Information Leakage in Code-based Masking: A Systematic Evaluation by Higher-Order Attacks

Wei Cheng, *Member, IEEE*, Sylvain Guilley, *Senior Member, IEEE*, and Jean-Luc Danger, *Member, IEEE*

***Abstract*—Code-based masking is a recent line of research on masking schemes aiming at provably counteracting side-channel attacks. It generalizes and unifies many masking schemes within a coding-theoretic formalization. In code-based masking schemes, the tuning parameters are the underlying linear codes, whose choice significantly affects the side-channel resilience. In this paper, we investigate the exploitability of the information leakage in code-based masking and present attack-based evaluation results of higher-order optimal distinguisher (HOOD). Particularly, we consider two representative instances of code-based masking, namely inner product masking (IPM) and Shamir's secret sharing (SSS) based masking. Our results do confirm the state-of-the-art theoretical derivatives in an empirical manner with numerically simulated measurements. Specifically, theoretical results are based on quantifying information leakage; we further complete the panorama with attack-based evaluations by investigating the exploitability of the leakage. Moreover, we classify all possible candidates of linear codes in IPM with 2 and 3 shares and $(3, 1)$-SSS based masking, and highlight both optimal and worst codes for them.**

Relying on our empirical evaluations, we therefore recommend investigating the coding-theoretic properties to find the best linear codes in strengthening instances of code-based masking. As for applications, our attack-based evaluation directly empowers designers, by employing optimal linear codes, to enhance the protection of code-based masking. Our framework leverages simulated leakage traces, hence allowing for source code validation or patching in case it is found to be attackable.

***Index Terms*—Side-Channel Attacks, Countermeasures, Leakage Exploitation, Mutual Information, Inner Product Masking, Shamir's Secret Sharing, Code-based Masking, Pre-silicon Simulation-based Evaluation.**

I. Introduction

SIDE-channel analyses (SCAs) are among the most powerful attacks against cryptographic implementations. Since the seminal works [1], [2], a very large amount of SCAs have been proposed by exploiting various observable physical leakages in practice, like power consumption [2], [3], electromagnetic emanations [4], [5], etc. In essential, SCAs attempt to extract the sensitive information from noisy measurements containing unintended emissions or leakages, where the measurements are correlated with internal states or behaviors of a cryptographic device.

Along with a large body of attacks, numerous countermeasures have been proposed to protect practical implementations against SCAs. Relying on different strategies and principles, two major lines of countermeasures are hiding and masking [6]. Specifically, the hiding approach attempts to balance the leakage of different key-dependent operations or data, resulting in less informative signals in side-channel measurements [7], [8]. In contrast, the masking approach randomizes the internal states by randomly splitting internal sensitive variables into several shares, which breaks the straightforward connection between the sensitive variables and the measurements. In particular, the latter is preferable since it is featured with the provable security rather than engineering intuitions of designers. More precisely, the key-recovery attack complexity is demonstrated to increase exponentially with the number of shares provided leakages of different shares are independent of each other and noisy enough [9], [10]. However, this exponential complexity provides only a lower bound (e.g., on the least number of measurements to achieve a successful attack) that is usually loose [11], [12]. Additionally, the lower bound is not able to indicate different side-channel resilience of distinct masking schemes, e.g., to compare different masking instances.

Therefore, more quantitative evaluations of side-channel resilience in masked scenarios play a significant role in understanding the concrete security level and verifying the effectiveness of the protections. In the following, we first revisit existing quantitative approaches and then their applications in assessing and comparing various code-based masking schemes described in the literature.

A. Quantitative Evaluations of Side-Channel Security

According to different leakage models and the abstraction levels of cryptographic implementations, quantitative evaluation strategies can be classified into four categories.

Firstly, the proof-based evaluation intends to prove the side-channel resistance of a *secure by design* masking scheme under abstract models like the probing model [13] and related variants [14], [10], [15], [16]. Typically, under independence assumption and large noise condition, several leakage models are equivalent (up to some numerical constants [16]) in providing formal security guarantees (bounds) of the masked implementation. However, as mentioned before, those bounds are usually quite loose, even in a simple scenario with Hamming weight leakages and independent Gaussian noises [12]. As a consequence, it is recommended to launch more quantitative evaluations in assessing the concrete side-channel security.

Secondly, the information-theoretic evaluation aims at measuring side-channel leakages by utilizing information-theoretic tools [17], [18]. The frequently used measures include Shannon conditional entropy, mutual information (MI), Kullback-

W. Cheng, S. Guilley and J.-L. Danger are with LTCI, Télécom Paris, Institut Polytechnique de Paris, 91120, Palaiseau, France. W. Cheng and S. Guilley are also with Secure-IC S.A.S., 75015, Paris, France. Email: {wei.cheng, jean-luc.danger}@telecom-paris.fr, sylvain.guilley@secure-ic.com.

Leibler (KL) divergence, etc. In fact, this category of evaluation measures the full distribution of leakages and provides insights on how much information an adversary can obtain. In essential, it usually provides information-theoretic bounds on the probability of success for any side-channel distinguishers given a set side-channel measurements [12], [19]. It is worth mentioning that not all distribution-based leakages can be exploited by side-channel distinguishers. For instance, correlation power analysis (CPA) is a typical non-profiling attack and each time it exploits only a few orders of moments of side-channel leakage. However, one of the major difficulties of using information-theoretic evaluation is how to estimate the leakage distribution accurately, for instance, when the number of measurements is not sufficiently enough.

Thirdly, the moment-based evaluation attempts to find the least order of moments of side-channel measurement that depend on the sensitives. Representative metrics including signal-to-noise ratio (SNR) [17] under proper definitions and the normalized inter-class variance (NICV) [20], etc. Particularly, NICV is connected to SNR in the sense that both of them evaluate the key-dependent variance of leakage. With proper definition, SNR can also be used to measure the leakage in presence of higher-order masking schemes.

Finally, the attack-based evaluation is at the core of side-channel security evaluation, which aims at assessing the probability of success of a specific side-channel distinguisher. Relying on large variety of side-channel distinguishers like correlation power analysis [21], mutual information analysis [22], [23], template attacks [24], [25], stochastic attacks [26], [27], higher-order optimal distinguisher (HOOD) [28], [29], etc., the attack-based evaluation provides more straightforward assessment of exploitable leakages. Indeed, those attacks usually provide an accurate number of measurements to achieve a successful attack. However, it is infeasible to exhaust all distinguishers to launch attack-based evaluation provided a limited resources and time.

In summary, the above evaluation strategies provide different levels of quantitative assessment [1]. To a large extent, those strategies are complementary to each other in assessing certain protected constructions or implementations, varying with different evaluation requirements and the necessary expertise in launching evaluations.

In the following, we introduce the code-based masking and review the corresponding security assessment through the above four evaluation strategies.

B. Code-based Masking Scheme and Security Evaluations

Essentially, the rationale of masking is to split the key-dependent variables into several shares and perform independent computations on masked variables only. Many masking schemes have been proposed after the simplest Boolean masking (BM) [9], including multiplicative masking [33], affine masking [34], inner product masking (IPM) [35], direct sum masking (DSM) [36], Shamir's secret sharing (SSS) based masking [37], [38], etc. Recently, code-based masking [39], [40] emerges and it unifies BM, IPM, DSM, SSS-based masking and some variants by a coding-theoretic approach. It employs two linear codes and different settings of the two codes correspond to its various instances.

Notably, the code-based masking shall be configured with *redundancy* to thwart both SCAs and fault injection attacks [38], [39]. The redundancy means that there are more shares that exceed the security threshold of recovering the sensitive variables. For instance, in an (n, t)-SSS based sharing where n is the number of shares and t denotes the security order, then if $n > t + 1$, the sharing is redundant. On the contrary, BM, IPM and DSM themselves are not redundant therefore cannot detect any faults during computations. In particular, an interesting extension of IPM for fault detection is presented in [41] by adding certain redundancy.

From a security perspective, Wang et al. [39] propose an efficient construction of secure gadgets equipped with proof-based evaluation for code-based masking; Cheng et al. [40] present both information-theoretic analysis and moment-based evaluation by using mutual information and signal-to-noise ratio as metrics, respectively; meanwhile, Costes et al. [42] perform an attack-based evaluation on some instances of code-based masking by using maximum likelihood based distinguishers. In particular, the latter two highlight that different linear codes have significant impact on the side-channel resistance of the corresponding code-based masking, while the proof-based evaluation cannot differentiate the impact of the linear codes. Moreover, Cheng et al. [40] also demonstrate how to choose optimal linear codes (tuning parameters) for all instances of code-based masking.

Although Costes et al. [42] show the impact of different codes on side-channel resilience, their attack-based evaluation is not complete yet. First, they identify some instances of SSS-based masking that are equivalent to Boolean masking or quasi-Boolean masking. Those special instances are not recommended for practical applications. However, two natural questions are that, *are there other instances even worse than (quasi-) Boolean instances and how to identify them?* Second, most of attack results in [42] are obtained in a small range of noise levels (e.g., the range of noise variance is $0.05 \leq \sigma^2 \leq 1.0$) [2]. Nevertheless, this range of noise is not sufficient for showing differences between some non-equivalent instances of the codes. At last, as shown by [40], there are a few optimal instances of linear codes from a leakage detection perspective, but those optimal instances are not fully verified by the attack-based evaluation. More generally, the leakage quantification approach proposed in [40] are generic for all instances of code-based masking, then the question is: *how much quantified leakage can be exploited by side-channel distinguishers?*

In view of the above questions, we leverage the attack-based evaluation on code-based masking by using numerically simulated measurements and answer them in a positive and quantitative manner.

[1] We omit in this paper the conformance-based leakage detection (e.g., by using Welch's t-test, χ^2-test, etc) [30], [31], [32] since it usually provides qualitative results.

[2] Only a few instances in [42, Fig. 4] are with a larger range of noise, say $0.25 \leq \sigma^2 \leq 4.0$.

C. Metrics in Attack-based Evaluation

Considering a key-recovery attack in SCA, the ultimate metric is the success rate (SR) indicating the probability that an adversary does succeed in recovering the secret key [17][3]. In particular, two interrelated problems in attack-based evaluation are, on one side, *how many side-channel measurements are needed for a successful attack?* or on the other side, *what is the probability of success given a certain number of measurements?* Therefore, it is preferable to show how the success rate evolves when the number of measurements increases. In a view to be as favorable as possible to the attacker (that is the worst case from a protection perspective), we consider synthetic (i.e., obtained by simulation) traces, which represent an ideal leakage.

Moreover, another attack metric is the guessing entropy (GE) [17], which measures the average rank of the correct key among all candidates based on distinguishing scores after an attack. The GE metric is complementary to success rate as it indicates how wrongly guessed keys behave before a successful attack, and it converges to 1 when the success rate goes to 100% stably in a sound attack.

In this work, we therefore utilize success rate (SR, but also noted P_s to recall it is a probability of success) as the primary metric in a way that it reflects the number of measurements (say q) to achieve a successful attack (e.g., $P_s \geq 95\%$). Implicitly, this metric integrates both SR and GE in attacks.

D. Contributions

In this work, we aim at completing the attack-based evaluation of side-channel resistance of the code-based masking and verifying the coding-theoretic leakage quantification approach by attacks. We highlight that all empirical results are based on numerically simulated measurements as in [45], [42] by considering the Hamming weight leakages with independent additive white Gaussian noises (AWGN). In particular, our contributions are as follows.

a) A complete HOOD-based evaluation of code-based masking: we provide an extensive evaluation on the side-channel resistance of the generalized code-based masking by simulated experiments (with similar settings as in [42]). The attacks are based on the higher-order optimal distinguisher as it is the best attack strategy following the Maximum Likelihood principle. We investigate both IPM and SSS-based masking, since they are representatives of non-redundant and redundant masking schemes, respectively. We highlight that the side-channel resistance of code-based masking is highly related to coding-theoretic properties wherein the dual distance and the (adjusted) kissing number are good indicators as shown in [40] from an information-theoretic perspective. Moreover, we consider a larger range of noises (e.g., σ^2 is up to 8.0) for all linear codes, resulting in a more extensive validation of our selection of optimal codes. Therefore, we verify, by HOOD-based attacks, that the coding-theoretic leakage quantification of code-based masking is sound.

b) Redundancy in code-based masking only decreases side-channel resistance: we leverage on attack-based evaluations to illustrate that the redundancy in sharing can only decrease the side-channel resistance of the corresponding masking schemes. Compared to the state-of-the-arts, our HOOD-based results challenge the evaluation launched in [45], but are in accordance with the ones in [42]. In particular, the authors showed in [45] that exploiting leakages from more shares in a horizontal attack [46] does not always lead to more efficient attacks, whereas we show there are always significant improvements by exploiting leakages from more shares. To verify this, we consider $(2, 1)$ and $(3, 1)$-SSS based masking, and show that exploiting leakages from all three shares always leads to more efficient attacks than that using two shares. Moreover, compared to [42], we extend the state-of-the-arts in two directions: 1) we show the best cases of the linear codes, that are recommended to use, and 2) we give the worst cases of the linear codes that are not recommended for practical applications.

The open sources of this paper are available on `Github` [4], all the data and scripts would allow other researchers to verify and reproduce the coding-theoretic results in this paper.

Outline. The remaining parts of this work are organized as follows. Sec. II recalls different side-channel distinguishers and gives the optimal one against masking. The main attack-based evaluations are in Sec. III and Sec. IV for IPM and SSS-based masking as non-redundant and redundant instances, respectively. The evaluation and discussion of redundancy in code-based masking are presented in Sec. V. Finally the conclusions are given in Sec. VI.

II. Side-channel Distinguishers

We first recall the side-channel distinguishers in unprotected scenarios (without masking, etc). Let $X \in \mathbb{K}$ be the secret variable which depends on the secrets in the cryptographic implementations. For instance, the sensitive variable is usually $X = S(T \oplus K)$, the output of Sbox given a plaintext (or ciphertext) T and a subkey K, e.g., in AES or PRESENT, then we may use $X(k)$ in order to indicate a specific key guess k in generating X.

Considering simulated measurements, we adopt the common scenario in which the intermediate variables leak in Hamming weight model with independent additive white Gaussian noise (AWGN). Therefore, we have $\mathcal{L}^j = w_H(X^j) + N^j$, $1 \leq j \leq q$ for q traces, where w_H denotes the Hamming weight function and $N^j \sim \mathcal{N}(0, \sigma^2)$ is the Gaussian noise with a standard deviation σ. The basic setting of side-channel analysis seen as a communication channel is illustrated in Fig. 1, where $\hat{K}$ is the estimation of the secret key K.

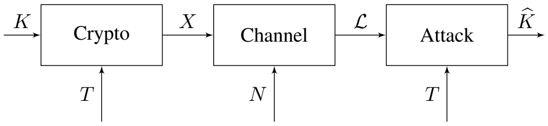

Figure 1. Side-channel seen as a communication channel.

[3] There are different orders of success rates when considering an adversary can launch key enumeration [43], [44] as a post-processing technique after perpetrating individual side-channel attacks. However, we focus on the first-order success rate by convention as it is more straightforward.

[4] https://github.com/Qomo-CHENG/Optimal_code_in_CBM_aes.

A. Different Distinguishers

In SCA, a key-recovery attack intends to extract the secret key from q traces by exploiting certain side-channel distinguishers. In particular, a distinguisher takes maximization over all key hypothesis and gives the most possible candidate(s) by:

$$\widehat{k} = \operatorname*{argmax}_{k \in \mathbb{K}} \widehat{\Delta}(k) = \operatorname*{argmax}_{k \in \mathbb{K}} \widehat{\Delta}(\mathcal{L}, X(k)). \tag{1}$$

Formally, we recall the definition of the side-channel distinguisher as follows.

Definition 1 (Side-Channel Distinguisher [47])**.** *Given a set of side-channel measurements $\mathcal{L}$ and known cryptographic inputs (or outputs) T, a side-channel distinguisher returns a theoretical value*

$$\Delta(k) = \Delta(\mathcal{L}, X(k)) \tag{2}$$

for any key guesses $k \in \mathbb{K}$ and the estimator $\widehat{\Delta}(k)$ converges to $\Delta(k)$ as $q \to \infty$, in the sense that the mean-squared error $\mathbb{E}\left[(\Delta(k) - \widehat{\Delta}(k))^2\right]$ approaches 0 when $q \to \infty$.

Note that we shall simplify $X(k)$ as X by implicitly indicating the link between the sensitive variable and the key hypothesis. In view of Def. 1, several classic side-channel distinguishers are presented as follows:

- Difference of Means (DoM): it is the original distinguisher proposed in the seminal work [2], known as Differential Power Analysis (DPA). Let $f_b(X)$ be the selection function which returns one specific bit of X, then we have

$$\begin{aligned} \Delta(k) &= |\mathbb{E}\left[\mathcal{L}|f_b(X) = 0\right] - \mathbb{E}\left[\mathcal{L}|f_b(X) = 1\right]|, \\ \widehat{\Delta}(k) &= |\frac{\sum_{j=1}^{q}(1 - f_b(X^j))\mathcal{L}^j}{\sum_{j=1}^{q}(1 - f_b(X^j))} - \frac{\sum_{j=1}^{q} f_b(X^j)\mathcal{L}^j}{\sum_{j=1}^{q} f_b(X^j)}|, \end{aligned} \tag{3}$$

  where the absolute value is always considered in maximization for each key hypothesis.
- Correlation Power Analysis (CPA) [21]: in which the distinguisher value is given by computing the Pearson correlation coefficient between the side-channel traces and the hypothetical leakages:

$$\begin{aligned} \Delta(k) &= |\rho(\mathcal{L}, f(X))| = \frac{|\mathrm{Cov}(\mathcal{L}, f(X))|}{\sigma_{\mathcal{L}}\sigma_{f(X)}} \\ &= \frac{|\mathbb{E}\left[\mathcal{L}f(X)\right] - \mathbb{E}\left[\mathcal{L}\right]\mathbb{E}\left[f(X)\right]|}{\sigma_{\mathcal{L}}\sigma_{f(X)}}, \\ \widehat{\Delta}(k) &= \frac{|\widehat{\mathrm{Cov}}(\mathcal{L}, f(X))|}{\widehat{\sigma}_{\mathcal{L}}\widehat{\sigma}_{f(X)}}, \end{aligned} \tag{4}$$

  where $f(\cdot)$ denotes the leakage function, e.g., in the Hamming weight leakage model $f(X) = w_H(X)$. In addition, the covariance is $\widehat{\mathrm{Cov}}(\mathcal{L}, f(X)) = \frac{1}{q}\sum_{j=1}^{q}\mathcal{L}^j f(X^j) - \frac{1}{q}\sum_{j=1}^{q}\mathcal{L}^j \cdot \frac{1}{q}\sum_{j=1}^{q} f(X^j)$, and two estimated variances are $\widehat{\sigma}^2_{\mathcal{L}} = \frac{1}{q}\sum_{j=1}^{q}(\mathcal{L}^j)^2 - (\frac{1}{q}\sum_{j=1}^{q}\mathcal{L}^j)^2$ and $\widehat{\sigma}^2_{f(X)} = \frac{1}{q}\sum_{j=1}^{q}(f(X^j))^2 - (\frac{1}{q}\sum_{j=1}^{q} f(X^j))^2$, respectively. The absolute value is taken for each key hypothesis.
- Mutual Information Analysis (MIA) [22], [48]: the mutual information is used as a metric for assessing the dependency between the side-channel traces and the hypothetical leakages in an information-theoretic sense:

$$\begin{aligned} \Delta(k) &= I(\mathcal{L}; X) = H(\mathcal{L}) - H(\mathcal{L}|X), \\ \widehat{\Delta}(k) &= \sum_{l}\sum_{x}\widehat{\Pr}(l, x)\log_2\frac{\widehat{\Pr}(l, x)}{\widehat{\Pr}(l)\widehat{\Pr}(x)}, \end{aligned} \tag{5}$$

  where I and H denote mutual information and (conditional) entropy, respectively.
- Maximum Likelihood (ML)-based attack [24], [28]: when the leakage distribution is known, the optimal strategy for launching such attack is to use the maximum likelihood rule:

$$\begin{aligned} \Delta(k) &= \Pr(\mathcal{L}|X(k)), \\ \widehat{\Delta}(k) &= \widehat{\Pr}(\mathcal{L}, |X(k)) = \prod_{j=1}^{q}\widehat{\Pr}(\mathcal{L}^j, |X^j(k)), \end{aligned} \tag{6}$$

  where side-channel measurements are assumed to be i.i.d. Therefore, the best key guess is made by:

$$\widehat{k} = \operatorname*{argmax}_{k \in \mathbb{K}} \widehat{\Delta}(k). \tag{7}$$

  Note that the ML rule is equivalent to Maximum a Posterior (MAP) rule with equiprobable keys. It is exactly the case as commonly assumed that K is uniformly distributed in $\mathbb{K}$.

Given a side-channel distinguisher, a primary question arises: whether the attack utilizing the distinguisher will be succeed eventually? Therefore, we define the soundness of distinguishers as follows.

Definition 2 (Soundness of a Distinguisher [17], [47])**.** *A side-channel distinguisher $\widehat{\Delta}(k)$ is said to be sound if the theoretical distinguisher value is maximized at the correct key hypothesis, namely,*

$$\Delta(k^*) > \Delta(k) \qquad \textit{for any } k \neq k^*. \tag{8}$$

Apparently, if a distinguisher is sound, the attack tends to succeed with success rate equal to 100% eventually given enough number of traces (e.g., when $q \to \infty$).

Remark 1. *For above classic distinguishers, CPA is sound [49], so as the DoM, since the latter can be seen as a special case of CPA [47] when $q \to \infty$. MIA is also proved to be sound under Gaussian noise [50], [51]. Moreover, ML-based distinguishers are sound by design, where the correct key guess will rank the first given enough amount of side-channel traces.*

B. Optimal Distinguisher in the Presence of Masking

We focus on code-based masking in this work, which generalizes several existing masking schemes. The communication view in the presence of masking is depicted in Fig. 2. Let $X \in \mathbb{K}$ and $Y \in \mathbb{K}^t$ be respectively the sensitive variable and t random masks. Then the sharing in code-based masking is:

$$Z = X\mathbf{G} + Y\mathbf{H} \in \mathbb{K}^n, \tag{9}$$

given that $t+1 \leq n$, where $\mathbf{G}$ and $\mathbf{H}$ are the generator matrices of two codes $\mathcal{C}$ and $\mathcal{D}$, respectively. As assumed previously, the sensitive variable is $X = S(T \oplus K)$.

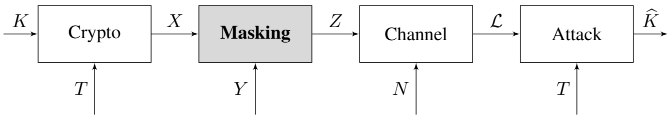

Figure 2. Side-channel seen as a communication channel in the presence of masking.

We first recall the *security order* under *probing model* to depict the side-channel resistance of a masking scheme.

Definition 3 (Security Order under Probing Model [13], [10])**.** *A protected cryptographic implementation is said to have a security order of t if any adversary who probes (or spies) up to t intermediate values of the computation fails to reveal any information on the sensitive variable.*

The security order under probing model, or probing model security is a fundamental property of masking schemes. Consider Boolean masking with n shares as an example, it can have the highest security order $t = n - 1$ if it is properly implemented [13].

In code-based masking, the two good indicators of its side-channel resilience are the dual distance $d_{\mathcal{D}}^{\perp}$ [36], [52], [53] and the kissing number $B_{d_{\mathcal{D}}^{\perp}}$ [53] (or the adjusted kissing number $B'_{d_{\mathcal{D}}^{\perp}}$ in redundant cases [40]). We recall their definitions:

Definition 4 (Dual Distance [54] and Kissing Number [55])**.** *Considering a linear code $\mathcal{C}$, its dual distance $d_{\mathcal{C}}^{\perp}$ is the minimum Hamming weight $w_H(u)$ of nonzero $u \in \mathbb{K}^n$, such that $\sum_{c\in\mathcal{C}}(-1)^{c\cdot u} \neq 0$. Accordingly, the kissing number B_d is the number of codewords in $\mathcal{C}$ at minimum distance d to any codewords, or equivalently: $B_d = |\{x \in \mathcal{C} \mid w_H(x) = d\}|$.*

Definition 5 (Adjusted Kissing Number [40])**.** *Let $\mathcal{C}$, $\mathcal{D}$ denote two linear codes, their adjusted kissing number B'_d is:*

$$\begin{aligned} B'_d = |\{(x, y) \in (\mathcal{D}\backslash\mathcal{C})^2 \mid x + y \in \mathcal{C}, \\ w_H(x) = w_H(y) = d\}|. \end{aligned} \tag{10}$$

Remark 2. *It is worth mentioning that the kissing number $B_{d_{\mathcal{D}}^{\perp}}$ is defined on the dual code $\mathcal{D}^{\perp}$, while the adjusted kissing number $B'_{d_{\mathcal{D}}^{\perp}}$ is on $\mathcal{C}^{\perp}$ and $\mathcal{D}^{\perp}$ in code-based masking. The latter is degraded to the former in non-redundant cases where we shall have $\mathcal{C}^{\perp} \cap \mathcal{D}^{\perp} = \{0\}$, e.g., in IPM and DSM. Furthermore, for the sake of simplicity, we use the same notations when the linear codes are expanded into the basefield $\mathbb{F}_2$ by using subfield representations [40].*

Remark 3. *As formally demonstrated in [52], [53], the security order of code-based masking schemes under probing model equals to $d_{\mathcal{D}}^{\perp} - 1$. Indeed, the dual distance of a code corresponds to the minimum number of linearly dependent coordinates in the code [54]. Therefore, the maximal number of linearly independent coordinates is $d_{\mathcal{D}}^{\perp} - 1$, corresponding to the security order in the probing model.*

Regarding the kissing number, it counts the number of codewords of Hamming weight equal to the dual distance in $\mathcal{D}^{\perp}$. In other words, it counts the frequency that $d_{\mathcal{D}}^{\perp}$ probes can obtain certain information about the sensitive variable. As verified in [53], [40], the mutual information between the sensitive variable and the noisy leakage is asymptotically linear with $B_{d_{\mathcal{D}}^{\perp}}$ when the Gaussian noise is high enough.

In particular, the bit-probing model [52], [53] corresponds to expand the linear codes into $\mathbb{F}_2$. As a result, the dual distance and the (adjusted) kissing number are also calculated in the basefield. Complete details can also be found in [56].

C. Simulation Settings and HOOD

As shown in Fig. 2, let $T \sim \mathcal{U}(\mathbb{K})$ denote plaintext or ciphertext uniformly drawn in $\mathbb{K}$, $K \sim \mathcal{U}(\mathbb{K})$ denote the secret key and $Y \sim \mathcal{U}(\mathbb{K}^t)$ be t random masks. The sensitive variable is $X = S(T \oplus K)$, which is the output of AES Sbox through this paper and $\mathbb{K} = \mathbb{F}_{2^8}$. Therefore, in the presence of code-based masking, we have $Z = (Z_1, Z_2, \ldots, Z_n) = X\mathbf{G} + Y\mathbf{H}$ for different initializations of $\mathbf{G}$ and $\mathbf{H}$ in specific instances.

Regarding the simulated measurements, we utilize the Hamming weight model with independent AWGN. For each share Z_i, we have $\mathcal{L}_i = w_H(Z_i) + N_i$, $1 \leq i \leq n$ for n shares and $N_i \sim \mathcal{N}(0, \sigma^2)$ is Gaussian noise. Given a dataset of q traces, we further denote all traces as $\mathcal{L} = (\mathcal{L}_i^j)$ for $1 \leq i \leq n$ and $1 \leq j \leq q$.

In our scenario, as the leakage model is assumed to be known, the best strategy for performing key-recovery attacks is to utilize the ML-based approach. Following the principle of ML-based attack, the higher-order optimal distinguisher (HOOD) is known as follows.

Lemma 1 (Higher-Order Optimal Distinguisher [29])**.** *Given a set of q measurements $\mathcal{L} = (\mathcal{L}_i^j) = f(Z_i^j) + N_i^j$ for $1 \leq i \leq n$ and $1 \leq j \leq q$ such that N_i^j are i.i.d. across $1 \leq j \leq q$ and independent across $1 \leq i \leq n$. When the leakage distribution is known (both the leakage function and the noise distribution), the d-th order optimal distinguisher is:*

$$\Delta(k) = \prod_{j=1}^{q} \sum_{y\in\mathbb{K}^t} \Pr(Y = y) \prod_{i=1}^{d} \Pr(\mathcal{L}_i^j | Z_i^j), \tag{11}$$

where the calculation of Z_i^j implicitly involves $Y = y$. Therefore, the key hypothesis is given by

$$\widehat{k} = \underset{k\in\mathbb{K}}{\operatorname{argmax}}\, \Delta(k). \tag{12}$$

In the sequel, we focus on attack-based evaluation of the code-based masking, particularly we target IPM with $n = 2$ and $n = 3$, and $(3, 1)$-SSS based masking.

III. Attacks against Non-Redundant Code-based Masking

Considering IPM as an instance of non-redundant code-based making, the generating matrices of $\mathcal{C}$ and $\mathcal{D}$ are:

$$\begin{aligned} \mathbf{G} &= \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \end{pmatrix} \\ \mathbf{H} &= \begin{pmatrix} \alpha_1 & 1 & 0 & \cdots & 0 \\ \alpha_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_t & 0 & 0 & \cdots & 1 \end{pmatrix} \end{aligned} \tag{13}$$

where $n = t + 1$ and $\alpha_i \in \mathbb{K}\backslash\{0\}$ for $1 \leq i \leq t$. In particular, by taking $\alpha_i = 1$ for $1 \leq i \leq t$ recovers the

Boolean masking. As a result, the generator matrix of $\mathcal{D}^\perp$ is: $\mathbf{H}^\perp = (1\ \alpha_1\ \alpha_2\ \cdots\ \alpha_t)$ with $d_\mathcal{D}^\perp = t+1$, indicating that IPM with n shares has a security order equal to $n-1$ under word-probing model [52], [57]. We denote $\alpha = (1, \alpha_1, \ldots, \alpha_t)$ the public parameters in IPM.

A. Optimal Distinguishers

Relying on Lemma 1, the HOOD is instantiated in the context of Hamming weight leakage with an AWGN as:

$$\begin{aligned}\Delta(k) &= \prod_{j=1}^{q} \sum_{y \in \mathbb{K}^{n-1}} \Pr(Y=y) \prod_{i=1}^{d} \Pr(\mathcal{L}_i^j | z_i^j) \\ &= \prod_{j=1}^{q} \sum_{y \in \mathbb{K}^{n-1}} \Pr(Y=y) \prod_{i=1}^{d} \mathcal{N}(\mathcal{L}_i^j | w_H(z_i^j), \sigma^2).\end{aligned} \tag{14}$$

Since Y is uniformly distributed (say $\Pr(Y=y) = \frac{1}{|\mathbb{K}^{n-1}|}$) required by a sound masking scheme, it is independent of each key hypothesis and hence has no impact on $\Delta(k)$. Taking logarithms further eases the numerical computations (avoiding float overflows), the HOOD is equivalent to the following distinguisher score [45], [42]:

$$S(k) = \sum_{j=1}^{q} \log \sum_{y \in \mathbb{K}^{n-1}} \prod_{i=1}^{d} \mathcal{N}(\mathcal{L}_i^j | w_H(z_i^j), \sigma^2), \tag{15}$$

then the key guess is determined by maximizing $S(k)$.

Formally, thanks to masking, an adversary cannot obtain anything about the sensitive variable if the order d of a HOOD is not strictly greater than the security order t. A prerequisite for launching a successful attack is $d > t$ in our scenario when targeting IPM, which is consistent with coding-theoretic conditions [40].

B. IPM with $n = 2$

Taking $n = 2$ gives $t = 1$, resulting that only one parameter in IPM is α_1 and $\mathbf{H} = (\alpha_1\ 1)$. There are 255 candidates for α_1 as it cannot be zero. In order to facilitate practical applications and a fair comparison with the state-of-the-art, we aim at the irreducible polynomial $g(\mathsf{X}) = \mathsf{X}^8 + \mathsf{X}^4 + \mathsf{X}^3 + \mathsf{X} + 1$ that is used in AES [5] to generate the finite field $\mathbb{K} = \mathbb{F}_{2^8}$.

As shown in [53], the two coding-theoretic properties that indicate the side-channel resistance of IPM are the dual distance $d_\mathcal{D}^\perp$ and the kissing number $B_{d_\mathcal{D}^\perp}$, and the optimal codes are those with the maximized $d_\mathcal{D}^\perp$ and the minimized $B_{d_\mathcal{D}^\perp}$. Herein, we first investigate the statistical properties of $d_\mathcal{D}^\perp$ and $B_{d_\mathcal{D}^\perp}$ among all linear code candidates. The distribution of $d_\mathcal{D}^\perp$ are enumerated in Tab. I and the corresponding choices of the codes with given $d_\mathcal{D}^\perp$ and $B_{d_\mathcal{D}^\perp}$ are in Tab. II, while in the latter we are only interested in the linear codes with the maximal and minimal values of $B_{d_\mathcal{D}^\perp}$ for each $d_\mathcal{D}^\perp$.

As shown in Tab. II, there are only 12 optimal linear codes which maximize $d_\mathcal{D}^\perp$ and minimize $B_{d_\mathcal{D}^\perp}$ at the same time.

Table I
DISTRIBUTION OF $d_\mathcal{D}^\perp$ FOR IPM WITH $n = 2$. NOTE THAT THE RATIO IS THE PERCENTAGE OF THE LINEAR CODES IN ALL CANDIDATES.

$d_\mathcal{D}^\perp = d$	$\|\{\alpha_1\}\|$ (ratio)	$\max\{B_d\}$	$\min\{B_d\}$
$d = 2$	35 (0.1373)	8	1
$d = 3$	146 (0.5725)	6	1
$d = 4$	74 (0.2902)	17	4

C. Experimental Results on 2-Share IPM

As mentioned previously, the simulated traces are $\mathcal{L} = (\mathcal{L}_i^j)$ for $1 \leq i \leq n$ and $1 \leq j \leq q$ where $\mathcal{L}_i^j = w_H(Z_i^j) + N_i^j$ denotes the leakage of i-th share in j-th trace. The evaluation metric is the minimum number of traces achieving $P_s \geq 95\%$, which varies along with different noise levels.

For linear codes of different $d_\mathcal{D}^\perp$ shown in Tab. II, we choose both the minimum and the maximum of $B_{d_\mathcal{D}^\perp}$ excluding the Boolean one. The evaluation results of IPM with $n = 2$ are shown in Fig. 3 by using up to $q = 100,000$ traces. Moreover, we include Boolean masking (BM) with $n = 2$ and $n = 3$ shares in comparison.

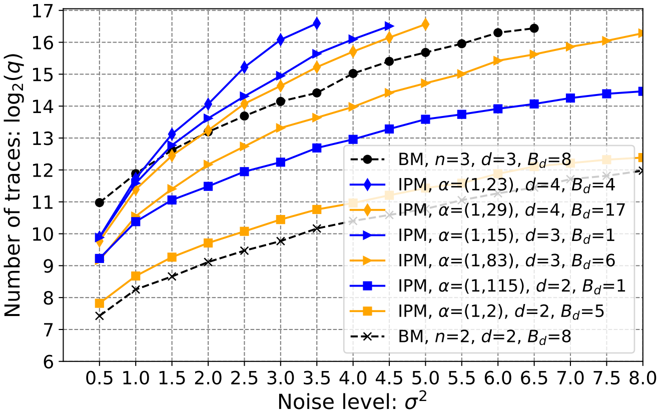

Figure 3. Attack-based evaluation of IPM with $n = 2$ shares. Taking two codes in each group with different $d_\mathcal{D}^\perp$ and/or $B_{d_\mathcal{D}^\perp}$.

The main takeaway point from Fig. 3 is that, IPM with the linear code of the maximized dual distance $d_\mathcal{D}^\perp$ and the minimized kissing number $B_{d_\mathcal{D}^\perp}$ indeed has the best achievable side-channel resistance. The attack-based evaluation also confirms: 1) all 2-share IPM are better than the first-order Boolean masking (with $n = 2$); 2) good choices of linear codes of 2-share IPM can even be better than the second-order Boolean masking (with $n = 3$) when the noise level is $\sigma^2 > 1.0$. The reason is that in IPM, the best cases of $d_\mathcal{D}^\perp$ is larger and $B_{d_\mathcal{D}^\perp}$ is smaller than that in the second-order Boolean masking, respectively; 3) it is also advantageous to adopt 2-share IPM rather than 3-share BM from a performance perspective. For instance, the clock cycles are $157,196$ vs $160,357$ as reported in [57] for an AES-128 implementations on an AVR architecture protected by the former and the latter, respectively.

a) Optimal Codes for 2-Share IPM: According to Tab. II, there are only 12 optimal codes with the best coding-theoretic properties. For the sake of brevity, we present four cases of optimal codes as in Fig. 4. The primary observation is that

[5] As a bonus, those linear codes in this work can be applied into AES straightforwardly. Moreover, we lay a common baseline for further comparison with the state-of-the-art linear codes in [35], [57], [52], [42].

Table II
CHOICES OF THE CODES FOR IPM WITH $n = 2$.

$d_{\mathcal{D}}^{\perp} = d$	B_d	$\lvert\{\alpha_1\}\rvert$	Candidates of α_1	Comments
$d = 2$	$B_d = 8$	1	$\{1\}$	Boolean masking
	$B_d = 1$	20	$\{16, 17, 34, 39, 60, 90, 115, 116, 119, 120, 133, 140, 180, 182, 201, 207, 215, 230, 234, 247\}$	One instance $\alpha = (1, 17)$ is reported in [57]
$d = 3$	$B_d = 6$	8	$\{3, 83, 101, 137, 158, 166, 202, 246\}$	
	$B_d = 1$	58	$\{14, 15, 19, 20, 40, 44, 48, 49, 52, 56, 61, 67$ $69, 75, 76, 80, 84, 94, 97, 99, 103, 112, 113, \ldots\}$	Three instances in [42] belong to this class, see details in Tab. VII
$d = 4$	$B_d = 17$	2	$\{29, 64\}$	$29 \overset{\text{equiv}}{\approx} 64$ [6]
	$B_d = 4$	12	$\{23, 46, 51, 54, 81, 92, 95, 102, 108, 162, 165, 184\}$	12 optimal codes in total. An optimal instance $(175, 237)$ given in [42] is equivalent to $(1, 54)$

those four codes have similar side-channel resistance from an adversary perspective who launches HOOD-based attacks. Note that the fluctuations among those four curves are due to the nature of numerical simulation with certain random seeds. Overall, those four codes perform closely against HOOD-based attacks.

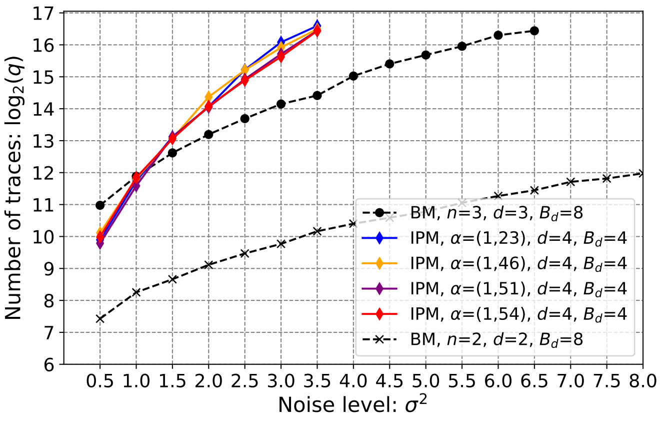

Figure 4. Comparison of four instances of optimal codes for 2-share IPM, according to the best coding-theoretic properties given by $d_{\mathcal{D}}^{\perp} = 4$ and $B_{d_{\mathcal{D}}^{\perp}} = 4$. Note that the number of traces is shown only in $\log_2$ scale for the sake of brevity.

We assume that the leakage distribution is known when launching such attacks, this scenario allows a worst-case evaluation of side-channel resistance of IPM. Furthermore, when other distinguishers are adopted in carrying out attacks, ML-based analysis also provides an upper bound on the success rate. In summary, our worst-case evaluation provides insights on how successful can an attack be in practice[7] and shows how to select optimal codes when applying IPM.

b) The Impact of $B_{d_{\mathcal{D}}^{\perp}}$: We have showed how to select optimal codes according to both $d_{\mathcal{D}}^{\perp}$ and $B_{d_{\mathcal{D}}^{\perp}}$, yet the solo role of $B_{d_{\mathcal{D}}^{\perp}}$ is not explicitly investigated. In the following, we compare several instances of the linear code in IPM with the same $d_{\mathcal{D}}^{\perp}$ while different $B_{d_{\mathcal{D}}^{\perp}}$.

As shown in Fig. 5, we set $d_{\mathcal{D}}^{\perp} = 2$ and $B_{d_{\mathcal{D}}^{\perp}} \in \{1, 2, 3, 5, 8\}$ where 2-share BM being a special case of IPM has $B_{d_{\mathcal{D}}^{\perp}} = 8$. Apparently, reducing $B_{d_{\mathcal{D}}^{\perp}}$ leads to a more difficult attack in the sense of the necessary number of traces to successfully extract the correct key. In addition, since we choose $d_{\mathcal{D}}^{\perp} = 2$, all those codes will not outperform 3-share BM in full range of noise levels.

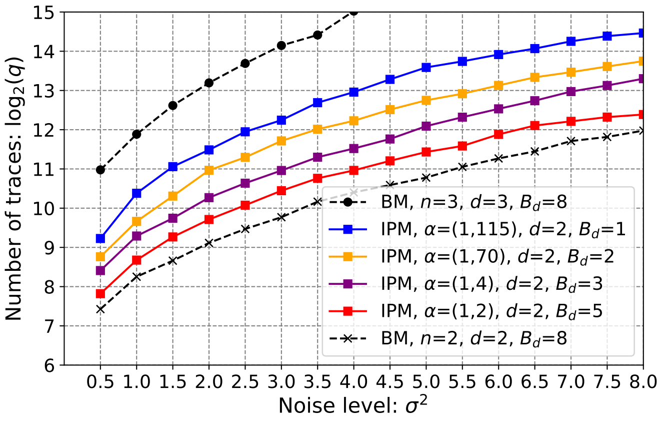

Figure 5. Illustrating the impact of $B_{d_{\mathcal{D}}^{\perp}}$ given the same $d_{\mathcal{D}}^{\perp}$ in 2-share IPM. Note that 2-share IPM instances have different kissing number $B_{d_{\mathcal{D}}^{\perp}}$.

Summing up, we demonstrate that $B_{d_{\mathcal{D}}^{\perp}}$ plays a significant role in indicating the side-channel resistance of IPM. More generally, it is integrated with the dual distance as indicators in evaluating side-channel security of non-redundant code-based masking like DSM, non-redundant SSS-based masking, etc.

D. Linear Codes for IPM with $n = 3$

Herein we present the classification of the linear codes of 3-share IPM. Taking $n = 3$, resulting that $t = 2$ and two free parameters in $\alpha = (1, \alpha_1, \alpha_2)$ are $\alpha_1, \alpha_2 \in \mathbb{K}\backslash\{0\}$. There are $255 \times 255 = 65,025$ candidates, where the number of candidates can be dramatically reduced by considering the equivalence of the linear codes [40]. Therefore, we take $\alpha_1 \leq \alpha_2$, which reduces the number of the codes to $32,640$.

The distribution of $d_{\mathcal{D}}^{\perp}$ are enumerated in Tab. III and the choices of the codes under given $d_{\mathcal{D}}^{\perp}$ and $B_{d_{\mathcal{D}}^{\perp}}$ are in Tab. IV. Note again that in Tab. IV we only focus on the maximal and minimal values of $B_{d_{\mathcal{D}}^{\perp}}$.

As shown in Tab. IV, there are only 3 optimal codes which maximize $d_{\mathcal{D}}^{\perp}$ and minimize $B_{d_{\mathcal{D}}^{\perp}}$ at the same time. Since the maximized dual distance is $d_{\mathcal{D}}^{\perp} = 8$, IPM with those optimal

[7] We precise that, here and in the sequel, "in practice" means when applied on synthetic traces consisting in Hamming weight leakage with AWGN noise, also when the leakage model of a real-world device is largely linear.

[7] We use $\overset{\text{equiv}}{\approx}$ to denote that two linear codes with the given parameters are equivalent over $\mathbb{F}_{2^\ell}$ or after the sub-field representation in $\mathbb{F}_2$, see [53], [55] for details.

Table III
DISTRIBUTION OF $d_{\mathcal{D}}^{\perp}$ FOR IPM WITH $n = 3$.

$d_{\mathcal{D}}^{\perp} = d$	$\lvert\{(\alpha_1, \alpha_2)\}\rvert$ (ratio)	$\max\{B_d\}$	$\min\{B_d\}$
$d = 3$	207 (0.0063)	8	1
$d = 4$	1730 (0.0530)	6	1
$d = 5$	7242 (0.2219)	7	1
$d = 6$	15304 (0.4689)	13	1
$d = 7$	7929 (0.2429)	12	1
$d = 8$	228 (0.0070)	20	6

codes should be comparable with the eighth-order BM (under the bit-probing model), namely $n = 8$ given certain levels of noise. Particularly, considering the security order in the bit-probing model [52], [53], the former and the latter share the same security order $t_b = d_{\mathcal{D}}^{\perp} - 1 = 7$. Therefore, it is recommended to apply IPM rather than BM with many more shares since as a rule of thumb, the implementation cost usually increases at least quadratically with n.

Remark 4. *Note that α_1, α_2 in $\alpha = (1, \alpha_1, \alpha_2)$ are interchangeable because of the equivalence of the linear codes. Therefore, other optimal codes shall be obtained easily.*

IV. ATTACKS ON REDUNDANT CODE-BASED MASKING

As a general rule, redundancy is indispensable for detecting faults in computations and operations. Code-based masking can be configured in a redundant way [39], [42] to thwart both side-channel analysis and fault injection attacks. Since IPM itself is not redundant, in the sequel, we focus on an instance of redundant code-based masking, namely SSS-based (polynomial) masking [37], [38] and perform HOOD-based evaluations on it. We also consider IPM-FD, proposed in [41], as a special case of redundant masking schemes and compared it with SSS-based masking in this section. We will show that although IPM remains more robust with the same number of shares n, it offers no protection against fault injection attacks as SSS-based masking and IPM-FD do.

Taking SSS-based masking as an example of redundant code-based masking, the parameters are denoted as $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ and the condition for α_i is that $\alpha_i \neq \alpha_j$ for any $i \neq j$. Then we have the following generator matrices [40] for the codes $\mathcal{C}$ and $\mathcal{D}$, respectively,

$$\mathbf{G} = \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix}$$
$$\mathbf{H} = \begin{pmatrix} \alpha_1^1 & \alpha_2^1 & \cdots & \alpha_n^1 \\ \alpha_1^2 & \alpha_2^2 & \cdots & \alpha_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1^t & \alpha_2^t & \cdots & \alpha_n^t \end{pmatrix} \tag{16}$$

where α_i for $1 \leq i \leq n$ are also called public points in SSS-based masking. The corresponding scheme is also denoted as (n, t)-SSS based masking.

From a coding-theoretic perspective, the SSS scheme is connected to the Reed-Solomon (RS) code [45]. Given the two generator matrices as in Eqn. 16, the rank of $\mathbf{H}$ equals t, so the dual distance of $\mathcal{D}$ is $t + 1$ [40]. Accordingly, the side-channel security order in the word-probing model is $t_w = t$, and in redundant cases, the optimal codes are those with the maximized dual distance $d_{\mathcal{D}}^{\perp}$ and the minimized adjusted kissing number $B'_{d_{\mathcal{D}}^{\perp}}$; while the worst codes are those with the minimized $d_{\mathcal{D}}^{\perp}$ and the maximized $B'_{d_{\mathcal{D}}^{\perp}}$.

A. Optimal Distinguishers

Recall the form of $\mathbf{H}$ in Eqn. 16 that, there are n public points to be determined in SSS-based masking. However, the masking itself is t-th order secure.

Similarly as in IPM, the optimal distinguisher is determined by applying the ML rule. Considering the same assumption on leakage distribution, we have:

$$\begin{aligned} \Delta(k) &= \prod_{j=1}^{q} \sum_{y \in \mathbb{K}^t} \Pr(Y = y) \prod_{i=1}^{d} \Pr(\mathcal{L}_i^j | z_i^j) \\ &= \prod_{j=1}^{q} \sum_{y \in \mathbb{K}^t} \Pr(Y = y) \prod_{i=1}^{d} \mathcal{N}(\mathcal{L}_i^j | w_H(z_i^j), \sigma^2). \end{aligned} \tag{17}$$

Taking logarithms to ease the numerical computations, the HOOD is therefore equivalent to the following distinguisher score [42]:

$$S(k) = \sum_{j=1}^{q} \log \sum_{y \in \mathbb{K}^t} \prod_{i=1}^{d} \mathcal{N}(\mathcal{L}_i^j | w_H(z_i^j), \sigma^2). \tag{18}$$

As mentioned in Sec. III-A, a prerequisite for a successful attack is $d > t$ when targeting SSS-based masking.

Remark 5. *It is worth mentioning that the distinguisher proposed in [45, Eqn. 13] is problematic. The reason is that, the summation within the logarithm is over $y \in \mathbb{K}^t$ rather than over $y \in \mathbb{K}^{n-1}$ when $n > t + 1$, namely in redundant cases. In fact, their results would match with ours if a correct formula for HOOD, e.g., Eqn. 18, is used instead.*

B. HOOD against $(3, 1)$-SSS based Masking

Considering $n = 3$ and $t = 1$, the generator matrices $\mathbf{G}$ and $\mathbf{H}$ are as follows.

$$\begin{aligned} \mathbf{G} &= \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \\ \mathbf{H} &= \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 \end{pmatrix}, \end{aligned} \tag{19}$$

where α_i for $1 \leq i \leq 3$ are not equal to each other. We can fix $\alpha_1 = 1$ by utilizing the equivalence of the linear codes. Additionally, we set $\alpha_2 < \alpha_3$ as in [40] and resulting that there are $32,131$ candidates (instead of $2,731,135$ codes for any pairwise different α_1, α_2 and α_3).

The distribution of $d_{\mathcal{D}}^{\perp}$ are exhausted in Tab. V and the choices of the codes under given $d_{\mathcal{D}}^{\perp}$ and $B_{d_{\mathcal{D}}^{\perp}}$ are in Tab. VI, by focusing on the maximal and minimal values of $B_{d_{\mathcal{D}}^{\perp}}$.

Remark 6. *In SSS-based masking, we should use the adjusted kissing number $B'_{d_{\mathcal{D}}^{\perp}}$ instead of $B_{d_{\mathcal{D}}^{\perp}}$. Typically, we have $B'_{d_{\mathcal{D}}^{\perp}} \geq B_{d_{\mathcal{D}}^{\perp}}$ in SSS-based masking as pointed out in [40]. However, we use $B_{d_{\mathcal{D}}^{\perp}}$ here since it follows the same trend as $B'_{d_{\mathcal{D}}^{\perp}}$. Note that given a specific $\mathcal{C}$, different choices of $\mathcal{D}$ with the same $B_{d_{\mathcal{D}}^{\perp}}$ may lead to different $B'_{d_{\mathcal{D}}^{\perp}}$.*

Table IV
CHOICES OF THE CODES FOR IPM WITH $n = 3$.

$d_{\mathcal{D}}^{\perp} = d$	B_d	$\lvert\{(\alpha_1, \alpha_2)\}\rvert$	Candidates of (α_1, α_2)	Comments
$d = 3$	$B_d = 8$	1	$\{(1,1)\}$	Boolean masking
	$B_d = 1$	151	$\{(1,16),(1,17),(1,34),(1,39),(1,60),(1,90),(1,115),(1,116),\ldots\}$	
$d = 4$	$B_d = 6$	3	$\{(2,3),(140,141),(246,247)\}$	
	$B_d = 1$	1227	$\{(1,14),(1,18),(1,19),(1,20),(1,21),(1,30),(1,41),(1,42),\ldots\}$	
$d = 5$	$B_d = 7$	8	$\{(1,176),(2,164),(5,143),(8,64),(8,232),(12,12),(29,232),(82,141)\}$	
	$B_d = 1$	4586	$\{(1,23),(1,31),(1,46),(1,47),(1,75),(1,77),(1,98),(1,107),\ldots\}$	One instance $(15, 233)$ is in [35]
$d = 6$	$B_d = 13$	2	$\{(1,130),(127,127)\}$	
	$B_d = 1$	7050	$\{(2,184),(3,45),(3,46),(3,47),(3,59),(3,65),(3,77),(3,81),\ldots\}$	
$d = 7$	$B_d = 12$	3	$\{(16,185),(56,142),(116,242)\}$	
	$B_d = 1$	645	$\{(3,53),(7,45),(7,49),(7,77),(7,99),(7,106),(7,107),(9,154),\ldots\}$	
$d = 8$	$B_d = 20$	3	$\{(94,109),(97,124),(147,161)\}$	
	$B_d = 6$	3	$\{(27,196),(91,204),(218,240)\}$	Only three codes are optimal

Table V
DISTRIBUTION OF $d_{\mathcal{D}}^{\perp}$ FOR $(3, 1)$-SSS BASED MASKING.

$d_{\mathcal{D}}^{\perp} = d$	$\lvert\{(\alpha_2, \alpha_3)\}\rvert$ (ratio)	$\max\{B_d\}$	$\min\{B_d\}$
$d = 2$	11460 (0.3567)	13	1
$d = 3$	20581 (0.6405)	19	1
$d = 4$	90 (0.0028)	73	37

C. Experimental Results

With the same setting as in evaluation of IPM, the simulated traces are $\mathcal{L} = (\mathcal{L}_i^j)$ for $1 \leq i \leq n$ and $1 \leq j \leq q$ where $\mathcal{L}_i^j = w_H(Z_i^j) + N_i^j$ denotes the leakage of i-th share in j-th trace. The sharing employs two codes generated by Eqn. 19.

For linear codes of different $d_{\mathcal{D}}^{\perp}$ shown in Tab. VI, we choose both the minimum and the maximum of $B_{d_{\mathcal{D}}^{\perp}}$. The evaluation results of $(3, 1)$-SSS based masking are shown in Fig. 6 by using up to $q = 100,000$ traces. Moreover, we also include Boolean masking with $n = 2$ and $n = 3$ shares in comparison.

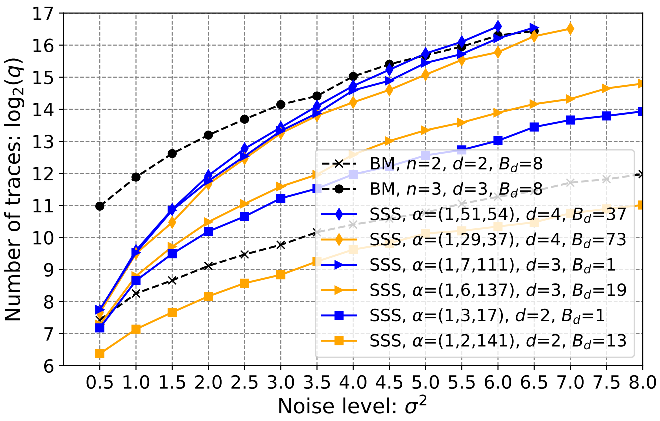

Figure 6. Attack-based evaluation of $(3, 1)$-SSS based masking. Taking two codes in each group with different $d_{\mathcal{D}}^{\perp}$ and/or $B_{d_{\mathcal{D}}^{\perp}}$.

From Fig. 6, the most important takeaway point is that the public points in $(3, 1)$-SSS based masking make a significant difference in side-channel resistance of the corresponding masking scheme. The fundamental reason is that in HOOD, the least orders of secret-dependent statistics of leakage dominate the amount of information that can be extracted from measurements. In this respect, different public points in SSS-based masking correspond to the linear codes with various dual distances and (adjusted) kissing numbers, that lead to distinct least orders of secret-dependent information (as explained in Remark 3). Note that another verification of the amount of exploitable leakage from a information-theoretic perspective can be found in [53], [40].

Furthermore, we can observe from Fig. 6 that: 1) with a dedicated selection of good linear codes, the side-channel resistance of the scheme can be improved significantly; 2) comparing with the attack-based evaluation on 2-share IPM, the side-channel security of $(3, 1)$-SSS based masking is degraded because of the redundancy, which is consistent with the information-theoretic evaluation in [40]; 3) similarly as in 2-share IPM, the best codes can provide comparable security level as 3-share BM when the noise level is higher enough (e.g., $\sigma^2 \geq 5.0$); 4) for the first time, we show that with bad choices of the code, the security level of $(3, 1)$-SSS based masking can be continuously lower than 2-share BM.

In the following, we further leverage the last two points by providing more instances of the optimal and the worst codes for $(3, 1)$-SSS based masking, respectively.

a) Optimal Codes for $(3, 1)$-SSS based Masking: According to Tab. VI, there are only three cases of optimal codes. The evaluation results are depicted in Fig. 7, showing that those codes lead to very close side-channel resistance.

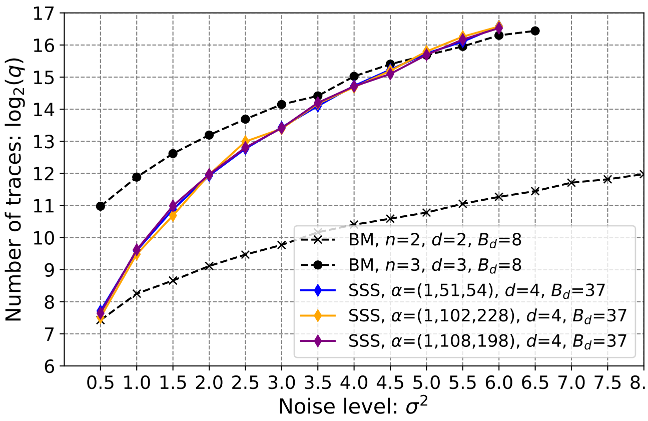

Figure 7. The optimal codes for $(3, 1)$-SSS based masking, in which $d_{\mathcal{D}}^{\perp}$ is maximized and $B_{d_{\mathcal{D}}^{\perp}}$ is minimized given a specific $d_{\mathcal{D}}^{\perp} = 4$.

To sum up, the optimal choices of public points in SSS-

Table VI
CHOICES OF THE CODES FOR $(3,1)$-SSS.

$d_{\mathcal{D}}^{\perp}=d$	B_d	$\lvert\{(\alpha_2,\alpha_3)\}\rvert$	Candidates of (α_2,α_3)	Comments
$d=2$	$B_d=13$	3	$\{(2,4),(2,141),(141,203)\}$	The worst cases
	$B_d=1$	5976	$\{(3,17),(3,34),(3,37),(3,39),(3,48),(3,49),(3,51),(5,60),\dots\}$	
$d=3$	$B_d=19$	3	$\{(6,137),(71,123),(105,158)\}$	
	$B_d=1$	435	$\{(7,23),(7,53),(7,111),(7,148),(7,198),(11,84),(11,94),(11,154),\dots\}$	An instance $(5,221,198)$ is in [42]
$d=4$	$B_d=73$	3	$\{(29,37),(64,131),(77,128)\}$	
	$B_d=37$	3	$\{(51,54),(102,228),(108,198)\}$	Only three codes are optimal

based masking can significantly improve its side-channel resistance that is much higher than 2-share BM. In particular, those optimal codes with the first-order security ($t=1$) can even provide comparable security as 3-share BM where $t=2$.

b) Worst Codes for $(3,1)$-SSS based Masking: From Tab. VI, there are several classes of the linear codes that are worse than 2-share BM, including the three worst cases. The evaluation results are plotted in Fig. 8. Interestingly, those worst codes make the SSS-based masking perform worse than BM in full range of noise levels.

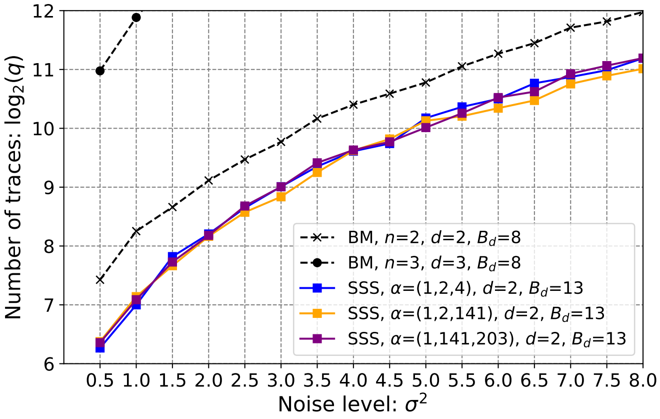

Figure 8. The worst codes for $(3,1)$-SSS based masking, where $d_{\mathcal{D}}^{\perp}$ is minimized and $B_{d_{\mathcal{D}}^{\perp}}$ is maximized given a specific $d_{\mathcal{D}}^{\perp}=2$.

To the best of our knowledge, we identify, for the first time, the worst cases of public points in SSS-based masking or more generally in the context of secret sharing schemes, when each share leaks certain noisy information. Comparing with the state-of-the-art [42], we confirm that the coding-theoretic approach proposed in [40] not only provides the optimal cases, but also identifies the worst cases of the public points in SSS-based masking. Both of them are instructive in designing redundant code-based masking in protecting cryptographic implementations in practice.

V. COMPARISONS: HOW REDUNDANCY MATTERS?

As shown in [40], the redundancy in code-based masking gives rise to more leakage from an information-theoretic sense when assessed by mutual information. However, more leakage detected by mutual information is not always exploited by side-channel distinguishers. Therefore, it is necessary to investigate from an attacking perspective that: *how does the redundancy impact the performance of distinguishers?*

A. Impact of Redundancy in SSS-based Masking

In this section, we demonstrate from an attack-based evaluation that, adding redundancy in code-based masking can only reduce the side-channel resistance of the corresponding masking scheme. To have a fair comparison, we consider two examples of $(3,1)$-SSS based masking. Specifically, in $(3,1)$-SSS based masking, the parameters are $\alpha=(1,\alpha_1,\alpha_2)$, while any 2-out-of-3 elements in α gives an instance of $(2,1)$-SSS based masking and there are three of them in total. Then those four instances of SSS-based masking are evaluated by HOOD-based attacks (refer to Eqn. 18 for the distinguisher). Note that $(2,1)$-SSS masking shall be equivalent to IPM [42], [40].

The first group of comparison is shown in Fig. 9, where we have $\alpha=(1,2,4)$. The first observation is that adding one share of redundancy always reduces the concrete side-channel security of code-based masking. Secondly, given the same security order ($t=1$) under the word-probing model, those $(2,1)$-SSS based instances outperforms $(3,1)$-SSS based one. The more redundancy can only further reduce the security level. Interestingly, as three instances of $(2,1)$-SSS based masking have the same security order under the bit-probing model, the difference exists in $B_{d_{\mathcal{D}}^{\perp}}$ only. That is, given the same $d_{\mathcal{D}}^{\perp}$ over $\mathbb{F}_2$, more redundancy leads to a greater value of $B_{d_{\mathcal{D}}^{\perp}}$, indicating a lower concrete security level.

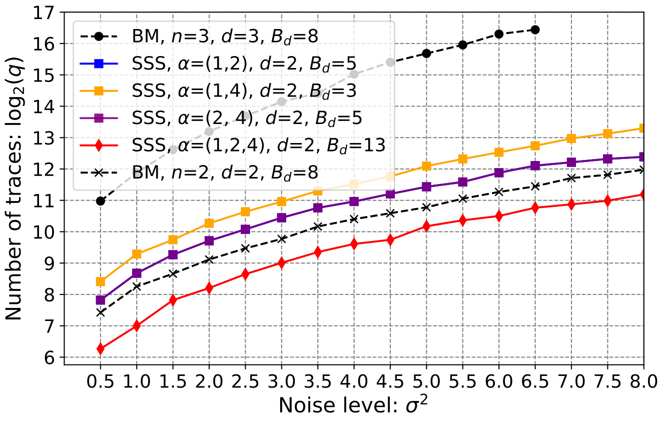

Figure 9. Illustrating the impact of redundancy by comparing $(2,1)$-SSS based masking with $(3,1)$-SSS based one, using $\alpha=(1,2,4)$ in the latter.

Another group of comparisons is presented in Fig. 10 with $\alpha=(1,3,17)$. It is worth noting that, both coding-theoretic parameters are different in two kinds of SSS-based maskings. Although one instance of $(2,1)$-SSS based masking is even better than the 3-share BM, the instance of $(3,1)$-SSS based masking gets much worse with one share of redundancy. In particular, the latter is even worse than the worst one among the three instances of $(2,1)$-SSS based masking. Overall, the attack-based evaluation results verify the impact of redundancy on the concrete security level of code-based masking.

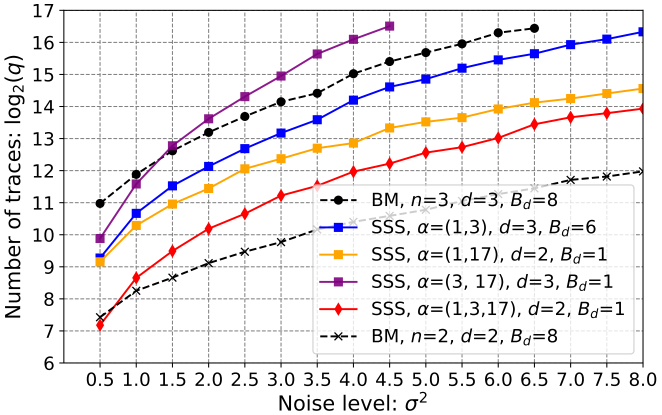

Figure 10. Illustrating the impact of redundancy by comparing $(2,1)$-SSS based masking with $(3,1)$-SSS based one, using $\alpha = (1,3,17)$ in the latter.

At last, we illustrate the impact of redundancy by presenting a comparison between the optimal codes in IPM and SSS-based masking. Those optimal codes are visualized in Fig. 11. In particular, the four (out of twelve) optimal codes for 2-share IPM and three optimal codes for $(3,1)$-SSS based masking are already shown in Fig. 4 and 7, respectively. Apparently, the redundancy can leverage an easier key-recovery attack in the sense of the necessary number of traces to succeed.

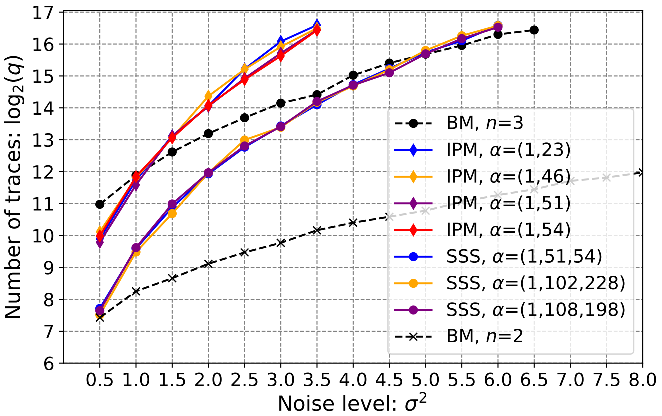

Figure 11. Illustrating the impact of redundancy by comparing 2-share IPM with $(3,1)$-SSS based masking, using $\alpha = (1,3,17)$ in the latter.

Those observations made in above two groups of comparison invoke the need of a trade-off between the amount of redundancy and the concrete security level in code-based masking. From a theoretical perspective, more redundancy can lead to more leakage, which is indicated by the two coding-theoretic properties. As a consequence, it is always advantageous to adopt non-redundant masking schemes rather than redundant ones when thwarting side-channel analysis. However, considering fault injection attacks (FIA) in real-world scenarios, a redundant masking scheme shall provide a combined countermeasure against both SCA and FIA.

More generally, above evaluation results show that only evaluating security orders under the probing model is not enough when assessing the concrete security level of a protected cryptographic implementation. Specifically, given the same side-channel security order (irrespective to word-level or bit-level), adding redundancy will always facilitate the adversaries in recovering secrets, and lower the concrete security in the sense of attacks. Therefore, we recommend further assessing the practical security of code-based masking by verifying both the dual distance and the (adjusted) kissing number in practice, since those coding-theoretic properties have been demonstrated by numerically simulated experiments in this work.

In summary, the attack-based evaluation confirms those theoretical findings in [40]. That is, we connect the dots in studying and improving code-based masking schemes. Particularly, we verify extensively by considering both IPM and SSS-based masking as instances of non-redundant and redundant instances of code-based masking.

B. Comparison with IPM-FD

In order to enhance IPM against fault injection attacks, another alternative for adding redundancy is proposed in [41] for fault detection, namely IPM-FD. In this subsection, we illustrate formal connections with SSS-based masking.

Consider the general case of IPM-FD with n shares and t random masks, denoted as (n,t)-IPM-FD, then the two generator matrices $\mathbf{G}$ and $\mathbf{H}$ are as follows:

$$\mathbf{G} = \left(\begin{array}{c|c} \boldsymbol{I}_{n-t} & \mathbf{0}_{(n-t)\times t} \end{array} \right) \in \mathbb{K}^{(n-t)\times n}$$

$$\mathbf{H} = \left(\begin{array}{ccc|c} \alpha_{1,1} & \cdots & \alpha_{1,n-t} & \\ \alpha_{2,1} & \cdots & \alpha_{2,n-t} & \\ \vdots & \cdots & \vdots & \boldsymbol{I}_t \\ \alpha_{t,1} & \cdots & \alpha_{t,n-t} & \end{array} \right) \in \mathbb{K}^{t\times n} \tag{20}$$

where $\boldsymbol{I}_t$ denotes identity matrix with size t and $\mathbf{0}_{(n-t)\times t}$ is a zero matrix. In particular, in IPM-FD, the tunable parameter is the family $\alpha = (\alpha_{i,j})$ where $1 \le i \le t$ and $1 \le j \le n-t$. Since the two codes $\mathcal{C}$ and $\mathcal{D}$ (generated by $\mathbf{G}$ and $\mathbf{H}$, resp.) are complementary, the side-channel resistance of IPM-FD is uniquely determined by the code $\mathcal{D}$ [53], [41].

Interestingly, by the equivalence of the linear codes, the two generator matrices of $\mathcal{D}$ in Eqn. 16 and 20 can be equivalent. For instance, taking $n = 3$ and $t = 1$, then the two instances of $\mathbf{H}$ are:

$$\begin{aligned} \mathbf{H}_{\text{SSS}} &= \left(\begin{array}{ccc} \alpha_1 & \alpha_2 & \alpha_3 \end{array} \right), \\ \mathbf{H}_{\text{IPM-FD}} &= \left(\begin{array}{ccc} \alpha_{1,1} & \alpha_{1,2} & 1 \end{array} \right), \end{aligned} \tag{21}$$

for $(3,1)$-SSS based masking and $(3,1)$-IPM-FD, respectively. Therefore, the two codes generated by $\mathbf{H}_{\text{SSS}}$ and $\mathbf{H}_{\text{IPM-FD}}$ are equivalent by setting $\alpha_2 = \alpha_{1,1}$, $\alpha_3 = \alpha_{1,2}$ and $\alpha_1 = 1$ (as in the previous subsection).

As in Sec. V-A, we enumerate all possible candidates for $(3,1)$-IPM-FD. Specifically, there are $32,640$ candidates in total: $20,581$ candidates have the dual distance equal to 3, and 90 candidates have $d_{\mathcal{D}}^{\perp} = 4$. They exactly match with those candidates in $(3,1)$-SSS based masking. However, there are more candidates in $(3,1)$-IPM-FD than in$(3,1)$-SSS based masking since $\alpha_{i,j}$ in Eqn. 20 can be equal to each other.

It is worth mentioning that those extra candidates all lead to linear codes with $d_{\mathcal{D}}^{\perp} = 2$ (as expected from a coding-theoretic perspective). The experimental results of taking two optimal linear codes in $(3,1)$-SSS based masking and $(3,1)$-IPM-FD are shown in Fig. 12. Apparently, those equivalent linear codes provide very similar resilience against HOOD-based attacks.

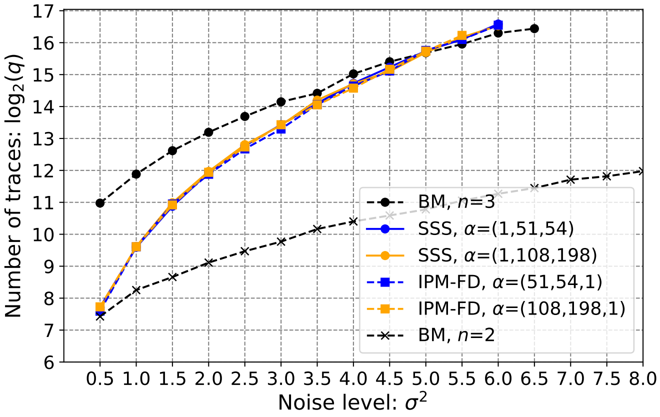

Figure 12. Comparison of equivalent linear codes used in $(3,1)$-SSS based masking and $(3,1)$-IPM-FD, by taking two optimal optimal codes in each scheme for illustration.

More generally with various n and t in IPM-FD and SSS-based masking, we highlight that there are three notable differences between the two schemes:

- the two linear codes $\mathcal{C}$ and $\mathcal{D}$ are complementary in the former, while there is no such condition in the latter;
- with the same choices of n and t, the former has more possible candidates of linear code than the latter;
- although both of them can detect faults from a coding-theoretic perspective, the ways of introducing redundancy are different in the two schemes.

C. Revisiting All Codes in the State-of-the-Art

Regarding the state-of-the-art, various instances of code-based masking have been presented in literature, accompanied with specific linear codes (which are tuning parameters) used in them. We therefore revisit all linear codes in the literature for a thorough comparison.

For the purpose of a fair comparison, we focus on instances of code-based masking in which the codes are generated over $\mathbb{F}_{2^8}$ by using AES's irreducible polynomial (see Sec. III-B). The results are detailed in Tab. VII. In particular, we present the best codes in several cases, along with the corresponding coding-theoretic properties.

The main takeaway point is that those optimal shall be used straightforwardly in practice, for instance, to protect AES implementations. We also provide instructive details for employing those codes in real circuits.

VI. Conclusions

In this work, we present an extensive attack-based evaluation on two representative instances of code-based masking, namely IPM and SSS-based masking. The higher-order optimal distinguisher is employed in numerically simulated evaluations. We highlight that various linear codes have significant impacts on the side-channel resilience of the corresponding scheme. Moreover, as an ultimate metric, the success rate of empirical attacks confirm the advantages of applying optimal instances of code-based masking.

Our attack-based evaluation completes the assessment of code-based masking assuming a known leakage model. Furthermore, compared with the state-of-the-art, we present the optimal codes (parameters) for both IPM and SSS-based masking in several cases. Those optimal codes should be of special interests to designers in devising more secure cryptographic circuits against side-channel attacks. As a part of future work, we will apply our theoretical and simulated analyses into practice and put forward practice-relevant evaluations of code-based masking on real-world circuits.

Appendix

A. Further Experimental Results.

For the sake of completeness, we provide here two more figures by using normal scale in y-axis, that are complementary to the results in Fig. 3 and 6, respectively. We highlight that with normal scale, it is more clear to illustrate the impact of different linear codes on the concrete security level of the corresponding instances.

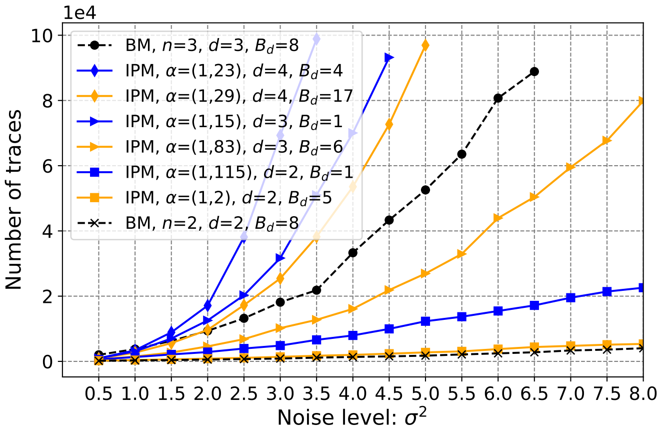

Figure 13. Attack-based evaluation of IPM with $n = 2$ shares with y-axis in normal scale. Note that we take two codes in each group with different $d_{\mathcal{D}}^{\perp}$ and/or $B_{d_{\mathcal{D}}^{\perp}}$.

Acknowledgments

This work has partly benefited from the bilateral MESRI-BMBF project "APRIORI" from the ANR cybersecurity 2020 call. It is also part of the Horizon 2020 "SPARTA" project under grant agreement number 830892. Besides, the authors acknowledge financial support of the French national Bank (BPI) under SECURYZR-V grant (Contract n° DOS0144216/00), a RISC-V centric platform integrating security co-processors.

Eventually, the authors would like to thank the anonymous reviewers and the Associate Editor, Prof. Georg Sigl, for their valuable comments that improved both the technical and the editorial quality of this paper.

Table VII
REVISITING ALL LINEAR CODES USED IN LITERATURE OVER $\mathbb{F}_{2^8}$, WITH REDUNDANCY WHEN $n > t+1$ WHILE NO REDUNDANCY WHEN $n = t+1$. NOTE THAT THE CODING-THEORETIC PARAMETERS ARE COMPUTED BY MAGMA [58], AND SOME NEEDED SCRIPTS ARE AVAILABLE FROM OUR OPEN SOURCES ON GITHUB, ALONG WITH GENERATED OUTPUTS.

Security Order t	**Num. of Shares** n	**Tunable Parameters** α **in Sharing**	**Masking Scheme**	**Coding-Theoretic Properties** (Def. 4 and 5)			**Comments**
				$d^{\perp}_{\mathcal{D}}$	$B_{d^{\perp}_{\mathcal{D}}}$	$B'_{d^{\perp}_{\mathcal{D}}}$	
$t = 1$	$\boldsymbol{n = 2}$ Non-redundant	$(1, 255)$ $(3, 7)$	IPM $(2, 1)$-SSS	3 3	2 2	2 2	[35]
		$(1, 17)$ $(1, 5)$ $(1, 7)$	IPM	2 3 4	1 4 8	1 4 8	[57] Three distinct codes
		$(221, 198)$, $(188, 189)$, $(237, 198)$ $\mathbf{(237, 175)}$	$(2, 1)$-SSS	3, 3, 3 4	1, 1, 1 4	1, 1, 1 4	[42]. Note that $\alpha = (237, 175)$ is optimal
		$\mathbf{(1, 23)}$, $\mathbf{(1, 46)}$, $\mathbf{(1, 51)}$, $\ldots$	IPM	**4, 4, 4**, $\ldots$	**4, 4, 4**, $\ldots$	**4, 4, 4**, $\ldots$	**This paper**. 12 optimal code in total, see Tab. II
	$n = 3$	$(5, 221, 198)$ $(237, 175, 221)$ $(237, 221, 198)$	$(3, 1)$-SSS	3 3 4	1 3 6	1 3 6	[42]
		$\mathbf{(1, 51, 54)}$, $\mathbf{(1, 102, 228)}$, $\mathbf{(1, 108, 198)}$	$(3, 1)$-SSS	**4, 4, 4**	**37, 37, 37**	**53, 53, 53**	**This paper**. Only 3 optimal codes, see Tab. VI
	$n = 4$	$(5, 237, 221, 198)$ $(237, 175, 221, 198)$ $(12, 80, 176, 237)$	$(4, 1)$-SSS	3 3 3	10 12 19	10 12 **53**	[42]
	$n = 5$	$(5, 237, 175, 221, 198)$	$(5, 1)$-SSS	2	2	2	[42]
$t = 2$	$\boldsymbol{n = 3}$ Non-redundant	$(1, 15, 233)$ $(13, 240, 163)$	IPM $(3, 2)$-SSS	5 6	1 2	1 2	[35]
		$(1, 146, 147)$, $(1, 188, 189)$	$(3, 2)$-SSS	3, 3	8, 8	8, 8	[42]. Both are equivalent to Boolean masking
		$\mathbf{(1, 27, 196)}$, $\mathbf{(1, 91, 204)}$, $\mathbf{(1, 218, 240)}$	IPM	**8, 8, 8**	**6, 6, 6**	**6, 6, 6**	**This paper**. Only 3 optimal codes, see Tab. IV
	$n = 5$	$(125, 246, 119, 104, 150)$, $(86, 23, 115, 107, 189)$ $(169, 63, 106, 49, 112)$ $(5, 237, 175, 221, 198)$	$(5, 2)$-SSS	4, 4 4 5	1, 1 2 6	1, 1 2 6	[45]
		$\mathbf{(1, 23, 71, 167, 235)}$	$(5, 2)$-SSS	**6**	**36**	**46**	**This paper**. We find only one optimal code by fixing $\alpha_1 = 1$ and $\alpha_2 = 23$

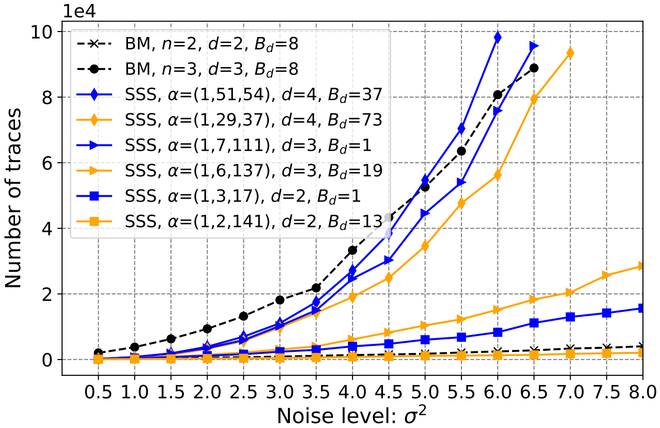

Figure 14. Attack-based evaluation of $(3, 1)$-SSS based masking with y-axis in normal scale. Note that we take two codes in each group with different $d^{\perp}_{\mathcal{D}}$ and/or $B_{d^{\perp}_{\mathcal{D}}}$.

REFERENCES

[1] P. C. Kocher, "Timing Attacks on Implementations of Diffie-Hellman, RSA, DSS, and Other Systems," in *Advances in Cryptology - CRYPTO '96, 16th Annual International Cryptology Conference, Santa Barbara, California, USA, August 18-22, 1996, Proceedings* (N. Koblitz, ed.), vol. 1109 of *Lecture Notes in Computer Science*, pp. 104–113, Springer, 1996.

[2] P. C. Kocher, J. Jaffe, and B. Jun, "Differential power analysis," in Wiener [59], pp. 388–397.

[3] C. Clavier, J.-S. Coron, and N. Dabbous, "Differential Power Analysis in the Presence of Hardware Countermeasures," in *CHES* (Ç. K. Koç and C. Paar, eds.), vol. 1965 of *Lecture Notes in Computer Science*, pp. 252–263, Springer, 2000.

[4] K. Gandolfi, C. Mourtel, and F. Olivier, "Electromagnetic analysis: Concrete results," in *Proceedings of the Third International Workshop on Cryptographic Hardware and Embedded Systems*, CHES '01, (London, UK, UK), pp. 251–261, Springer-Verlag, 2001.

[5] J.-J. Quisquater and D. Samyde, "ElectroMagnetic Analysis (EMA): Measures and Counter-Measures for Smard Cards," in *Smart Card Programming and Security (E-smart 2001)* (I. Attali and T. P. Jensen, eds.), vol. 2140 of *LNCS*, pp. 200–210, Springer-Verlag, September 2001. Nice, France. ISSN 0302-9743.

[6] S. Mangard, E. Oswald, and T. Popp, *Power Analysis Attacks: Revealing the Secrets of Smart Cards*. Springer, December 2006. ISBN 0-387-30857-1, http://www.dpabook.org/.

[7] K. Tiri and I. Verbauwhede, "A Logic Level Design Methodology for a Secure DPA Resistant ASIC or FPGA Implementation," in *2004 Design, Automation and Test in Europe Conference and Exposition (DATE 2004), 16-20 February 2004, Paris, France*, pp. 246–251, IEEE Computer Society, 2004.

[8] Z. Chen and Y. Zhou, "Dual-Rail Random Switching Logic: A Countermeasure to Reduce Side Channel Leakage," in *CHES*, vol. 4249 of *LNCS*, pp. 242–254, Springer, October 10-13 2006. Yokohama, Japan, http://dx.doi.org/10.1007/11894063_20.

[9] S. Chari, C. S. Jutla, J. R. Rao, and P. Rohatgi, "Towards Sound Approaches to Counteract Power-Analysis Attacks," in Wiener [59], pp. 398–412.

[10] E. Prouff and M. Rivain, "Masking against Side-Channel Attacks: A Formal Security Proof," in *Advances in Cryptology - EUROCRYPT 2013, 32nd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Athens, Greece, May 26-30, 2013. Proceedings* (T. Johansson and P. Q. Nguyen, eds.), vol. 7881 of *Lecture Notes in Computer Science*, pp. 142–159, Springer, 2013.

[11] A. Duc, S. Faust, and F. Standaert, “Making Masking Security Proofs Concrete - Or How to Evaluate the Security of Any Leaking Device,” in Oswald and Fischlin [60], pp. 401–429.

[12] É. de Chérisey, S. Guilley, O. Rioul, and P. Piantanida, “Best Information is Most Successful — Mutual Information and Success Rate in Side-Channel Analysis,” *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, vol. 2019, no. 2, pp. 49–79, 2019.

[13] Y. Ishai, A. Sahai, and D. Wagner, “Private Circuits: Securing Hardware against Probing Attacks,” in *CRYPTO*, vol. 2729 of *Lecture Notes in Computer Science*, pp. 463–481, Springer, August 17–21 2003. Santa Barbara, California, USA.

[14] A. Duc, S. Dziembowski, and S. Faust, “Unifying Leakage Models: From Probing Attacks to Noisy Leakage,” in *Advances in Cryptology - EUROCRYPT 2014 - 33rd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Copenhagen, Denmark, May 11-15, 2014. Proceedings* (P. Q. Nguyen and E. Oswald, eds.), vol. 8441 of *Lecture Notes in Computer Science*, pp. 423–440, Springer, 2014.

[15] S. Dziembowski, S. Faust, and M. Skorski, “Noisy leakage revisited,” in *Advances in Cryptology - EUROCRYPT 2015 - 34th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Sofia, Bulgaria, April 26-30, 2015, Proceedings, Part II* (E. Oswald and M. Fischlin, eds.), vol. 9057 of *Lecture Notes in Computer Science*, pp. 159–188, Springer, 2015.

[16] T. Prest, D. Goudarzi, A. Martinelli, and A. Passelègue, “Unifying Leakage Models on a Rényi Day,” in *Advances in Cryptology - CRYPTO 2019 - 39th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2019, Proceedings, Part I* (A. Boldyreva and D. Micciancio, eds.), vol. 11692 of *Lecture Notes in Computer Science*, pp. 683–712, Springer, 2019.

[17] F.-X. Standaert, T. Malkin, and M. Yung, “A Unified Framework for the Analysis of Side-Channel Key Recovery Attacks,” in *EUROCRYPT*, vol. 5479 of *LNCS*, pp. 443–461, Springer, April 26-30 2009. Cologne, Germany.

[18] C. Whitnall and E. Oswald, “A Comprehensive Evaluation of Mutual Information Analysis Using a Fair Evaluation Framework,” in *CRYPTO* (P. Rogaway, ed.), vol. 6841 of *Lecture Notes in Computer Science*, pp. 316–334, Springer, 2011.

[19] W. Cheng, Y. Liu, S. Guilley, and O. Rioul, “Attacking masked cryptographic implementations: Information-theoretic bounds,” *CoRR*, vol. abs/2105.07436, 2021.

[20] S. Bhasin, J.-L. Danger, S. Guilley, and Z. Najm, “Side-Channel Leakage and Trace Compression using Normalized Inter-Class Variance,” *IACR Cryptology ePrint Archive*, vol. 2014, p. 1020, 2014.

[21] É. Brier, C. Clavier, and F. Olivier, “Correlation power analysis with a leakage model,” in *Cryptographic Hardware and Embedded Systems - CHES 2004: 6th International Workshop Cambridge, MA, USA, August 11-13, 2004. Proceedings* (M. Joye and J. Quisquater, eds.), vol. 3156 of *Lecture Notes in Computer Science*, pp. 16–29, Springer, 2004.

[22] B. Gierlichs, L. Batina, P. Tuyls, and B. Preneel, “Mutual information analysis,” in *CHES, 10th International Workshop*, vol. 5154 of *Lecture Notes in Computer Science*, pp. 426–442, Springer, August 10-13 2008. Washington, D.C., USA.

[23] É. de Chérisey, S. Guilley, A. Heuser, and O. Rioul, “On the optimality and practicability of mutual information analysis in some scenarios,” *Cryptography and Communications*, vol. 10, no. 1, pp. 101–121, 2018.

[24] S. Chari, J. R. Rao, and P. Rohatgi, “Template attacks,” in Kaliski *et al.* [61], pp. 13–28.

[25] E. Oswald and S. Mangard, “Template Attacks on Masking — Resistance Is Futile,” in *CT-RSA* (M. Abe, ed.), vol. 4377 of *Lecture Notes in Computer Science*, pp. 243–256, Springer, 2007.

[26] W. Schindler, K. Lemke, and C. Paar, “A Stochastic Model for Differential Side Channel Cryptanalysis,” in *CHES* (LNCS, ed.), vol. 3659 of *LNCS*, pp. 30–46, Springer, Sept 2005. Edinburgh, Scotland, UK.

[27] B. Gierlichs, K. Lemke-Rust, and C. Paar, “Templates vs. Stochastic Methods,” in *CHES*, vol. 4249 of *LNCS*, pp. 15–29, Springer, October 10-13 2006. Yokohama, Japan.

[28] A. Heuser, O. Rioul, and S. Guilley, “Good Is Not Good Enough - Deriving Optimal Distinguishers from Communication Theory,” in *Cryptographic Hardware and Embedded Systems - CHES 2014 - 16th International Workshop, Busan, South Korea, September 23-26, 2014. Proceedings* (L. Batina and M. Robshaw, eds.), vol. 8731 of *Lecture Notes in Computer Science*, pp. 55–74, Springer, 2014.

[29] N. Bruneau, S. Guilley, A. Heuser, and O. Rioul, “Masks Will Fall Off – Higher-Order Optimal Distinguishers,” in *Advances in Cryptology – ASIACRYPT 2014 - 20th International Conference on the Theory and Application of Cryptology and Information Security, Kaoshiung, Taiwan, R.O.C., December 7-11, 2014, Proceedings, Part II* (P. Sarkar and T. Iwata, eds.), vol. 8874 of *Lecture Notes in Computer Science*, pp. 344–365, Springer, 2014.

[30] J. Cooper, G. Goodwill, J. Jaffe, G. Kenworthy, and P. Rohatgi, “Test Vector Leakage Assessment (TVLA) Methodology in Practice,” Sept 24–26 2013. International Cryptographic Module Conference (ICMC), Holiday Inn Gaithersburg, MD, USA.

[31] L. Mather, E. Oswald, J. Bandenburg, and M. Wójcik, “Does my device leak information? an a priori statistical power analysis of leakage detection tests,” in *Advances in Cryptology - ASIACRYPT 2013 - 19th International Conference on the Theory and Application of Cryptology and Information Security, Bengaluru, India, December 1-5, 2013, Proceedings, Part I* (K. Sako and P. Sarkar, eds.), vol. 8269 of *Lecture Notes in Computer Science*, pp. 486–505, Springer, 2013.

[32] T. Schneider and A. Moradi, “Leakage assessment methodology - A clear roadmap for side-channel evaluations,” in *Cryptographic Hardware and Embedded Systems - CHES 2015 - 17th International Workshop, Saint-Malo, France, September 13-16, 2015, Proceedings* (T. Güneysu and H. Handschuh, eds.), vol. 9293 of *Lecture Notes in Computer Science*, pp. 495–513, Springer, 2015.

[33] J. D. Golić and C. Tymen, “Multiplicative masking and power analysis of AES,” in Kaliski *et al.* [61], pp. 198–212.

[34] G. Fumaroli, A. Martinelli, E. Prouff, and M. Rivain, “Affine Masking against Higher-Order Side Channel Analysis,” in *Selected Areas in Cryptography* (A. Biryukov, G. Gong, and D. R. Stinson, eds.), vol. 6544 of *Lecture Notes in Computer Science*, pp. 262–280, Springer, 2010.

[35] J. Balasch, S. Faust, and B. Gierlichs, “Inner Product Masking Revisited,” in Oswald and Fischlin [60], pp. 486–510.

[36] J. Bringer, C. Carlet, H. Chabanne, S. Guilley, and H. Maghrebi, “Orthogonal Direct Sum Masking - A Smartcard Friendly Computation Paradigm in a Code, with Builtin Protection against Side-Channel and Fault Attacks,” in *Information Security Theory and Practice. Securing the Internet of Things - 8th IFIP WG 11.2 International Workshop, WISTP 2014, Heraklion, Crete, Greece, June 30 - July 2, 2014. Proceedings* (D. Naccache and D. Sauveron, eds.), vol. 8501 of *Lecture Notes in Computer Science*, pp. 40–56, Springer, 2014.

[37] L. Goubin and A. Martinelli, “Protecting AES with Shamir’s Secret Sharing Scheme,” in Preneel and Takagi [62], pp. 79–94.

[38] E. Prouff and T. Roche, “Higher-Order Glitches Free Implementation of the AES Using Secure Multi-party Computation Protocols,” in Preneel and Takagi [62], pp. 63–78.

[39] W. Wang, P. Méaux, G. Cassiers, and F. Standaert, “Efficient and Private Computations with Code-Based Masking,” *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, vol. 2020, no. 2, pp. 128–171, 2020.

[40] W. Cheng, S. Guilley, C. Carlet, J. Danger, and S. Mesnager, “Information Leakages in Code-based Masking: A Unified Quantification Approach,” *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, vol. 2021, no. 3, pp. 465–495, 2021.

[41] W. Cheng, C. Carlet, K. Goli, J. Danger, and S. Guilley, “Detecting faults in inner product masking scheme,” *J. Cryptogr. Eng.*, vol. 11, no. 2, pp. 119–133, 2021.

[42] N. Costes and M. Stam, “Redundant code-based masking revisited,” *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, vol. 2021, no. 1, pp. 426–450, 2021.

[43] N. Veyrat-Charvillon, B. Gérard, M. Renauld, and F.-X. Standaert, “An Optimal Key Enumeration Algorithm and Its Application to Side-Channel Attacks,” in *Selected Areas in Cryptography* (L. R. Knudsen and H. Wu, eds.), vol. 7707 of *Lecture Notes in Computer Science*, pp. 390–406, Springer, 2012.

[44] R. Poussier, F. Standaert, and V. Grosso, “Simple key enumeration (and rank estimation) using histograms: An integrated approach,” in Gierlichs and Poschmann [63], pp. 61–81.

[45] H. Chabanne, H. Maghrebi, and E. Prouff, “Linear repairing codes and side-channel attacks,” *IACR Trans. Cryptogr. Hardw. Embed. Syst.*, vol. 2018, no. 1, pp. 118–141, 2018.

[46] A. Battistello, J. Coron, E. Prouff, and R. Zeitoun, “Horizontal Side-Channel Attacks and Countermeasures on the ISW Masking Scheme,” in Gierlichs and Poschmann [63], pp. 23–39.

[47] S. Guilley, A. Heuser, and O. Rioul, “A Key to Success - Success Exponents for Side-Channel Distinguishers,” in *Progress in Cryptology - INDOCRYPT 2015 - 16th International Conference on Cryptology in India, Bangalore, India, December 6-9, 2015, Proceedings* (A. Biryukov and V. Goyal, eds.), vol. 9462 of *Lecture Notes in Computer Science*, pp. 270–290, Springer, 2015.

[48] N. Veyrat-Charvillon and F. Standaert, “Mutual Information Analysis: How, When and Why?,” in *Cryptographic Hardware and Embedded*

Systems - CHES 2009, 11th International Workshop, Lausanne, Switzerland, September 6-9, 2009, Proceedings (C. Clavier and K. Gaj, eds.), vol. 5747 of *Lecture Notes in Computer Science*, pp. 429–443, Springer, 2009.

[49] S. Guilley, P. Hoogvorst, R. Pacalet, and J. Schmidt, "Improving Side-Channel Attacks by Exploiting Substitution Boxes Properties," in *BFCA* (Presse Universitaire de Rouen et du Havre, ed.), pp. 1–25, 2007. May 02–04, Paris, France, http://www.liafa.jussieu.fr/bfca/books/BFCA07.pdf.

[50] A. Moradi, N. Mousavi, C. Paar, and M. Salmasizadeh, "A Comparative Study of Mutual Information Analysis under a Gaussian Assumption," in *WISA (Information Security Applications, 10th International Workshop)*, vol. 5932 of *Lecture Notes in Computer Science*, pp. 193–205, Springer, August 25-27 2009. Busan, Korea.

[51] E. Prouff and M. Rivain, "Theoretical and practical aspects of mutual information-based side channel analysis," *International Journal of Applied Cryptography (IJACT)*, vol. 2, no. 2, pp. 121–138, 2010.

[52] R. Poussier, Q. Guo, F. Standaert, C. Carlet, and S. Guilley, "Connecting and Improving Direct Sum Masking and Inner Product Masking," in *Smart Card Research and Advanced Applications - 16th International Conference, CARDIS 2017, Lugano, Switzerland, November 13-15, 2017, Revised Selected Papers* (T. Eisenbarth and Y. Teglia, eds.), vol. 10728 of *Lecture Notes in Computer Science*, pp. 123–141, Springer, 2017.

[53] W. Cheng, S. Guilley, C. Carlet, S. Mesnager, and J.-L. Danger, "Optimizing Inner Product Masking Scheme by a Coding Theory Approach," *IEEE Trans. Inf. Forensics Secur.*, vol. 16, pp. 220–235, 2021.

[54] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Elsevier, Amsterdam, North Holland, 1977. ISBN: 978-0-444-85193-2.

[55] W. Cheng, Y. Liu, S. Guilley, and O. Rioul, "Towards Finding Best Linear Codes for Side-Channel Protections," in *Proceedings of 10th International Workshop on Security Proofs for Embedded Systems* (K. Heydemann, U. K\"uhne, and F. T. Zhang, eds.), Kalpa Publications in Computing, pp. 1–16, EasyChair, 2021.

[56] W. Cheng, *What Can Information Guess? Towards Information Leakage Quantification in Side-Channel Analysis*. Theses, Institut Polytechnique de Paris, Dec. 2021.

[57] J. Balasch, S. Faust, B. Gierlichs, C. Paglialonga, and F. Standaert, "Consolidating Inner Product Masking," in *Advances in Cryptology - ASIACRYPT 2017 - 23rd International Conference on the Theory and Applications of Cryptology and Information Security, Hong Kong, China, December 3-7, 2017, Proceedings, Part I* (T. Takagi and T. Peyrin, eds.), vol. 10624 of *Lecture Notes in Computer Science*, pp. 724–754, Springer, 2017.

[58] University of Sydney (Australia), "Magma Computational Algebra System." http://magma.maths.usyd.edu.au/magma/, Accessed on 2021-08-22.

[59] M. J. Wiener, ed., *Advances in Cryptology - CRYPTO '99, 19th Annual International Cryptology Conference, Santa Barbara, California, USA, August 15-19, 1999, Proceedings*, vol. 1666 of *Lecture Notes in Computer Science*, Springer, 1999.

[60] E. Oswald and M. Fischlin, eds., *Advances in Cryptology - EUROCRYPT 2015 - 34th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Sofia, Bulgaria, April 26-30, 2015, Proceedings, Part I*, vol. 9056 of *Lecture Notes in Computer Science*, Springer, 2015.

[61] B. S. Kaliski, Jr., Ç. K. Koç, and C. Paar, eds., *Cryptographic Hardware and Embedded Systems - CHES 2002, 4th International Workshop, Redwood Shores, CA, USA, August 13-15, 2002, Revised Papers*, vol. 2523 of *Lecture Notes in Computer Science*, Springer, 2003.

[62] B. Preneel and T. Takagi, eds., *Cryptographic Hardware and Embedded Systems - CHES 2011 - 13th International Workshop, Nara, Japan, September 28 – October 1, 2011. Proceedings*, vol. 6917 of *LNCS*, Springer, 2011.

[63] B. Gierlichs and A. Y. Poschmann, eds., *Cryptographic Hardware and Embedded Systems - CHES 2016 - 18th International Conference, Santa Barbara, CA, USA, August 17-19, 2016, Proceedings*, vol. 9813 of *Lecture Notes in Computer Science*, Springer, 2016.