}$$

For short $\Upsilon_d(\overline{\mathcal{E}}) \triangleq \Upsilon((\overline{\mathcal{E}}_{i,j})_{0 \leqslant i,j \leqslant d})$ *when for all* i, j *we have* $\overline{\mathcal{E}}_{i,j} = \overline{\mathcal{E}}$*.*

In fact, Υ_d corresponds to the domination polynomial of the rook graph [Mer24]. It can be sandwiched explicitly as follows:

Proposition 3

$$\max\{\prod_{i=0}^d (1 - \prod_{j=0}^d \mathcal{E}_{i,j}), \prod_{j=0}^d (1 - \prod_{i=0}^d \mathcal{E}_{i,j})\} \leqslant \Upsilon((\overline{\mathcal{E}}_{i,j})_{0 \leqslant i,j \leqslant d}) \tag{50}$$

$$\leqslant \min\{\prod_{i=0}^d (1 - \prod_{j=0}^d \mathcal{E}_{i,j}) + \prod_{j=0}^d (1 - \prod_{i=0}^d \mathcal{E}_{i,j}), 1\}. \tag{51}$$

In particular when for all i, j, $\overline{\mathcal{E}}_{i,j} = \overline{\mathcal{E}}$ *it yields:*

$$\left(1 - \mathcal{E}^{d+1}\right)^{d+1} \leqslant \Upsilon_d(\overline{\mathcal{E}}) \leqslant \min\{2\left(1 - \mathcal{E}^{d+1}\right)^{d+1}, 1\}. \tag{52}$$

Stephan Mertens [Mer24, Thm. 4] provides a recursive formula for this polynomial, which gives an efficient way to compute exhaustively the coefficients of Υ_d. The coefficients of the first polynomials Υ_d for $d \leqslant 5$ are shown in Table 2.

Rationale for the Rook Domination Polynomial. Within type 3 subsequences of ISW the cross-wise terms $G_i H_j$ are computed so that each pair (G_i, H_j) leaks. After degradation into an erasure channel, for each pair (i, j) the degraded adversary $\mathcal{A}$ either probes (G_i, H_j) or receives an erasure symbol. Let $E_{i,j}$ be the event that $\mathcal{A}$ probes the pair (G_i, H_j). By the secret sharing property, the sensitive value leaks if and only if we probe each G_i or each H_j.

Table 2. Some explicit values of $\Upsilon_d(\overline{\mathcal{E}})$ from small values of d.

d	$\Upsilon_d(\overline{\mathcal{E}})$
0	$\overline{\mathcal{E}}$
1	$6\overline{\mathcal{E}}^2\mathcal{E}^2 + 4\overline{\mathcal{E}}^3\mathcal{E} + \overline{\mathcal{E}}^4$
2	$48\overline{\mathcal{E}}^3\mathcal{E}^6 + 117\overline{\mathcal{E}}^4\mathcal{E}^5 + 126\overline{\mathcal{E}}^5\mathcal{E}^4 + 84\overline{\mathcal{E}}^6\mathcal{E}^3 + 36\overline{\mathcal{E}}^7\mathcal{E}^2 + 9\overline{\mathcal{E}}^8\mathcal{E} + \overline{\mathcal{E}}^9$
3	$488\overline{\mathcal{E}}^4\mathcal{E}^{12} + 2640\overline{\mathcal{E}}^5\mathcal{E}^{11} + 6712\overline{\mathcal{E}}^6\mathcal{E}^{10} + 10864\overline{\mathcal{E}}^7\mathcal{E}^9 + 12726\overline{\mathcal{E}}^8\mathcal{E}^8 + 11424\overline{\mathcal{E}}^9\mathcal{E}^7$ $+8008\overline{\mathcal{E}}^{10}\mathcal{E}^6 + 4368\overline{\mathcal{E}}^{11}\mathcal{E}^5 + 1820\overline{\mathcal{E}}^{12}\mathcal{E}^4 + 560\overline{\mathcal{E}}^{13}\mathcal{E}^3 + 120\overline{\mathcal{E}}^{14}\mathcal{E}^2 + 16\overline{\mathcal{E}}^{15}\mathcal{E} + \overline{\mathcal{E}}^{16}$
4	$6130\overline{\mathcal{E}}^5\mathcal{E}^{20} + 58300\overline{\mathcal{E}}^6\mathcal{E}^{19} + 269500\overline{\mathcal{E}}^7\mathcal{E}^{18} + 808325\overline{\mathcal{E}}^8\mathcal{E}^{17} + 1778875\overline{\mathcal{E}}^9\mathcal{E}^{16}$ $+3075160\overline{\mathcal{E}}^{10}\mathcal{E}^{15} + 4349400\overline{\mathcal{E}}^{11}\mathcal{E}^{14} + 5154900\overline{\mathcal{E}}^{12}\mathcal{E}^{13} + 5186300\overline{\mathcal{E}}^{13}\mathcal{E}^{12}$ $+4454400\overline{\mathcal{E}}^{14}\mathcal{E}^{11} + 3268360\overline{\mathcal{E}}^{15}\mathcal{E}^{10} + 2042950\overline{\mathcal{E}}^{16}\mathcal{E}^9 + 1081575\overline{\mathcal{E}}^{17}\mathcal{E}^8$ $+480700\overline{\mathcal{E}}^{18}\mathcal{E}^7 + 177100\overline{\mathcal{E}}^{19}\mathcal{E}^6 + 53130\overline{\mathcal{E}}^{20}\mathcal{E}^5 + 12650\overline{\mathcal{E}}^{21}\mathcal{E}^4 + 2300\overline{\mathcal{E}}^{22}\mathcal{E}^3$ $+300\overline{\mathcal{E}}^{23}\mathcal{E}^2 + 25\overline{\mathcal{E}}^{24}\mathcal{E}^1 + \overline{\mathcal{E}}^{25}$
5	$92592\overline{\mathcal{E}}^6\mathcal{E}^{30} + 1356480\overline{\mathcal{E}}^7\mathcal{E}^{29} + 9859140\overline{\mathcal{E}}^8\mathcal{E}^{28} + 47187180\overline{\mathcal{E}}^9\mathcal{E}^{27} + 167284836\overline{\mathcal{E}}^{10}\mathcal{E}^{26}$ $+469268496\overline{\mathcal{E}}^{11}\mathcal{E}^{25} + 1086623400\overline{\mathcal{E}}^{12}\mathcal{E}^{24} + 2137381200\overline{\mathcal{E}}^{13}\mathcal{E}^{23} + 3642777000\overline{\mathcal{E}}^{14}\mathcal{E}^{22}$ $+5453014080\overline{\mathcal{E}}^{15}\mathcal{E}^{21} + 7235196885\overline{\mathcal{E}}^{16}\mathcal{E}^{20} + 8558765100\overline{\mathcal{E}}^{17}\mathcal{E}^{19} + 9057864300\overline{\mathcal{E}}^{18}\mathcal{E}^{18}$ $+8591124600\overline{\mathcal{E}}^{19}\mathcal{E}^{17} + 7305959610\overline{\mathcal{E}}^{20}\mathcal{E}^{16} + 5567447160\overline{\mathcal{E}}^{21}\mathcal{E}^{15} + 3796214400\overline{\mathcal{E}}^{22}\mathcal{E}^{14}$ $+2310778800\overline{\mathcal{E}}^{23}\mathcal{E}^{13} + 1251676800\overline{\mathcal{E}}^{24}\mathcal{E}^{12} + 600805260\overline{\mathcal{E}}^{25}\mathcal{E}^{11} + 254186856\overline{\mathcal{E}}^{26}\mathcal{E}^{10}$ $+94143280\overline{\mathcal{E}}^{27}\mathcal{E}^9 + 30260340\overline{\mathcal{E}}^{28}\mathcal{E}^8 + 8347680\overline{\mathcal{E}}^{29}\mathcal{E}^7 + 1947792\overline{\mathcal{E}}^{30}\mathcal{E}^6 + 376992\overline{\mathcal{E}}^{31}\mathcal{E}^5$ $+58905\overline{\mathcal{E}}^{32}\mathcal{E}^4 + 7140\overline{\mathcal{E}}^{33}\mathcal{E}^3 + 630\overline{\mathcal{E}}^{34}\mathcal{E}^2 + 36\overline{\mathcal{E}}^{35}\mathcal{E}^1 + \overline{\mathcal{E}}^{36}$

Let $E \triangleq (\cap_{i=0}^{d} \cup_{j=0}^{d} E_{i,j}) \cup (\cap_{j=0}^{d} \cup_{i=0}^{d} E_{i,j})$. Equation (49) defines $\Upsilon \triangleq \mathbb{P}(E)$. If one represents the $E_{i,j}$ in a checkerboard, the sensitive value is probed if and only if at least one event $E_{i,j}$ occurs at least once in each line $(\cap_{i=0}^{d} \cup_{j=0}^{d} E_{i,j})$ or at least once in each column $(\cap_{j=0}^{d} \cup_{i=0}^{d} E_{i,j})$. This results in Eq. (49).

Considering that $E_{i,j}$ occurs when a rook is placed at coordinates (i, j) of the checkerboard, the event E is realized if and only if every position in the checkerboard is attacked (or occupied) by at least one rook. The domination polynomial of the rook graph counts the number of such configurations of m rooks. Therefore, it allows one to compute Υ_d when all $E_{i,j}$ are equiprobable.

We now show the security of type 3 subsequences using the domination polynomial of the rook graph:

Lemma 11 (Type 3 Subsequences). *Consider the channels* $((G_i, H_j) \to Y_{i,j})_{0 \leqslant i,j \leqslant d}$ *and let* $\mathbf{Y} \triangleq (Y_{i,j})_{0 \leqslant i,j \leqslant d}$. *Then one has*

$$\overline{\mathcal{E}}(X \to \mathbf{Y}) \leqslant \Upsilon((\overline{\mathcal{E}}((G_i, H_j) \to Y_{i,j}))_{0 \leqslant i,j \leqslant d}). \tag{53}$$

4.4 Security of Type 4 Subsequences

Type 4 subsequences consider the compression layer in multiplication gadgets. At this stage, the sensitive variable is masked in $(d+1)^2$ shares. In the compression layer, the shares are grouped in $d+1$ groups each of size $d+1$, and are recombined to obtain a d-th order encoding of the sensitive variable. Formally, let $(V_{i,j})$ be an encoding in $(d+1)^2$ shares of $f(X)$ where f is a given function. Let $T_{i,0} = V_{i,0}$ and $T_{i,j} = T_{i,j-1} \oplus V_{i,j}$. In particular $(T_{i,d})_{i=0}^{d}$ is a d-th order encoding of $f(X)$.

Lemma 12 (Cumulative Sum). *Consider the cumulative sum function $h_d : (x_0, \ldots, x_d) \in \mathbb{F}^{d+1} \mapsto (x_0, x_0 + x_1, \ldots, x_0 + \ldots + x_d) \in \mathbb{F}^{d+1}$. $(V_{i,0}, \ldots, V_{i,d})$ is a d-th order sharing of $T_{i,d}$ and since $T_{i,j} - T_{i,j-1} = V_{i,j}$, a channel from $(T_{i,j-1}, V_{i,j})$ can be seen as a channel from the pair $(T_{i,j}, T_{i,j-1})$. Let*

$$\mathbf{T}_i = (T_{i,0}, \ldots, T_{i,d+1}) \to \boxed{p_{\mathbf{Y}|\mathbf{T}_i} \triangleq \prod_{j=0}^{d} p_{Y_{i,j}|(T_{i,j-1},T_{i,j})}} \to \mathbf{Y}_i = (Y_{i,0}, \ldots, Y_{i,d})$$

with the convention that $T_{i,-1} = 0$. Then the channel rewrites,

$$T_{i,d} \to \boxed{\mathrm{Mask}_d} \to \mathbf{V_i} = (V_{i,0}, \ldots, V_{i,d}) \to \boxed{h_d} \to \mathbf{T}_i \to \boxed{P_{\mathbf{Y}|\mathbf{T}_i}} \to \mathbf{Y}_i \tag{54}$$

and $\overline{\mathcal{E}}(T_{i,d} \to \mathbf{Y}_i) \leqslant \overline{\mathcal{E}}((T_{i,d-1}, V_{i,d}) \to Y_{i,d})$.

We obtain now the following security result for type 4 subsequences:

Lemma 13 (Type 4 Subsequences). *Consider $((T_{i,j-1}, V_{i,j}) \to Y_{i,j})_{0 \leqslant i,j \leqslant d}$ and let $\mathbf{Y} = (Y_{i,j})_{0 \leqslant i,j \leqslant d}$ then, $\overline{\mathcal{E}}(X \to \mathbf{Y}) \leqslant \prod_{i=0}^{d} \overline{\mathcal{E}}((T_{i,d-1}, V_{i,d}) \to Y_{i,d})$.*

4.5 Security for the Whole Circuit

Combining Lemmas 10, 11, 13 for the different subsequences together with the single-letterization Lemma 6 we obtain a security guarantee for the whole circuit:

Theorem 2 (Direct Security Proof). *Consider an implementation with n_i subsequences of type i ($i = 1, 2, 3, 4$) and a $\overline{\mathcal{E}}$-noisy adversary with respect to* CDC *with q queries. Let $\mathbf{Y}$ be the vector of all corresponding side-informations acquired by the chosen channel adversary. Then one has*

$$0 \leqslant \overline{\mathcal{E}}(K \to \mathbf{Y}) \leqslant 1 - \left(\left(1 - \overline{\mathcal{E}}^{d+1}\right)^{n_1+n_2+n_4} \left(1 - \Upsilon_d(\overline{\mathcal{E}})\right)^{n_3}\right)^q \leqslant 1. \tag{55}$$

The upper bound can be weakened via the union bound *to*

$$\overline{\mathcal{E}}(K \to \mathbf{Y}) \leqslant q \left((n_1 + n_2 + n_4) + 2n_3(d+1)^{d+1}\right) \overline{\mathcal{E}}^{d+1}. \tag{56}$$

Also, using the asymptotic equivalent of the domination polynomial of the rook graph the upper bound of Eq. (55) is asymptotically equivalent to

$$q \left(n_1 + n_2 + \left(2(d+1)^{d+1} - (d+1)!\right) n_3 + n_4\right) \overline{\mathcal{E}}^{d+1}. \tag{57}$$

Remark 13. Equation (56) is of a similar form as [MS23b, Thm. 5] but the constants n_i are not scaled by a term depending on the field size $|\mathcal{X}|$, contrary to the t_i occurring in [MS23b, Thm. 7].

De Chérisey *et al.* obtained a lower bound on the minimum number of queries required by the adversary to achieve a given advantage in terms of SR in the unprotected setting with MI [dCGRP19, Thm. 2, Eq. 4]. Béguinot *et al.* derived a tight bound for masked encoding with MI [BCG+23, Coro. 2] and maximal leakage [BLR+23, Coro. 1]. Liu *et al.* derived a tight bound for masked encoding with Sibson's α-information of order 2 [LBC+23, Thm. 2]. The combination of Theorem 2 with Proposition 1 yields a lower bound on the minimum number of queries required by the adversary to achieve a given advantage in terms of SR, GE or TVI for the entire protected computations (not only encodings).

Theorem 3 (Lower Bound on the Number of Queries). *Let*

$$\lambda(\overline{\mathcal{E}}, d) = \Big(\ln\Big(\big(1-\overline{\mathcal{E}}^{d+1}\big)^{n_1+n_2+n_4}\big(1-\Upsilon_d(\overline{\mathcal{E}})\big)^{n_3}\Big)\Big)^{-1} \tag{58}$$

$$= \Big((n_1+n_2+n_4)\log\big(1-\overline{\mathcal{E}}^{d+1}\big) + n_3\log\big(1-\Upsilon_d(\overline{\mathcal{E}})\big)\Big)^{-1} \tag{59}$$

$$\approx \Big(\big(n_1+n_2+n_4+n_3(2(d+1)^{d+1}-(d+1)!)\big)\,\overline{\mathcal{E}}^{d+1}\Big)^{-1}. \tag{60}$$

The number of queries to achieve $\mathbb{P}_{s,o}(K|\mathbf{Y}) = \mathbb{P}_{s,o}$, $G(K|\mathbf{Y}) = G$ *or* $\Delta(K;Y) = \Delta$ *is at least:*

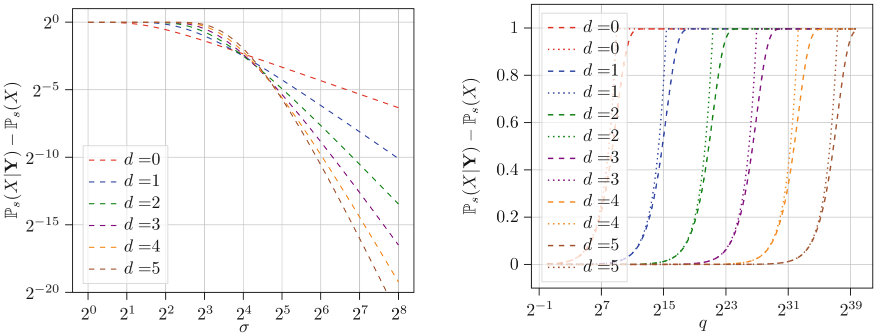
$$\begin{aligned} q_{\mathrm{sr}} &\geqslant \lambda(\overline{\mathcal{E}}, d)\ln\big((1-\mathbb{P}_{s,o})^{-1}\lambda_{\mathrm{SR}_o}\big),\\ q_{\mathrm{ge}} &\geqslant \lambda(\overline{\mathcal{E}}, d)\ln\big((G-1)^{-1}\lambda_{\mathrm{GE}}\big),\\ q_{\mathrm{tvi}} &\geqslant \lambda(\overline{\mathcal{E}}, d)\ln\big(\Delta^{-1}\lambda_{\mathrm{TVI}}\big). \end{aligned} \tag{61}$$

Theorem 3 is illustrated by Fig. 3. It shows how it behaves for a fixed number of queries and increasing level of noise in Fig. 3a and how it behaves for a fixed value of the CDC and increasing number of queries q in Fig. 3b. We can observe on Fig. 3a that there is an optimal masking order with respect to Theorem 3 depending on the noise level. Further, Fig. 3b shows that the bound benefits from the single letterization of Lemma 5 and Lemma 6 as it is an "S"-shaped curve. The weakened version of Lemma 6 into a linear bound corresponds to the dotted line.

4.6 Comparison With Bounds Based on Mutual Information

Theorem 2 provides bounds on all figures of merits that we compare with bounds based on MI. Let $K \to Y$ be a side-channel of a uniform key $K \sim \mathcal{U}(\mathbb{F}_2^n)$. Let $Y = \mathrm{HW}(K) + N$ where $N \sim \mathcal{N}(0, \sigma^2)$ is an additive Gaussian noise. Let $\mathbf{Y} = (Y_1, \ldots, Y_q)$ be a side-channel leakage with q queries.

No Constraint on the Noise Level. A strength of Theorem 2 is that the bound on the CDC (Eq. (55)) is always less than 1. In particular, our bound does not require any constraint on the noise level to apply. This is by contrast with MI based bound such as [MS23b].



(a) One query ($q = 1$) and variable σ under Hamming weight leakage model.

(b) $\overline{\mathcal{E}} = 2 \times 10^{-3}$ and increasing number of queries q.

Fig. 3. Plots of Theorem 3 for one subsequence of type 3 ($n_1 = n_2 = n_4 = 0, n_3 = 1$) and different masking order. The y-axis corresponds to the adversary's advantage in terms of success rate.

Scaling with the Number of Queries. De Chérisey *et al.* [dCGRP19, Lemma 5] established an upper bound on the MI for side-channels that grows linearly with respect to q i.e., $I(K;\mathbf{Y}) \leqslant qI(K;Y)$. However, the MI $I(K;\mathbf{Y})$ is bounded above by the constant $\log|\mathcal{K}|$. Therefore, the linear bound cannot be tight for large values of q. Finding an alternative practical non-linear bound on the MI that remains bounded when q increases remains an interesting open question. However, our bound $\overline{\mathcal{E}}(K \to \mathbf{Y}) \leqslant 1 - (1 - \overline{\mathcal{E}}(K \to Y))^q \leqslant q\overline{\mathcal{E}}(K \to Y)$ of Lemma 6 for the CDC is nonlinear with respect to q and can be weakened into a linear bound using Boole's inequality. This linear approximation is tight when $\overline{\mathcal{E}} \to 0$.

Attack with One Trace (SPA) and Multiple Traces (DPA). On the one hand, Béguinot *et al.* [BLR+23, Eq. 87] showed that as $I(K;\mathbf{Y}) \to 0$, we have

$$\mathbb{P}_s(K|\mathbf{Y}) - \mathbb{P}_s(K) \leqslant \sqrt{\frac{2(2^n-1)}{2^{2n}} \frac{qI(K;Y)}{\log e}}. \tag{62}$$

On the other hand, Proposition 1 yields

$$\mathbb{P}_s(K|\mathbf{Y}) - \mathbb{P}_s(K) \leqslant (1 - (1 - \overline{\mathcal{E}}(K \to Y))^q)\lambda_{\text{SR}} \approx q\overline{\mathcal{E}}(K \to Y)\lambda_{\text{SR}}. \tag{63}$$

While $\delta_{\text{MI}} \propto \frac{1}{\sigma^2}$ [BCPZ16, Appendix A] and $\delta_{\text{CDC}} \propto \frac{1}{\sigma}$ (Proposition 2), the square root in the MI-based security bounds (e.g., Eq. (62)) makes both bounds on the SR advantage to be $O(\sigma^{-1})$. As a consequence, the CDC-based bound is comparable with the MI-based bound for single trace attack ($q = 1$) and at large noise.

However, in the context of a DPA where the goal is to lower bound the minimum number of queries q to achieve a given figure of merit the MI based bound Eq. (62) will be in $O(\sigma^2)$ while the CDC based bound Eq. (62) is only $O(\sigma)$. In the presence of masking the same observation applies to Theorem 3 and [MS23b, Theorem 7] where σ is replaced by σ^{d+1}. In conclusion CDC may not be the most suitable informational noisiness measure to capture the leakage in a DPA with many traces. However, since the factorization is optimal this leads to the important conclusion that such a loss is inherent to a reduction from noisy leakages to the random probing model. For this reason informational bounds based on MI remain an interesting tool for side-channel analysis.

5 Indirect Proofs *via* CDC With Random Probing

In this Section, we explain how existing proof in the random probing model can be combined with the CDC. We show how a security proof in the random probing model can be lifted into the noisy leakage model by improving [DDF14, Thm. 1] in Subsect. 5.1. The resulting bound is shown to yield an upper bound on the side-channel adversary's advantage in Subsect. 5.2. The asymptotic behavior of the security bound is analyzed in Subsect. 5.3. This analysis confirms theoretically the finding from Battistello *et al.* [BCPZ16] that increasing indefinitely the masking order of ISW gadgets whose noise rate is not constant can be detrimental to security.

5.1 Lifting Security Proof in the Random Probing Model to Noisy Leakage

We first refer back the existing security proofs in the random probing model. Cassiers *et al.* [CFOS21] showed how to derive tight security bounds in the random probing model using *probe distribution table* (PDT). Belaïd *et al.* [BCP+20,BRT21] also derived security proof in the random probing model based on an *expansion strategy* of small gadgets. Their improvements are based on the fact that when an adversary probes more than t wires it does not necessarily learn any information about the sensitive variable. The t-threshold probing security ensures that no subset of at most t wires leak information. However, some subsets of more than t wires does not leak information either. By carefully determining the subsets of leaking wires and the subset of non-leaking wires the so-called *probability of simulation failure* can be reduced.

Our goal is to show how we can lift a proof in the random probing model to the noisy leakage model using CDC. To do so we show how to improve [DDF14, Thm. 1]. In the same way the above-mentioned proof in the random probing model can be lifted to the noisy leakage model using the CDC.

In this section, the circuit Γ is decomposed in $|\Gamma|$ regions (gadgets) whose numbers of wires is specified by the sequence (l_i). We assume that Γ is secure in the region probing model, i.e. any set of at most t (probed) wires in each region of the circuit is independent with the secret key.

Theorem 4 (Indirect Security Proof). *Let $\mathcal{A}$ be a $\overline{\mathcal{E}}$-noisy adversary with respect to* CDC *with q queries. Let $\mathbf{Y}$ be the vector of all corresponding side-information acquired by the chosen channel adversary. Then one has*

$$\overline{\mathcal{E}}(K \to \mathbf{Y}) \leqslant \text{fail}(t, (l_i), \overline{\mathcal{E}}, q) \triangleq 1 - \prod_{i=1}^{|\Gamma|} \Big(1 - Q_B(t, l_i, \overline{\mathcal{E}})\Big)^q \leqslant q \sum_{i=1}^{|\Gamma|} Q_B(t, l_i, \overline{\mathcal{E}}). \tag{64}$$

Remark 14. Here the wires leak while the gates leak in Sect. 4.

Theorem 4 is very generic, it is "agnostic" to the countermeasure implemented to achieve security against t-threshold probing adversary in the region probing model. Masure & Standaert observed in [MS23b, Tab. 1] that [DDF14, Thm. 1] does not provide incentive to noisier leakage. Theorem 4 does not suffer from this weakness. In particular, the adversary's advantage vanishes as the noise level increases. Theorem 4 depends on five parameters:

1. The noise level as quantified by $\overline{\mathcal{E}}$.
2. The security order of the gadgets as quantified by t.
3. The "temporal attack surface" quantified by the number of queries q.
4. The attack surface of the adversary within a gadget as measured by the number l of wires potentially probed within the gadgets.
5. The attack surface of the adversary on the whole circuit as measured by $|\Gamma|$.

5.2 Bounds on SCA Advantage

If a cryptographic algorithm is insecure against a black-box adversary then it is also insecure against a side-channel adversary. For this reason we are interested in the advantage of the side-channel adversary compared to its black-box counterpart. In Proposition 4 we upper bound this advantage, contrary to the usual bound for side-channel attacks it depends on the computational assumption made on the adversary through the term $\mathbb{P}^{\text{BB}}_{s,o}(q)$.

Proposition 4. *Let $\mathcal{A}$ be a $\overline{\mathcal{E}}$-noisy with respect to* CDC *adversary with q queries. Let $\mathbb{P}^{\text{BB}}_{s,o}(q)$ and $\mathbb{P}^{\text{SCA}}_{s,o}(q, \overline{\mathcal{E}})$ be the respective* SR_o *of the best black box and side-channel adversary with q queries. Then,*

$$\begin{cases} \mathbb{P}^{\text{SCA}}_{s,o}(q, \overline{\mathcal{E}}) - \mathbb{P}^{\text{BB}}_{s,o}(q) & \leqslant \big(1 - \mathbb{P}^{\text{BB}}_{s,o}(q)\big) \quad \text{fail}(t, (l_i), \overline{\mathcal{E}}, q) \\ \text{GE}^{\text{BB}}(q) - \text{GE}^{\text{SCA}}(q, \overline{\mathcal{E}}) & \leqslant \big(\text{GE}^{\text{BB}}(q) - 1\big) \ \text{fail}(t, (l_i), \overline{\mathcal{E}}, q). \end{cases} \tag{65}$$

5.3 Properties of the Failure Probability Function

We analyze how our bound depends on the order of an implemented countermeasure. Consider a masking countermeasure of order d, and for illustrative purposes assume that $t = d$ (which would obtained for the implementation of [RP10b] separated by leak-free refresh). Also assume that the l_i associated with the multiplication gadgets grow quadratically with respect to d (e.g., ISW). We use the shorthand notations $l(d) = \max_i l_i(d)$, $\mathrm{fail}(d, \overline{\mathcal{E}}) = \mathrm{fail}(t(d), (l_i(d)), \overline{\mathcal{E}}, 1)$ and $\mathrm{fail}(d, \overline{\mathcal{E}}, q) = \mathrm{fail}(t(d), (l_i(d)), \overline{\mathcal{E}}, q)$. We can define an optimal masking order with respect to Theorem 4:

$$d^*(\overline{\mathcal{E}}) \triangleq \arg\min_{d \in \mathbb{N}} \mathrm{fail}(d, \overline{\mathcal{E}}, q). \tag{66}$$

Since $\mathrm{fail}(d, \overline{\mathcal{E}}, q) = 1 - \left(1 - \mathrm{fail}(d, \overline{\mathcal{E}})\right)^q$ we have $d^*(\overline{\mathcal{E}}) = \arg\min_{d \in \mathbb{N}} \mathrm{fail}(d, \overline{\mathcal{E}})$ so that the optimal masking order is independent of the number of queries.

It is not true in general that $\mathrm{fail}(d, \overline{\mathcal{E}})$ is a decreasing function of the masking order d. Though we show that it essentially holds in the limits of high noise.

Proposition 5. *For all $d_1 < d_2$ there exist a noise threshold $\overline{\mathcal{E}}_0$ such that for all noise level $\overline{\mathcal{E}} \leqslant \overline{\mathcal{E}}_0$, $\mathrm{fail}(d_1, \overline{\mathcal{E}}) > \mathrm{fail}(d_2, \overline{\mathcal{E}})$, which indicates that $d^*(\overline{\mathcal{E}}) \xrightarrow{\overline{\mathcal{E}} \to 0} \infty$.*

Proposition 5 means that while the bound is not decreasing with respect to the masking order d there always exists a noise level for which masking at higher order is more interesting.

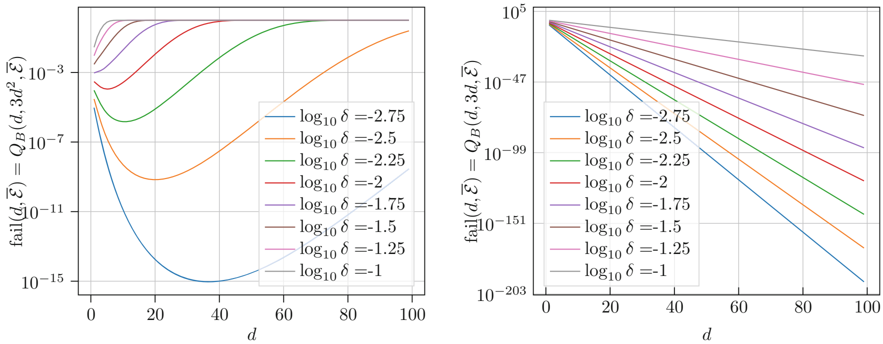
Proposition 6. *If $l(d)/t(d) \xrightarrow{d \to \infty} \infty$ then $\mathrm{fail}(d, \overline{\mathcal{E}}) \xrightarrow{d \to \infty} 1$. Therefore, there exist a finite optimal masking order with respect to the noise level $d^*(\overline{\mathcal{E}}) < \infty$ and $\mathrm{fail}(d, \overline{\mathcal{E}})$ cannot be reduced further than $\mathrm{fail}(d^*, \overline{\mathcal{E}}) > 0$. This in turn implies that the adversary's advantage cannot be made arbitrarily small.*

Proposition 6 means that for a fixed level of noise increasing indefinitely the masking order is detrimental to the security bounds. As a consequence, there exists a finite optimal masking order with respect to the noise level.

Proposition 7. *If $\overline{\mathcal{E}} < \overline{\mathcal{E}}_0 \leqslant \frac{t}{l}$ then $Q_B(t, l, \overline{\mathcal{E}}) \leqslant \exp\left(-l d_{\mathrm{KL}}\left(t/l \,\|\overline{\mathcal{E}}\right)\right) \leqslant \exp\left(-l d_{\mathrm{KL}}\left(\overline{\mathcal{E}}_0 \,\|\overline{\mathcal{E}}\right)\right)$ where $d_{\mathrm{KL}}(p\|q) \triangleq D_{\mathrm{KL}}(\mathcal{B}(p)\|\mathcal{B}(q)) = p \log \frac{p}{q} + \bar{p} \log \frac{\bar{p}}{\bar{q}}$. In this case $d^*(\overline{\mathcal{E}}) = \infty$.*

Proposition 7 shows that when the gadget size grows at most linearly with respect to the security parameter t then masking provides an exponential gain with respect to the gadget size. The coefficient in the exponential is lower bounded by a binary divergence between a *protection rate* $\frac{t}{l}$ and a *leakage rate* $\overline{\mathcal{E}}$.

The original expression of [DDF14, Thm. 1] can be misinterpreted as it seems that the probability of error converges to 0 when d increases. But when the gadget size l grows quadratically with respect to d then the maximum value of d tolerated in the proof is upper bounded by a function inversely proportional to the noise level. Hence, it is not true to say that the advantage of the adversary decreases exponentially with respect to d. This confirms the observations from [BCPZ16] where Battistello *et al.* observed that the noise is expected to decrease linearly with the number of shares and that this assumption is not met in practice. Figure 4a derived from Theorem 4 shows that for quadratic gadget and a fixed level of noise, the advantage of the adversary is either increasing or decreasing and then increasing with respect to the masking order depending on the noise level. For linear gadget, Fig. 4b shows that masking does provide an exponential gain with respect to the adversaries advantage. Note that this is a weakness of the ISW gadget whose noise rate is not constant and not a weakness of the bound. This emphasizes the importance to obtain gadget with improved noise rate [CS19]. In particular quasi-linear masking scheme [GJR18, GPRV22, CDGT24] and Toom-Cook based gagdets [Pla22] are promising approaches.



(a) Quadratic gadget (multiplication) with $t = d$ and $l = 3d^2$.

(b) Linear gadget (linear operation) with $t = d$ and $l = 3d$.

Fig. 4. Evolution of Theorem 4 for different masking order d for quadratic and linear gadget.

Remark 15. Proposition 6 may appear as contradictory with [DFS19, Coro. 2]. It turns out that [DFS19, Coro. 2] is incorrectly derived from [DFS19, Thm. 3].

6 Conclusion

We showed how the complementary Doeblin coefficient (CDC) can be used to reduce optimally a noisy adversary to a random probing adversary. This allows us to exhibit the unavoidable inherent cost of a reduction from noisy leakages to the random probing model. We derived a set of properties of the CDC which makes it a sound leakage measure that is easy to manipulate and showed that it is amenable to evaluation in a multivariate setting.

The CDC yields security bounds for all figures of merits that scale well with the number of side-channel queries (single letterization property). As a byproduct we also lower bounded the minimum number of queries to achieve a given figure of merit.

Furthermore, security bounds in terms of CDC are easily derived using the Prouff-Rivain subsequence decomposition or can be naturally combined with existing security proofs in the random probing model for any type of countermeasures. We analyzed the asymptotic behavior of the obtained bounds in terms of countermeasure order and confirmed the existence of an optimal masking order with respect to the security bounds.

Overall, we believe that these contributions are essential to ground the security of masked implementations in the noisy leakage model on solid foundations. As perspectives, we would like to obtain direct security proof of code-based masking implementation using CDC leveraging a new appropriate subsequence decomposition. Investigating formal security proof of masking combined with shuffling [ABG+22] using CDC could also be relevant as a way to reduce the physical noise requirement to obtain relevant security parameters.

Acknowledgements. We are grateful to Stephan Mertens that provided us an efficient way to compute the coefficients of the rook domination polynomials appearing in Type 3 subsequence. We also thank him for sharing us the explicit values of the coefficients of the domination polynomials up to $d = 20$. We are also grateful to Loïc Masure, François-Xavier Standaert, Matthieu Rivain and Thomas Prest for our very insightful discussions. We would also like to thank the anonymous reviewers for their in-depth reviews of the article and their extremely valuable suggestions. Secure-IC acknowledges partial funding from the European Union's Horizon Europe research and innovation programme through the ALLEGRO project, under grant agreement No. 101070009. Julien Béguinot is a PhD candidate funded by Institut Mines-Télécom through the *Future & Ruptures* program. Wei Cheng is also partially supported by National Key R&D Program of China No. 2022YFB3103800.

References

[AARR02] Agrawal, D., Archambeault, B., Rao, J.R., Rohatgi, P.: The EM side—channel(s). In: Kaliski, B.S., Koç, K., Paar, C. (eds.) CHES 2002. LNCS, vol. 2523, pp. 29–45. Springer, Heidelberg (2003). https://doi.org/10.1007/3-540-36400-5_4

[AB00] Abdalla, M., Bellare, M.: Increasing the lifetime of a key: a comparative analysis of the security of re-keying techniques. In: Okamoto, T. (ed.) ASIACRYPT 2000. LNCS, vol. 1976, pp. 546–559. Springer, Heidelberg (2000). https://doi.org/10.1007/3-540-44448-3_42

[ABG+22] Azouaoui, M., Bronchain, O., Grosso, V., Papagiannopoulos, K., Standaert, F.-X.: Bitslice masking and improved shuffling: how and when to mix them in software? IACR Trans. Cryptogr. Hardw. Embed. Syst. **2022**(2), 140–165 (2022)

[BB11] Bloch, M.R., Barros, J.: Physical-Layer Security: From Information Theory to Security Engineering. Cambridge University Press (2011)

[BBD+16] Barthe, G.: Strong non-interference and type-directed higher-order masking. In: Weippl, E.R. Katzenbeisser, S., Kruegel, C., Myers, A.C., Halevi, S. (eds.) Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, 24–28 October 2016, pp. 116–129. ACM (2016)

[BCG+23] Béguinot, J., et al.: Removing the field size loss from Duc et al.'s conjectured bound for masked encodings. In: Kavun, E.B., Pehl, M. (eds.) COSADE

2023. LNCS, vol. 13979, pp. 86–104. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-29497-6_5

[BCGR24] Béguinot, J., Cheng, W., Guilley, S., Rioul, O.: Formal security proofs via Doeblin coefficients: optimal side-channel factorization from noisy leakage to random probing. IACR Cryptology ePrint Archive, p. 199 (2024). https://eprint.iacr.org/2024/199.pdf

[BCO04] Brier, E., Clavier, C., Olivier, F.: Correlation power analysis with a leakage model. In: Joye, M., Quisquater, J.-J. (eds.) CHES 2004. LNCS, vol. 3156, pp. 16–29. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-28632-5_2

[BCP+20] 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.) CRYPTO 2020, Part I. LNCS, vol. 12170, pp. 339–368. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-56784-2_12

[BCPZ16] 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.) CHES 2016. LNCS, vol. 9813, pp. 23–39. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-662-53140-2_2

[BFO+22] Brian, G., et al.: The mother of all leakages: how to simulate noisy leakages via bounded leakage (almost) for free. IEEE Trans. Inf. Theory **68**(12), 8197–8227 (2022)

[BHM+19] Bronchain, O., Hendrickx, J.M., Massart, C., Olshevsky, A., Standaert, F.-X.: Leakage certification revisited: bounding model errors in side-channel security evaluations. In: Boldyreva, A., Micciancio, D. (eds.) CRYPTO 2019, Part I. LNCS, vol. 11692, pp. 713–737. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-26948-7_25

[BLR+23] Béguinot, J., Liu, Y., Rioul, O., Cheng, W., Guilley, S.: Maximal leakage of masked implementations using Mrs. Gerber's lemma for min-entropy. CoRR, abs/2305.06276 (2023)

[BRT21] 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.) EUROCRYPT 2021, Part II. LNCS, vol. 12697, pp. 313–343. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-77886-6_11

[CDGT24] Carlet, C., Daif, A., Guilley, S., Tavernier, C.: Quasi-linear masking against SCA and FIA, with cost amortization. IACR Trans. Cryptogr. Hardw. Embed. Syst. **2024**(1), 398–432 (2024)

[CFOS21] Cassiers, G., Faust, S., Orlt, M., Standaert, F.-X.: Towards tight random probing security. In: Malkin, T., Peikert, C. (eds.) CRYPTO 2021. LNCS, vol. 12827, pp. 185–214. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-84252-9_7

[CJRR99] Chari, S., Jutla, C.S., Rao, J.R., Rohatgi, P.: Towards sound approaches to counteract power-analysis attacks. In: Wiener, M. (ed.) CRYPTO 1999. LNCS, vol. 1666, pp. 398–412. Springer, Heidelberg (1999). https://doi.org/10.1007/3-540-48405-1_26

[CL10] Chestnut, S., Lladser, M.E.: Occupancy distributions in Markov chains via Doeblin's ergodicity coefficient. In: DMTCS Proceedings, 21st International Meeting on Probabilistic, Combinatorial, and Asymptotic Methods in the Analysis of Algorithms (AofA 2010) (2010)

[CRR02] Chari, S., Rao, J.R., Rohatgi, P.: Template attacks. In: Kaliski, B.S., Koç, K., Paar, C. (eds.) CHES 2002. LNCS, vol. 2523, pp. 13–28. Springer, Heidelberg (2003). https://doi.org/10.1007/3-540-36400-5_3

[CS19] Cassiers, G., Standaert, F.-X.: Towards globally optimized masking: from low randomness to low noise rate or probe isolating multiplications with reduced randomness and security against horizontal attacks. IACR Trans. Cryptogr. Hardw. Embed. Syst. **2019**(2), 162–198 (2019)

[CS21] Coron, J.-S., Spignoli, L.: Secure wire shuffling in the probing model. In: Malkin, T., Peikert, C. (eds.) CRYPTO 2021. LNCS, vol. 12827, pp. 215–244. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-84252-9_8

[dCGRP19] de Chérisey, É., Guilley, S., Rioul, O., Piantanida, P.: Best information is most successful – mutual information and success rate in side-channel analysis. IACR Trans. Cryptogr. Hardw. Embed. Syst. **2019**(2), 49–79 (2019)

[DDF14] Duc, A., Dziembowski, S., Faust, S.: Unifying leakage models: from probing attacks to noisy leakage. In: Nguyen, P.Q., Oswald, E. (eds.) EUROCRYPT 2014. LNCS, vol. 8441, pp. 423–440. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-642-55220-5_24

[DDF19] Duc, A., Dziembowski, S., Faust, S.: Unifying leakage models: from probing attacks to noisy leakage. J. Cryptol. **32**(1), 151–177 (2019)

[DFS15] Duc, A., Faust, S., Standaert, F.-X.: Making masking security proofs concrete. In: Oswald, E., Fischlin, M. (eds.) EUROCRYPT 2015, Part I. LNCS, vol. 9056, pp. 401–429. Springer, Heidelberg (2015). https://doi.org/10.1007/978-3-662-46800-5_16

[DFS19] Duc, A., Faust, S., Standaert, F.-X.: Making masking security proofs concrete (or how to evaluate the security of any leaking device), extended version. J. Cryptol. **32**(4), 1263–1297 (2019)

[DKL+98] Dhem, J.-F., Koeune, F., Leroux, P.-A., Mestré, P., Quisquater, J.-J., Willems, J.-L.: A practical implementation of the timing attack. In: Quisquater, J.-J., Schneier, B. (eds.) CARDIS 1998. LNCS, vol. 1820, pp. 167–182. Springer, Heidelberg (2000). https://doi.org/10.1007/10721064_15

[Dob56] Dobrushin, R.L.: Central limit theorem for nonstationary Markov chains. i. Theory Probab. Its Appl. **1**(1), 65–80 (1956)

[Doe37] Doeblin, W.: Le cas discontinu des probabilités en chaîne. Publ. Fac. Sci. Univ. Masaryk (Brno) **236**, 1–13 (1937). munispace.muni.cz/library/catalog/book/1837

[DP08] Dziembowski, S., Pietrzak, K.: Leakage-resilient cryptography. In: 49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, 25–28 October 2008, Philadelphia, PA, USA, pp. 293–302. IEEE Computer Society (2008)

[EVG22] Esposito, A.R., Vandenbroucque, A., Gastpar, M.: On Sibson's α-mutual information. In: IEEE International Symposium on Information Theory, ISIT 2022, Espoo, Finland, 26 June–1 July 2022, pp. 2904–2909. IEEE (2022)

[GGK20] Gohari, A., Günlü, O., Kramer, G.: Coding for positive rate in the source model key agreement problem. IEEE Trans. Inf. Theory **66**(10), 6303–6323 (2020)

[GJR18] Goudarzi, D., Joux, A., Rivain, M.: How to securely compute with noisy leakage in quasilinear complexity. In: Peyrin, T., Galbraith, S. (eds.) ASIACRYPT 2018, Part II. LNCS, vol. 11273, pp. 547–574. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-03329-3_19

[GMO01] Gandolfi, K., Mourtel, C., Olivier, F.: Electromagnetic analysis: concrete results. In: Koç, Ç.K., Naccache, D., Paar, C. (eds.) CHES 2001. LNCS, vol. 2162, pp. 251–261. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-44709-1_21

[GPRV22] Goudarzi, D., Prest, T., Rivain, M., Vergnaud, D.: Probing security through input-output separation and revisited quasilinear masking. IACR Cryptology ePrint Archive, p. 45 (2022)

[ISW03] Ishai, Y., Sahai, A., Wagner, D.: Private circuits: securing hardware against probing attacks. In: Boneh, D. (ed.) CRYPTO 2003. LNCS, vol. 2729, pp. 463–481. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-540-45146-4_27

[IUH22] Ito, A., Ueno, R., Homma, N.: Perceived information revisited new metrics to evaluate success rate of side-channel attacks. IACR Trans. Cryptogr. Hardw. Embed. Syst. **2022**(4), 228–254 (2022)

[IWK20] Issa, I., Wagner, A.B., Kamath, S.: An operational approach to information leakage. IEEE Trans. Inf. Theory **66**(3), 1625–1657 (2020)

[KGG+18] Kocher, P., et al.: Spectre attacks: exploiting speculative execution. CoRR, abs/1801.01203 (2018)

[KJJ99] Kocher, P., Jaffe, J., Jun, B.: Differential power analysis. In: Wiener, M. (ed.) CRYPTO 1999. LNCS, vol. 1666, pp. 388–397. Springer, Heidelberg (1999). https://doi.org/10.1007/3-540-48405-1_25

[KR19] Kalai, Y.T., Reyzin, L.: A survey of leakage-resilient cryptography. In: Goldreich, O. (ed.) Providing Sound Foundations for Cryptography: On the Work of Shafi Goldwasser and Silvio Micali, pp. 727–794. ACM (2019)

[LBB19] Lagasse, J., Bartoli, C., Burleson, W.P.: Combining clock and voltage noise countermeasures against power side-channel analysis. In: 30th IEEE International Conference on Application-Specific Systems, Architectures and Processors, ASAP 2019, New York, NY, USA, 15–17 July 2019, pp. 214–217. IEEE (2019)

[LBC+23] Liu, Y., et al.: Improved alpha-information bounds for higher-order masked cryptographic implementations. In: IEEE Information Theory Workshop, ITW 2023, Saint-Malo, France, 23–28 April 2023, pp. 81–86. IEEE (2023)

[LH20] Lee, J.H., Han, D.-G.: Security analysis on dummy based side-channel countermeasures - case study: AES with dummy and shuffling. Appl. Soft Comput. **93**, 106352 (2020)

[LRP07] Lemke-Rust, K., Paar, C.: Gaussian mixture models for higher-order side channel analysis. In: Paillier, P., Verbauwhede, I. (eds.) CHES 2007. LNCS, vol. 4727, pp. 14–27. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-74735-2_2

[Mak20] Makur, A.: Coding theorems for noisy permutation channels. IEEE Trans. Inf. Theory **66**(11), 6723–6748 (2020)

[Mas94] Massey, J.L.: Guessing and entropy. In: Proceedings of 1994 IEEE International Symposium on Information Theory, p. 204, June 1994

[Mer24] Mertens, S.: Domination polynomial of the rook graph. J. Integer Sequences **27** (2024). Article 24.3.7. arXiv preprint arXiv:2401.00716

[MR04] Micali, S., Reyzin, L.: Physically observable cryptography. In: Naor, M. (ed.) TCC 2004. LNCS, vol. 2951, pp. 278–296. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-24638-1_16

[MS23a] Makur, A., Singh, J.: Doeblin coefficients and related measures. arXiv preprint arXiv:2309.08475 (2023)

[MS23b] Masure, L., Standaert, F.-X.: Prouff and Rivain's formal security proof of masking, revisited - tight bounds in the noisy leakage model. In: Handschuh, H., Lysyanskaya, A. (eds.) CRYPTO 2023, Part III. LNCS, vol. 14083, pp. 343–376. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-38548-3_12

[MSS09] Moradi, A., Shalmani, M.T.M., Salmasizadeh, M.: Dual-rail transition logic: a logic style for counteracting power analysis attacks. Comput. Electric. Eng. **35**(2), 359–369 (2009)

[PGMP19] Prest, T., Goudarzi, D., Martinelli, A., Passelègue, A.: Unifying leakage models on a Rényi day. In: Boldyreva, A., Micciancio, D. (eds.) CRYPTO 2019. LNCS, vol. 11692, pp. 683–712. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-26948-7_24

[Pla22] Plançon, M.: Exploiting algebraic structures in probing security. IACR Cryptology ePrint Archive, p. 1540 (2022)

[PR13] Prouff, E., Rivain, M.: Masking against side-channel attacks: a formal security proof. In: Johansson, T., Nguyen, P.Q. (eds.) EUROCRYPT 2013. LNCS, vol. 7881, pp. 142–159. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-38348-9_9

[PW23] Polyanskiy, Y., Wu, Y.: Information Theory, From Coding to Learning, 1st edn. Cambridge University Press (2023)

[RP10a] Rivain, M., Prouff, E.: Provably secure higher-order masking of AES. IACR Cryptology ePrint Archive, 441 (2010). Extended version of [RP10b]

[RP10b] Rivain, M., Prouff, E.: Provably secure higher-order masking of AES. In: Mangard, S., Standaert, F.-X. (eds.) CHES 2010. LNCS, vol. 6225, pp. 413–427. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15031-9_28

[RSV+11] Renauld, M., Standaert, F.-X., Veyrat-Charvillon, N., Kamel, D., Flandre, D.: A formal study of power variability issues and side-channel attacks for nanoscale devices. In: Paterson, K.G. (ed.) EUROCRYPT 2011. LNCS, vol. 6632, pp. 109–128. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-20465-4_8

[Sen73] Seneta, E.: On the historical development of the theory of finite inhomogeneous Markov chains. Math. Proc. Cambridge Philos. Soc. **74**(3), 507–513 (1973)

[SMY09] Standaert, F.-X., Malkin, T.G., Yung, M.: A unified framework for the analysis of side-channel key recovery attacks. In: Joux, A. (ed.) EUROCRYPT 2009. LNCS, vol. 5479, pp. 443–461. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-01001-9_26

[UHIM24] Ueno, R., Homma, N., Inoue, A., Minematsu, K.: Fallen sanctuary: a higher-order and leakage-resilient rekeying scheme. IACR Trans. Cryptogr. Hardw. Embed. Syst. **2024**(1), 264–308 (2024)

[UKM+17] Unterluggauer, T., et al.: Leakage bounds for Gaussian side channels. In: Eisenbarth, T., Teglia, Y. (eds.) CARDIS 2017. LNCS, vol. 10728, pp. 88–104. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75208-2_6

[VCMKS12] Veyrat-Charvillon, N., Medwed, M., Kerckhof, S., Standaert, F.-X.: Shuffling against side-channel attacks: a comprehensive study with cautionary note. In: Wang, X., Sako, K. (eds.) ASIACRYPT 2012. LNCS, vol. 7658, pp. 740–757. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-34961-4_44

[Ver15] Verdú, S.: α-mutual information. In: 2015 Information Theory and Applications Workshop, ITA 2015, San Diego, CA, USA, 1–6 February 2015, pp. 1–6. IEEE (2015)

[WZ73] 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)