A Formal Security Proof of Masking

Reduction from Strong Noisy Leakage to Probing Model without Random Probing and Application to LR Primitive

Rei Ueno[1][0000−0002−9754−6792], Akiko Inoue[2][0000−0002−0173−7245], Kazuhiko Minematsu[2][0000−0002−3427−6772], Akira Ito[3][0000−0002−4602−7570], and Naofumi Homma[3][0000−0003−0864−3126]

[1] Kyoto University, Japan. ueno.rei.2e@kyoto-u.ac.jp
[2] NEC, Japan. a_inoue@nec.com, k-minematsu@nec.com
[3] Tohoku University, Japan. akira.ito.b1@tohoku.ac.jp, naofumi.homma.c8@tohoku.ac.jp

Abstract. This paper provides upper bounds on the success rate (SR) of side-channel attacks (SCAs) on masked implementations. We present a formal security proof of additive masking—including both Boolean and arithmetic—through new reductions from *strong noisy leakage (SNL)* to the probing model. Unlike existing proofs relying on noisy leakage (NL) and random probing (RP), our proof introduces a novel security notion named leakage energy (LE), which enables a stronger bound. Our proof reveals the *necessary and sufficient* condition for asymptotic security of additive masking in both the NL and mutual information frameworks, which includes a resolution to an open problem in TCC 2016. Our claims are validated through numerical evaluations. As an application of our theorems, we propose a binary block-cipher based leakage-resilient primitive based on a variant of XEX, which claims d-th order SCA security of arithmetic masking by design under some assumptions, enabling efficient OCB-style authenticated encryption with implementation cost of $O(d)$.

Keywords: Boolean masking · Arithmetic masking · Probing model · Noisy leakage · Mutual information · Leakage resilience · XEX.

1 Introduction

1.1 Background

Masking is a major countermeasure against power/electromagnetic side-channel attacks (SCAs) [48]. In a d-th order masking, secret variable z is decomposed into $d+1$ random shares s_i $(0 \le i \le d)$ such that $z = s_0 + s_1 + \cdots + s_d$, where the addition is defined by the cryptosystem. In typical symmetric ciphers including AES, it is XOR, which is called Boolean masking. A masking defined over integer ring modulo m is called arithmetic masking. Masked implementation stores the shares on the device instead of secret value itself, and performs cryptographic operation using them based on multiparty computation on the device. Masking

shall force attackers to observe leakages of all shares for key recovery, while each observation suffers from noise. Thus, the masking mitigates the leakage and renders SCA difficult due to the noise accumulation.

Its formal security was first analyzed in [23], focusing on one-bit differential power analysis (DPA) and assuming significant noise. It was followed by many studies for strengthening the noise condition, generalization to other SCAs, and extension to arithmetic masking [7,8,34,36,45,53,59]. The formal security proof is done recently by bounding the SCA success rate (SR) or adversarial advantage using a function of mutual information (MI) negligible in masking order d [7,34, 45, 53, 59], as MI represents how much information leaks from device. Deriving stronger bounds is important not only from theoretical perspectives but also from practical security evaluation of masked implementation, because it is useful to determine the key lifetime of a masked implementation for a given device [67].

Currently, the d-probing model (or d-threshold probing) [43] is promising to define and verify d-th order masked implementation. However, it evaluates a masked implementation with a binary decision of secure or insecure against d-probing attack, but can neither discuss nor evaluate the SR of a $(d+1)$-th order SCA, whereas real-world SCA would be unbounded in terms of d. In [34], Duc *et al.* presented an SR upper bound by MI using a reduction from noisy leakage (NL) [33] to probing model through random probing (RP) model [33][4]. The NL notion expresses the principle of SCA using the *total variation distance* (or *statistical distance*) of probability distributions; that is, the side-channel leakage supposedly depends on the intermediate value. However, it was difficult to directly manipulate the NL in mathematical proofs. Accordingly, many studies have focused on the reductions and security proofs to discuss the SR in RP [6, 9–12, 15, 19, 33, 70]. By contrast, the reduction *via* RP incurs a significant loss, resulting in a weak bound, as evaluated in comparison with actual SCAs in [25, 45]. In addition, the SR bound derived through RP holds only if the side-channel measurement is significantly noisy such that $\max_i I(S_i; \boldsymbol{L}_i) < 2^{-2l+1}$, where $I(S_i; \boldsymbol{L}_i)$ is the MI and l is the bit-length of partial key ($l = 8$ for AES).

In [45], Ito *et al.* gave another bound and proof based on communication-channel modeling of higher-order SCA with Fano's inequality [31, Theorem 2.10.1]. The bound was stronger than that by Duc *et al.*, and holds under a significantly strong condition of $\max_i I(S_i; \boldsymbol{L}_i) < (2 \ln 2)^{-1}$, limited to Boolean masking. In [53], Masure *et al.* derived similar (slightly weaker) bound for both Boolean and arithmetic maskings. In [7], Béguinot *et al.* presented yet another bound for Boolean masking claimed to be stronger. These communication-channel-based bounds (which are defined in this paper as ones based on Fano's inequality) are significantly stronger than reduction-based ones. However, their relationship remains undetermined. Many studies on implementation of masking, including circuit compiler and practical evaluations, have been conducted in the NL and

[4] We use the term *reduction* for derivation of inequality among masking security/adversarial notions as in literature like [8, 34], while the meaning would be different from the provable security framework in cryptology [66]. In this paper, *reduction-based bounds/proofs* are defined as those based on reductions among (S)NL and RP.

RP, which cannot incorporate these stronger bounds in [7, 45, 53]. In addition to a theoretical perspective, it is important to clarify the relationship to realize masked implementation with appropriate parameters under stronger bounds.

Arithmetic masking is currently attracting much attention, because prime-field masking is promising for realizing secure implementations of lattice-based cryptography [1–3, 18, 27–30, 37, 56, 57]. More importantly, arithmetic masking, especially prime-field masking, is claimed to offer higher security than Boolean masking [36]. However, its current application in symmetric cryptography is limited, because most ciphers including AES are defined over a binary field. Although some researchers have developed prime-field ciphers to exploit the prime-field masking [4, 20, 52], a dedicated non-binary cipher may face limitations in deployment, especially in integrating it in existing modes of operations. Moreover, the methodology for evaluating the security of prime-field ciphers is still underdeveloped, as Beyne and Verbauwhede recently showed in [17] that prime-field ciphers do not achieve their claimed security with respect to a cryptanalysis dedicated to prime field. Meanwhile, arithmetic masking might be primarily developed for Addition-Rotation-XOR (ARX) ciphers (*e.g.*, ChaCha20 [13]), but it requires Boolean masking for some parts. Thus, although arithmetic masking would achieve higher security than Boolean masking of the same order, its usage in symmetric ciphers is underdeveloped. It is excellent if we have a methodology to exploit the security of arithmetic masking with conventional binary ciphers, which enables higher SCA security with minimal performance overhead.

1.2 Technical overview, main ideas, and our contribution

Tight security reduction with new adversarial notions. Focusing on the encoding of masking (*i.e.*, smallest but essential gadget, see Remark 1), we newly define a reduction to d-probing model using a communication-channel, unlike the approach by Duc *et al.* using RP and the Chernoff bound [34]. Using it, we establish a new and tight reduction among masking security/adversarial notions. For this purpose, we propose two new notions: *strong noisy leakage (SNL)* and *leakage energy (LE)*. Figure 1 illustrates the relation among masking security notions, where the proposed notions and reductions are highlighted in red. Importantly, SNL is defined to enable a Fourier/Walsh–Hadamard analysis in combination with Parseval's theorem, yielding a stronger bound than existing reduction-based ones. Our reduction establishes another path different from RP, which enables us to avoid lossy reductions relating to the RP. Our security reduction is concluded with Fano's inequality (as same as [25, 45]), which reveals the relationship between reduction-based and communication-channel-based proofs.

Generalization of existing bounds and security proof of masking under strongest condition. We derive bounds that holds for both Boolean and arithmetic maskings. They are followed by a formal security proof, stating an exponential decay of adversarial advantage by increase in masking order d under the strongest condition of $\max_i I(S_i; \boldsymbol{L}_i) < \log \gamma$, where γ is the smallest non-trivial divisor of field/ring characteristic (if characteristic is prime, γ is the

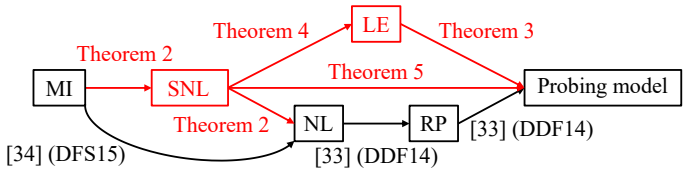


Fig. 1: Reductions among MI and masking security notions/models.

prime)[5]. In addition, we derive the *necessary and sufficient* condition for the asymptotic security by proving the limit of $I(Z; \boldsymbol{X})$ as $d \to \infty$. In the proof, we utilize a new concept of *degraded leakage* and demonstrate the importance of *subgroup leakage* with formal treatment; namely, the information of subgroup to which each share belong plays an important role in SCA. We further provide a numerical evaluation of the bound and SCA success rate for a validation.

Application to leakage-resilient (LR) primitive design. In our proof and evaluation, we confirm a higher SCA security of arithmetic masking than Boolean. To utilize the advantage of arithmetic masking with conventional block ciphers (BCs) in practice, we propose a new LR primitive named $\mathsf{gXEX\text{-}A}^d$, which is an arithmetic variant of XEX. It claims an inherent d-th order SCA security with arithmetic masking by design (*i.e.*, without explicit adoption of masking. Using $\mathsf{gXEX\text{-}A}^d$, we also present AOCB, an OCB-style authenticated encryption (AE) [63]. AOCB achieves a provable security of AE and leakage resilience of *leveled implementation* when combined with [54], assuming that the underlying BC (typically AES) is black-box secure. Its implementation cost is $O(d)$ in order to achieve d-th order SCA security. As a d-th order masking usually requires an implementation cost of $O(d^2)$, our LR primitive breaks the barrier and would yield d-th order SCA security with unprecedented implementation performance.

1.3 Paper organization

Section 2 introduces the mathematical notations and basics of masking. Section 3 reviews the existing security notions and proofs for masking. Section 4 proposes our communication channel and new security notions. Section 5 presents our reduction and security proof. Section 6 demonstrates a numerical evaluation of masking schemes. Finally, Section 7 proposes an LR primitive and AE that achieves d-th order SCA security by design and validate it experimentally.

[5] Our proof includes a resolution to open problem about noise requirement for asymptotic security of masking by Dziembowski *et al.* in TCC 2016 [36]. Although Jahandideh *et al.* claimed to prove the conjecture independently in Asiacrypt 2025 [47], our proof is based on a different approach. It enables us to discuss the noise requirement in MI beyond NL, its necessity and sufficiency, asymptotic behavior, and the properties of leakage that render masking (in)secure in Section 5.6.

2 Preliminaries

2.1 Notations

We denote sets by a calligraphic letter (*e.g.*, $\mathcal{X}$). A random variable and an element of the set are denoted by italic uppercase and lowercase letters (*i.e.*, X and x), respectively. We denote the probability measure by Pr. We denote a probability density or mass function of X by p_X. A sequence of n random variables is expressed by $X^n = (X_0, X_1, \ldots, X_{n-1})$. For a measurable function f, $\mathbb{E}_X f(X)$ denotes the expectation of $f(X)$ with respect to X. We denote the binary and natural logarithms by log and ln, respectively. The entropy of X is defined as $H(X) = -\mathbb{E}_X \log p_X(X)$, and the mutual information between X and Y is defined as $I(X;Y) = \mathbb{E}_{X,Y} \log(p_{X,Y}(X,Y)/p_X(X)p_Y(Y))$. Further, $D_{\mathrm{KL}}(p \parallel q) = \mathbb{E}_X \log(p(X)/q(X))$ denotes the Kullback–Leibler (KL) divergence between the probability distributions p and q. Let i, e, and π denote the imaginary unit, Euler's number, and the circle constant, respectively. Let $\mathbb{Z}_m = \mathbb{Z}/m\mathbb{Z}$ denote the integer ring modulo m. An m-point discrete Fourier transform (DFT) of a function $f : \mathbb{Z}_m \to \mathbb{R}$ is denoted by $\hat{f} : \mathbb{Z}_m \to \mathbb{C}$. In addition, Walsh–Hadamard transform (WHT) of a function $f : \mathbb{Z}_2^l \to \mathbb{R}$ is also denoted by $\hat{f} : \mathbb{Z}_2^l \to \mathbb{R}$, where l is the bit length. See Section A for the details of DFT and WHT. Let $\langle \cdot \mid \cdot \rangle : \mathbb{Z}_2^l \times \mathbb{Z}_2^l \to \mathbb{Z}_2$ be an inner product, which is used for the WHT twiddle factor. Decimal representation is used for both elements of $\mathbb{Z}_m$ and $\mathbb{Z}_2^l = \mathbb{F}_2^l$.

2.2 Additive masking

The basic idea of additive masking is decomposing the secret variable into random shares over the additive group of field/ring defining the cryptosystem to mitigate the correlation between the variable and leakage. In a d-th order masking, a secret variable z is represented using $d+1$ shares $s_0, s_1, \ldots, s_d$ such that $z = s_0 + s_1 + \cdots + s_d$, where s_i is called share and is supposedly sampled from a uniform distribution. If the addition is defined by XOR, it is called Boolean masking, which has been major for usual binary ciphers including AES. Maskings for non-binary fields/rings is called arithmetic maskings, which operates over the ring $\mathbb{Z}_m$. Arithmetic masking with $m = 2^l$ (where l is the bit length) has been typically used in addition-XOR-rotation (ARX) ciphers. In addition, the prime-field masking is over $\mathbb{Z}_q = \mathbb{F}_q$ (where q is a prime), which has attracted much attention recently owing to the gaining presence of lattice-based cryptography and the claim of its higher SCA security than Boolean masking [4, 36, 52].

The security of masking is based on Shamir's secret sharing [64]; namely, if the attacker obtains up to only d shares out of $d+1$, the unshared variable can never be recovered in an information-theoretic sense. In the masked implementation, we must compute the cryptographic function using shares in such a way that the original secret variables are never recovered by d probes on the implementation, which is formally defined as *d-probing model* in [43], as described in Section 3.1.

3 Existing notions and proofs of masking

3.1 The probing model

The d-probing model (or d-threshold probing) formulates an adversary that directly probes d wires (*i.e.*, intermediate variables) in the masked circuit [43]. Although it was originally defined using a simulation of joint probability distribution, we introduce an intuitive definition that has been commonly used [62].

Definition 1 (Probing model [43, 62], informal). *A masked circuit is d-probing (or d-threshold probing) secure if any d combination of probes in the circuit neither reveals the unshared secret variable.*

If we consider an encoding of masking (namely, memory elements storing shares) without computation, it is trivially d-probing secure if the number of shares is $d + 1$ and their leakages are independent of each other. The probing model and its extensions (*e.g.*, NI/SNI [5] and PINI [21]) have been definitely useful to define a sound masking and verify the d-th order security for a given masked circuit. However, they evaluate masked circuits by binary decision of secure or insecure against d-th order attack, but can neither discuss nor evaluate the SR of $(d+1)$-th order SCA. To discuss the SR, we need to translate the probing model to another notion *via* an upper-bound based on probability-related metric (*i.e.*, have a reduction to probing model) In [33], Duc *et al.* reduced the RP (introduced below) to the probing model *via* the Chernoff bound. In this paper, we present another translation using an information-theoretic bound based on Fano's and data processing inequalities.

3.2 Reduction-based proof using noisy leakage and random probing

In [33], Duc *et al.* presented the noisy leakage (NL) model, which represents a fact that, if the attacker obtains a side-channel leakage $\boldsymbol{L}_i$ of a masking share S_i, the attacker can perform a guess of S_i better than random guessing. Thus, the NL model reflects the principle and physical phenomenon of SCA; that is, the physical leakage depends on the secret intermediate value/share. Formally, it is defined *via* the *total variation distance* or *statistical distance*[6] between conditional probability distribution $p_{S_i|\boldsymbol{L}_i}$ and non-conditional probability distribution p_{S_i} (*i.e.*, a uniform distribution over $\mathcal{S}$) as follows:

Definition 2 (δ-noisy leakage (δ-NL) [59]). *Let δ be a non-negative real number. A leakage is δ-noisy and an adversary with the leakage is δ-NL if*

$$\max_i \mathbb{E}_{\boldsymbol{L}} \left[\sum_s \left| p_{S_i|\boldsymbol{L}_i}(s \mid \boldsymbol{L}) - p_{S_i}(s) \right| \right] \leq \delta. \tag{1}$$

[6] In the literature, NL is defined as the total variation distance itself; namely, the left-hand side of (1) is scaled by 1/2. In our definition, the left-hand side of (1) is doubled for the scaling consistency with other notions.

A major reason why the NL is significant would be that it is a lower bound of MI (*i.e.*, MI is reduced to NL) through Pinsker's inequality [39, 58] as

$$\begin{aligned}\left(\mathbb{E}_{\boldsymbol{L}}\left[\sum_s \left|p_{S_i|\boldsymbol{L}_i}(s \mid \boldsymbol{L}) - p_{S_i}(s)\right|\right]\right)^2 &= \left(\int_{\mathcal{L}} \sum_s \left|p_{S_i,\boldsymbol{L}_i}(s,\boldsymbol{l}) - p_{S_i}(s)p_{\boldsymbol{L}_i}(\boldsymbol{l})\right| \mathrm{d}\boldsymbol{l}\right)^2 \\ &\leq 2\ln(2) D_{\mathrm{KL}}(p_{S_i,\boldsymbol{L}_i} \parallel p_{S_i} p_{\boldsymbol{L}_i}) \\ &= 2\ln(2) I(S_i; \boldsymbol{L}_i), \end{aligned} \tag{2}$$

as mentioned in [34]. Another important characteristic of NL is that it is reduced to the probing model through the RP model [33].

Definition 3 (Random probing (RP) [33]). *Let ρ be a non-negative real number in the range of $[0, 1]$. A ρ-RP attacker observes each intermediate bit value (i.e., wires in the circuit) with probability of ρ.*

RP has been used in mathematical analyses of masking instead of NL; RP enables us to discuss the formal security of masking through its reduction from NL to probing model. In fact, in [34], Duc *et al.* proved that the success rate (SR) of $(d+1)$-th order SCA on d-th order masking is bounded from above as

$$\mathrm{SR} \leq 1 - \left(1 - \left(m\sqrt{\delta \ln 2}\right)^{d+1}\right)^n \left(1 - \frac{1}{m}\right), \tag{3}$$

where m is the cardinality of ring/field and n is the number of traces[7]. However, the bound in (3) is non-negligibly weak as shown in [25, 45], and is trivial unless leakage is significantly noisy such that $I(Z; \boldsymbol{X}) \leq 2^{-2l+1}$ for Boolean masking, meaning that the condition for asymptotic security is also weak.

As another important result, Dziembowski *et al.* showed that the condition of NL for asymptotic security is given by $\delta/2 < 1 - 1/p$ for prime-field masking and $\delta/2 < 1/2$ for arbitrary maskings [36, Proposition 1][8], which implies that prime-field masking is more secure than Boolean masking. Further, they conjectured that the noise requirement for asymptotic security is determined by the size of the largest non-trivial subgroup of field/ring, as resolved in Theorem 7 and [47].

3.3 Communication-channel-based proof

In 2019, de Chérisey *et al.* established a communication-channel model representing SCA to derive an SR upper-bound using Fano's inequality [31]. Subsequently, in 2022 [45], Ito *et al.* extended the model to Boolean masking and derived an SR upper-bound of $(d+1)$-th order SCA as a function of d as

$$\xi(\mathrm{SR}) \leq n \log\left(1 + (2^l - 1)(2\ln 2)^{d+1} \prod_{i=0}^{d} I(S_i; \boldsymbol{L}_i)\right),$$

[7] Bound (3) is a corrected version by Béguinot *et al.* in [8] with our definition of δ.

[8] Because of difference in the definition of NL between this paper and [36], the bound is given for $\delta/2$ instead of δ.

where $\xi : [0,1] \to \mathbb{R}_0^+$ is defined as $\xi(r) = H(K) - (1-r)\log(m-1) - H_2(r)$ (here, $H(K) = l$ is the entropy of K and H_2 is the binary entropy function). This bound enables a formal proof of asymptotic security of masking as $d \to \infty$. The proof guarantees the asymptotic security if $I(S_i; \boldsymbol{L}_i) < (2\ln 2)^{-1}$ holds for any i and d, which is a significantly stronger condition than existing reduction-based one. Independently, in [53], Masure *et al.* developed similar (slightly weaker) bound for additive masking not limited to Boolean by the *XOR lemma* [36]. Furthermore, in [7], Béguinot *et al.* derived another bound using the entropy power inequality [31, Theorem 17.7.3] over binary field, claimed to be stronger. These bounds are significantly stronger than those by reduction-based proofs, whereas the relation of these proofs to the NL and RP models remains undetermined.

4 Our adversarial models and notions

4.1 Communication channel and probing model

The communication channel model that represents SCA has been used for discussing the upper bound on SR [25, 45]. Figure 2 displays the communication channel model representing n-trace $(d+1)$-th order SCA on d-th order masked implementation. This model is generalized from that in [45] such that it does not depend on specific cipher, selection function, masking, or plaintext distribution. The symbols are defined as follows:

- $n \in \mathbb{N}$: Number of traces.
- $d \in \mathbb{N}$: Masking order.
- $i \in \mathbb{N}$: Index of masking shares ($0 \le i \le d$).
- $\mathbb{G}$: Field/ring of the masking. $\mathbb{G} = \mathbb{F}_2^l$ for l-bit Boolean masking and $\mathbb{G} = \mathbb{Z}_m$ for arithmetic masking, where m is the order (*i.e.*, period) of additive group.
- $K \in \mathbb{G}$: Secret key, which is assumed to be sampled from a uniform distribution over $\mathcal{K}$, but is fixed for one n-trace SCA.
- $Z^n \in \mathbb{G}$: Sequence of n secret intermediate variable targeted by the SCA. For example, $Z = \mathrm{Sbox}(K \oplus M)$ for major SCAs on AES, where M is plaintext.
- $(S_0, S_1, \dots, S_d)^n \in (\mathbb{G}^{d+1})^n$: Masking shares of Z. Note that $Z = S_0 + S_1 + \cdots + S_d$ holds and each S_i is uniformly distributed over $\mathcal{S} = \mathcal{Z} = \mathbb{G}$.
- $(\boldsymbol{L}_0, \boldsymbol{L}_1, \dots, \boldsymbol{L}_d)^n \in ((\mathbb{R}^\ell)^{d+1})^n$: Leakages of S_j. Each $\boldsymbol{L}_j$ is defined as an ℓ-dimensional vector over $\mathbb{R}$, where ℓ is the number of sampling points. As in literature, $\boldsymbol{L}_0, \boldsymbol{L}_1, \dots, \boldsymbol{L}_d$ are assumed to be mutually independent. Let $\boldsymbol{X} = (\boldsymbol{L}_0, \boldsymbol{L}_1, \dots, \boldsymbol{L}_d)$ be a side-channel trace, which is a $(d+1)$-dimensional vector of all leakages (namely, $\ell \times (d+1)$-dimensional vector in total).
- $\hat{K} \in \mathbb{G}$: Secret key estimated by the SCA. The adversary estimates the secret key from n traces $\boldsymbol{X}^n$ (and plaintexts M^n).

In this paper, we assume that $I(S_i; \boldsymbol{L}_i) > 0$ (equivalently, $p_{S_i|\boldsymbol{L}_i}$ is not a uniform distribution) for all i to exclude a case that masking is trivially secure. On this channel, the following theorem about SR holds as same as [25, 45].

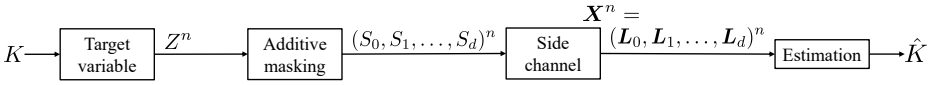


Fig. 2: Communication channel of n-trace SCA on d-th order additive masking.

Theorem 1 (Upper bound of SR by MI [25, 45]). *Let* SR *be the success rate of an optimal SCA. Let $I(Z; \boldsymbol{X})$ be the mutual information between Z and $\boldsymbol{X}$. If the number of traces is n and $\mathcal{Z} = \mathcal{K} = \mathbb{Z}_m$, the SR is upper-bounded as*

$$\xi(\mathrm{SR}) \leq nI(Z; \boldsymbol{X}), \tag{4}$$

where $\xi(r) : [0, 1] \to \mathbb{R}_0^+; r \mapsto H(K) - (1-r)\log(m-1) - H_2(r)$. Here, $H(K) = \log m$ is the entropy of K; and H_2 is the binary entropy function.

Proof. Despite the difference of communication channel model, Theorem 1 is proven in the same manner as [25, 45]. Owing to the communication channel in Figure 2, suppose that a Markov chain of $K \leftrightarrow Z^n \leftrightarrow \boldsymbol{X}^n \leftrightarrow \hat{K}$ holds. Here, the data processing inequality [31, Theorem 2.8.1] on this Markov chain states $I(K; \hat{K}) \leq I(Z^n; \boldsymbol{X}^n)$. Then, Fano's inequality [31, Theorem 2.10.1] states

$$\begin{aligned} I(K; \hat{K}) &= H(K) - H(K \mid \hat{K}) \\ &\geq H(K) - (1 - \mathrm{SR})\log(m-1) - H_2(\mathrm{SR}) = \xi(\mathrm{SR}). \end{aligned}$$

Because $I(Z^n; \boldsymbol{X}^n) = nI(Z; \boldsymbol{X})$ holds owing to the independence assumption, we conclude the inequality of (4) by combining these inequalities. □

In the inequality of (4), function ξ is strictly convex and takes the minimum of $\xi(1/m) = 0$ and maximum of $\xi(1) = \log m$, representing a random guess and full-bit key recovery, respectively. Based on Figure 2 and Equation (4), we formally define the security of cryptographic implementation.

Definition 4 (SCA security). *Consider the communication channel in Figure 2. Let μ be a non-negative real number. Let* SR *be the SR of n-trace SCA on a probing secure cryptographic implementation. The cryptographic implementation is (SR, n, μ)-SCA secure if it holds*

$$\xi(\mathrm{SR}) \leq n\mu.$$

We may write (SR, n, μ)-SCA security as μ-SCA security for short, as our analysis is mainly devoted to μ (see [67] for an analysis with regard to SR and n). From Theorem 1, any implementation is $I(Z; \boldsymbol{X})$-SCA secure. We should derive upper-bounds on $I(Z; \boldsymbol{X})$ to discuss the security of masked implementation.

When $d = 0$, the implementation is not masked; that is, $S_0 = Z$ and $\boldsymbol{L}_0 = \boldsymbol{X}$. When $d > 0$, the side-channel trace $\boldsymbol{X}$ is a vector comprising the leakages of shares (*i.e.*, $\boldsymbol{L}_i$) but does not directly contain the leakage of Z. To analyze the security of masking, we should discuss MI and probability distributions relating

to shares and their leakages, as the attacker obtains an information amount of $I(S_i; \boldsymbol{L}_i)$ for all i per side-channel trace on average. Therefore, our goal is to derive a non-trivial upper bound on $I(Z; \boldsymbol{X})$ by $I(S_i; \boldsymbol{L}_i)$ and masking order d. Such an attacker with leakage of $I(S_i; \boldsymbol{L}_i)$ is formally defined as follows:

Definition 5 (Adversarial notion by mutual information (MI) [34]). *Consider the communication channel in Figure 2. Let ϵ be a non-negative real number. A leakage (and an adversary with the leakage) are ϵ-MI if it holds*

$$\max_i I(S_i; \boldsymbol{L}_i) \leq \epsilon.$$

4.2 Formulation of d-probing using information-theoretic metric

For a d-probing secure implementation, any SCA of order up to d always fails (*i.e.*, SR $= 1/m$), as any collection of d shares brings no information about Z (*i.e.*, $I(Z; \boldsymbol{L}_0, \boldsymbol{L}_1, \ldots, \boldsymbol{L}_{i-1}, \boldsymbol{L}_{i+1}, \ldots, \boldsymbol{L}_d) = 0$ holds for any i) owing to the information-theoretic security of Shamir's secret sharing [64]. Therefore, we should discuss the SR of $(d+1)$-th order SCA on d-probing secure implementation to analyze its security. If we bound the SR of $(d+1)$-th order SCA on d-probing secure implementation when considering the leakages in Figure 2, it represents the adversarial advantage on it.

Thus, we say that a model/notion is reduced to d-probing model if SR of $(d+1)$-th order SCA can be upper-bounded accordingly. Our goal is to reduce the MI to the SCA notion for any d; namely, to bound μ by ϵ and the number of probe d (*i.e.*, masking order). This is formally stated as follows:

Definition 6 (Reduction to d-probing model). *Consider the communication channel in Figure 2. Let $\epsilon = \max_i I(S_i; \boldsymbol{L}_i)$ and d be the masking order. Let SR_d be the SR of n-trace $(d+1)$-th order SCA on d-probing secure implementation. The ϵ-MI security is reduced to d-probing model if ϵ-MI security is reduced to $f(\epsilon, d)$-SCA security for a function f; that is, a function f exists such that*

$$\xi(\mathrm{SR}_d) \leq n f(\epsilon, d).$$

Reduction from a notion/model other than MI is defined similarly.

Definition 6 implies that $(d+1)$-probing attackers on d-probing secure implementation obtains $\boldsymbol{L}_i$ by putting probes on S_i for all i. For a non-trivial reduction, f should be a decreasing function in d. More desirably, f should be negligible in d; for example, $f(\epsilon, d) = O(\epsilon^d)$ $(d \to \infty)$ holds for $0 \leq \epsilon < 1$.

Remark 1 (Leakages of encoding and computation). The communication channel represents SCAs focusing on the selection function output, although the real-world adversary may exploit all leakages during computation. This implies that our definition and analysis are devoted solely to *masking encoding*, although the probing model was originally presented for evaluating the security of *masked circuit* (see [34, Sections 3.2 and 3.3]). In other words, we here focus on the

smallest masking gadget—memory element storing shares. However, we stress that it is essential for masking as any implementation must use it. Furthermore, we develop non-trivial bound and proof in analyzing even this smallest one, which is followed by more complex analyses for practice. Extension of our modeling and bounds to circuits (*e.g.*, type 1–4 subsequences [59]) is an important future study.

Remark 2 (Physical defaults). Masking security order sometimes degrades due to share-interaction such as glitches [51], coupling [26], and microarchitectural interaction [40]. An easy way to consider such share-interaction with our models and analyses would be to replace the probing model with advanced ones, such as glitch-extended and robust probing models [38,60]. If share-interaction exists, interacted shares are not independent and considered as one share, indicating a security order degradation. For a given masked circuit, advanced probing models would capture the share-interaction as the degraded order. Thus, physical defaults would be reflected by evaluating the degraded order with our bounds.

4.3 Strong noisy leakage (SNL) model

As mentioned in Section 3, a major reason why the NL is significant would be that it is a lower bound on MI (*i.e.*, MI is reduced to NL). In other words, if we have another metric that is a lower bound on MI, we can discuss the security of masking based on it. A stronger bound would be desirable. Thus, we define the strong noisy leakage (SNL), yielding a stronger lower bound on MI than NL.

Definition 7 (Strong noisy leakage (SNL)). *Consider the communication channel in Figure 2. Let λ be a non-negative real number. A leakage is λ-strongly noisy and an adversary with the leakage is λ-SNL if it holds*

$$\max_i \mathbb{E}_{\boldsymbol{L}} \left[\left(\sum_s \left| p_{S_i|\boldsymbol{L}_i}(s \mid \boldsymbol{L}) - p_{S_i}(s) \right| \right)^2 \right] \leq \lambda.$$

Unlike Equation (2), here the squaring of the total variation distance is taken before the expectation. Theorem 2 states that SNL is an upper bound on NL and a lower bound on MI. Namely, Theorem 2 states a reduction from SNL to NL; that is, a λ-SNL adversary implies $\lambda^{1/2}$-NL, and also states a reduction from MI to SNL; namely, an ϵ-MI adversary implies $2\ln(2)\epsilon$-SNL.

Theorem 2 (Order of NL, SNL, and MI). *Consider the communication channel in Figure 2. Let $\delta = \max_i \mathbb{E}_{\boldsymbol{L}} \left[\sum_s \left| p_{S_i|\boldsymbol{L}_i}(s \mid \boldsymbol{L}) - p_{S_i}(s) \right| \right]$, $\lambda = \max_i \mathbb{E}_{\boldsymbol{L}} \left[\left(\sum_s \left| p_{S_i|\boldsymbol{L}_i}(s \mid \boldsymbol{L}) - p_{S_i}(s) \right| \right)^2 \right]$, and $\epsilon = \max_i I(S_i; \boldsymbol{L}_i)$. It holds*

$$\delta^2 \leq \lambda \leq 2\ln(2)\epsilon.$$

Proof. We first prove the right inequality. Pinsker's inequality states that

$$\sum_s \left| p_{S_i|\boldsymbol{L}_i}(s \mid \boldsymbol{L}) - p_{S_i}(s) \right| \leq \sqrt{2\ln(2) D_{\mathrm{KL}}(p_{S_i|\boldsymbol{L}_i} \parallel p_{S_i})},$$

Table 1: Comparison of distance-based adversarial notions

Notion	Definition	Major purpose/usage
Bias [59]	Expected L_2 distance	Analysis of classes of noisy leakage function and leaking computation
NL [33]	Expected L_1 distance	Reduction to RP
SNL (This study)	Expected squared L_1 distance	Fourier/Walsh–Hadamard analysis of $I(Z;\boldsymbol{X})$ using Parseval's theorem

for each i. Take squaring and expectation in terms of $\boldsymbol{L}$ of both hands, and then take max in terms of i. We conclude

$$\max_i \mathbb{E}_{\boldsymbol{L}}\left[\left(\sum_s \left|p_{S_i|\boldsymbol{L}_i}(s \mid \boldsymbol{L}) - p_{S_i}(s)\right|\right)^2\right] \le \max_i \mathbb{E}_{\boldsymbol{L}}\left[2\ln(2) D_{\mathrm{KL}}(p_{S_i|\boldsymbol{L}_i} \parallel p_{S_i})\right] = 2\ln(2)\max_i I(S_i;\boldsymbol{L}_i).$$

The equality holds if $p_{S_i|\boldsymbol{L}_i}(s \mid \boldsymbol{l})$ is uniform distribution over $\mathcal{S}$ for any $\boldsymbol{l}$.

The left inequality $\delta^2 \le \lambda$ is proven by Jensen's inequality [31, Theorem 2.6.2], because squaring is a convex function. The equality holds if $p_{S_i|\boldsymbol{L}_i}(s \mid \boldsymbol{l})$ is one-hot or uniform distribution over $\mathcal{S}$ for any $\boldsymbol{l}$. □

Comparison of distance-based notions. To highlight the significance of SNL, Table 1 lists distance-based adversarial notions in masking, where L_1 distance means distance between $p_{S_i|\boldsymbol{L}_i}$ and p_{S_i} induced by L_1 norm (same for L_2). As mentioned in [33], the choice of distance would not be essentially important in an asymptotic sense owing to the equivalence of norms over finite-dimensional space (*e.g.*, stated by [59, Equation (1)]). Nevertheless, the difference of definition would enable us to utilize distinct proof techniques and hence the strength of resulting bound would vary. In fact, the squaring in SNL is particularly unique and important, which directly relates the distance to LE and enables us to leverage Fourier/Walsh–Hadamard analysis *via* Parseval's theorem.

4.4 Leakage energy (LE) model

To prove the reduction from SNL to probing model by spectral analysis, we propose an intermediate notion named *leakage energy (LE)*, which bridges the SNL and probing model. In Section 5.1, we prove that the energy of the probability distribution is an upper bound on MI for *general* additive masking based on [45] (whereas it was proven only for Boolean masking in [45]). Here, Perceval's theorem states that the energy is expressed by its power spectrum in the frequency domain of DFT or WHT. Accordingly, we introduce LE for a formal treatment.

Definition 8 (Leakage energy (LE)). *Consider the communication channel in Figure 2. Let τ be a non-negative real number. A leakage and an adversary with the leakage are τ-LE if it holds*

$$\max_i \sum_{w\neq 0} \mathbb{E}_{\boldsymbol{L}}\left[\left|\hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L})\right|^2\right] = \max_i \sum_{w} \mathbb{E}_{\boldsymbol{L}}\left[\left|\hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L})\right|^2\right] - 1 \le \tau,$$

where $\hat{p}$ is m-point DFT for arithmetic masking and WHT for Boolean masking.

Note that the direct current component (*i.e.*, $w = 0$) is excluded from the LE because $\hat{p}_{S_i|\boldsymbol{L}_i}(0 \mid \cdot) = 1$ always holds owing to the law of total probability. In other words, the power spectral density at $w = 0$ is a constant offset and those at $w \neq 0$ are essential to evaluate the SR bound and the SCA security.

5 Reductions and security proof of masking

5.1 Reduction from LE to probing model

According to [45, Lemma 4.4], we first derive an upper bound on $I(Z; \boldsymbol{X})$ by the energy of probability distribution of $p_{S_i|\boldsymbol{L}_i}$ for general additive masking.

Lemma 1 (Upper bound on $I(Z; \boldsymbol{X})$ by energy). *Consider the communication channel in Figure 2. It holds*

$$I(Z; \boldsymbol{X}) \leq \log \left(\sum_{w} \prod_{i=0}^{d} \mathbb{E}_{\boldsymbol{L}} \left[\left| \hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L}) \right|^2 \right] \right), \tag{5}$$

where $\hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L})$ is m-point DFT for arithmetic masking and $\hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L})$ is WHT for Boolean masking.

Proof. For Boolean masking, Lemma 1 is proven in [45, Lemma 4.4], where $\hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L})$ is a WHT. We prove the case of arithmetic masking by replacing the WHT with m-point DFT. As in the proof of [45, Lemma 4.4], we upper-bound $I(Z; \boldsymbol{X})$ by the energy of $p_{Z|\boldsymbol{X}}$. According to Jensen's inequality, we have

$$\begin{aligned} I(Z; \boldsymbol{X}) &= H(Z) - H(Z \mid \boldsymbol{X}) = \log m + \mathbb{E}_{Z,\boldsymbol{X}} \log p_{Z|\boldsymbol{X}}(Z \mid \boldsymbol{X}) \\ &\leq \log \left(m \mathbb{E}_{Z,\boldsymbol{X}} p_{Z|\boldsymbol{X}}(Z \mid \boldsymbol{X}) \right) = \log \left(m \mathbb{E}_{\boldsymbol{X}} \left[\sum_{z} \left(p_{Z|\boldsymbol{X}}(z \mid \boldsymbol{X}) \right)^2 \right] \right). \end{aligned} \tag{6}$$

Parseval's theorem states that the energy can be expressed using the power spectrum of the signal (*i.e.*, $|\hat{p}_{Z|\boldsymbol{X}}(w \mid \boldsymbol{X})|^2$); namely, it holds

$$\sum_{z} \left(p_{Z|\boldsymbol{X}}(z \mid \boldsymbol{X}) \right)^2 = \frac{1}{m} \sum_{w} \left| \hat{p}_{Z|\boldsymbol{X}}(w \mid \boldsymbol{X}) \right|^2. \tag{7}$$

Here, $Z = S_0 + S_1 + \cdots + S_d$ holds by the definition of arithmetic masking. Recall that the probability distribution of sum of random variables is given by convolution of their probability distributions, which means that it holds

$$p_{Z|\boldsymbol{X}}(z \mid \boldsymbol{X}) = \sum_{z = s_0 + s_1 + \cdots + s_d} \prod_{i=0}^{d} p_{S_i|\boldsymbol{L}_i}(s_i \mid \boldsymbol{L}_i). \tag{8}$$

The right-hand side is cyclic convolution with a period of m. Using m-point DFT, in the frequency domain, Equation (8) is equivalently expressed by

$$\hat{p}_{Z|\boldsymbol{X}}(w \mid \boldsymbol{X}) = \prod_{i=0}^{d} \hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L}_i). \tag{9}$$

By combining Equations (6), (7) and (9), we conclude Equation (5). □

Using Theorem 1 and Lemma 1, we prove the reduction from LE to d-probing model as Theorem 3.

Theorem 3 (Reduction from LE to d-probing). *Consider the communication channel in Figure 2. Let* $\tau = \max_i \sum_{w\neq 0} \mathbb{E}_{\boldsymbol{L}}\left[\left|\hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L})\right|^2\right]$. *A* τ*-LE secure implementation of* d*-th order masking implies* $\log(1+\tau^{d+1})$*-SCA security. Namely, the SR of* n*-trace SCA on* d*-probing secure masked implementation, denoted by* SR_d*, is bounded from above as*

$$\xi(\mathrm{SR}_d) \leq n \log\left(1+\tau^{d+1}\right).$$

Proof. From Theorem 1 and Lemma 1, with Jensen's inequality, we conclude

$$\begin{aligned}\xi(\mathrm{SR}_d) &\leq n \log\left(\sum_w \prod_{i=0}^{d} \mathbb{E}_{\boldsymbol{L}}\left[\left|\hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L})\right|^2\right]\right) \\ &\leq n \log\left(1+\left(\max_i \sum_{w\neq 0} \mathbb{E}_{\boldsymbol{L}}\left[\left|\hat{p}_{S_i|\boldsymbol{L}}(w \mid \boldsymbol{L}_i)\right|^2\right]\right)^{d+1}\right) = n\log\left(1+\tau^{d+1}\right),\end{aligned}$$

as required. □

5.2 Reduction from SNL to LE

We then demonstrate the reduction from SNL to LE using Lemma 2, which states that the absolute amplitude of DFT/WHT is bounded from above by the total variation distance between $p_{S_i|\boldsymbol{L}_i}$ and p_{S_i}. We provide a novel proof leveraging the characteristics of DFT/WHT.

Lemma 2 (Upper bound on DFT/WHT absolute amplitude by total variation distance). *Consider Figure 2. For any* i *and* $\boldsymbol{L}$*, it holds that*

$$\begin{cases} \left|\hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L})\right| = 1 & \text{if } w = 0, \\ \left|\hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L})\right| \leq \sum_{s=0}^{m-1} \left|p_{S_i|\boldsymbol{L}_i}(s \mid \boldsymbol{L}) - p_{S_i}(s)\right| & \text{if } w \neq 0. \end{cases}$$

Proof. The proof does not depend on i. We first consider an arithmetic masking with a period of m. Recall that DFT $\hat{p}_{S_i|\boldsymbol{L}_i}$ is defined as

$$\hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L}) = \sum_{s=0}^{m-1} p_{S_i|\boldsymbol{L}_i}(s \mid \boldsymbol{L}) \mathrm{e}^{-\mathrm{i}s\Omega_w},$$

where $\Omega_w = 2\pi w/m$ is the normalized frequency. When $w = 0$, it holds that $\hat{p}_{S_i|\boldsymbol{L}_i}(0 \mid \boldsymbol{L}) = \sum_{s=0}^{m-1} p_{S_i|\boldsymbol{L}_i}(s \mid \boldsymbol{L}) = 1$ according to the law of total probability.

Consider the case that $w \neq 0$. According to the characteristic of twiddle factor (*i.e.*, a non-trivial root of unity solving $x^m - 1 = 0$), $\sum_{s=0}^{m-1} \mathrm{e}^{-\mathrm{i}s\Omega_w} = 0$

holds for any $w \neq 0$. When $w \neq 0$, the DFT is equivalent to

$$\begin{aligned}\hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L}) &= \sum_{s=0}^{m-1} p_{S_i|\boldsymbol{L}_i}(s \mid \boldsymbol{L}) \mathrm{e}^{-\mathrm{i}s\Omega_w} - \frac{1}{m}\sum_{s=0}^{m-1} \mathrm{e}^{-\mathrm{i}s\Omega_w} \\ &= \sum_{s=0}^{m-1} \left(p_{S_i|\boldsymbol{L}_i}(s \mid \boldsymbol{L}) - \frac{1}{m} \right) \mathrm{e}^{-\mathrm{i}s\Omega_w}.\end{aligned}$$

Here, p_{S_i} is supposedly a uniform distribution over $\mathcal{S}$; namely, $p_{S_i}(s) = 1/|\mathcal{S}| = 1/m$ holds for any $s \in \mathcal{S}$. Hence, according to the triangle inequality over the complex field, $|\hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L})|$ is upper-bounded as

$$\left|\hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L})\right| \leq \sum_{s=0}^{m-1} \left| \left(p_{S_i|\boldsymbol{L}_i}(s \mid \boldsymbol{L}) - \frac{1}{m} \right) \mathrm{e}^{-\mathrm{i}s\Omega_w} \right| = \sum_{s=0}^{m-1} \left| p_{S_i|\boldsymbol{L}_i}(s \mid \boldsymbol{L}) - p_{S_i}(s) \right|,$$

which completes the proof. In the case of Boolean masking (*i.e.*, $\hat{p}_{S_i|\boldsymbol{L}_i}$ is defined using WHT), it is proven by replacing the twiddle factor to $(-1)^{\langle s|w \rangle}$. □

Remark 3 (Meaning of the proof from Fourier analysis). A uniform distribution (*i.e.*, constant function) in the time domain corresponds to the delta function in the frequency domain. Owing to the linearity of DFT (and WHT), this means that subtracting a uniform distribution from a signal has no impact on DFT and WHT (*i.e.*, $\hat{p}_{S_i}(w) = 0$) when $w \neq 0$, but has an impact on the direct current at $w = 0$. Thus, by decomposing the probability distribution into the frequency domain, when $w \neq 0$, we discuss a signal of

$$\left|\hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L})\right| = \left|\hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L}) - \hat{p}_{S_i}(w)\right| = \left|\mathrm{DFT}\left[p_{S_i|\boldsymbol{L}_i}(\cdot \mid \boldsymbol{L}) - p_{S_i}(\cdot)\right](w)\right|,$$

which corresponds to the total variation distance in the NL and SNL.

Using Lemma 2, we prove the security reduction from SNL to LE.

Theorem 4 (Reduction from SNL to LE). *Consider the communication channel in Figure 2. Let* $\lambda = \max_i \sum_s \mathbb{E}_{\boldsymbol{L}}[(|p_{S_i|\boldsymbol{L}_i}(s \mid \boldsymbol{L}) - p_{S_i}(s)|)^2]$. *An* $(m-1)\lambda$*-LE adversary implies* λ*-SNL adversary. Namely, it holds*

$$\max_i \sum_{w \neq 0} \mathbb{E}_{\boldsymbol{L}} \left[\left|\hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L})\right|^2 \right] \leq (m-1)\lambda.$$

Proof. From Lemma 2, for any i, we have

$$\sum_{w \neq 0} \mathbb{E}_{\boldsymbol{L}} \left[\left|\hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L})\right|^2 \right] \leq (m-1)\mathbb{E}_{\boldsymbol{L}} \left[\left(\sum_{s=0}^{m-1} \left| p_{S_i|\boldsymbol{L}_i}(s \mid \boldsymbol{L}) - p_{S_i}(s) \right| \right)^2 \right].$$

Take max in terms of i. This completes the proof. □

5.3 Reduction from MI/SNL to d-probing

Theorem 5 (Reduction from MI/SNL to d-probing). *Consider the communication channel in Figure 2. Let* $\epsilon = \max_i I(S_i; \boldsymbol{L}_i)$ *and* $\lambda = \max_i \sum_s \mathbb{E}_{\boldsymbol{L}}[|p_{S_i \boldsymbol{L}_i}(s \mid \boldsymbol{L}) - p_{S_i}(s)|^2]$. *A* λ*-SNL secure and an* ϵ*-MI secure implementations with* d*-th order masking in probing model are* $\log(1 + (m-1)\lambda^{d+1})$*-SCA secure and* $\log(1 + (m-1)(2\ln 2)^{d+1}\epsilon^{d+1})$*-SCA secure, respectively. Namely, the SR of* n*-trace* $(d+1)$*-th order SCA on* d*-th order masking represented in Figure 2, denoted by* SR_d*, is bounded from above as*

$$\begin{aligned}\xi(\mathrm{SR}_d) &\le n \log\left(1 + (m-1)\lambda^{d+1}\right) \\ &\le n \log\left(1 + (m-1)(2\ln 2)^{d+1}\epsilon^{d+1}\right). \end{aligned} \tag{10}$$

Proof. It is obvious from Theorems 1 and 2 and Lemmas 1 and 2.[9] □

Note that the derived SR bound by MI is the same as [45, Lemma 4.6] for Boolean masking; however, our new proof using Fourier analysis generalizes the result to arithmetic masking schemes and clarifies the relationship between the communication channel in Figure 2 and masking security/adversarial notions including the NL through our security reductions.

Tightness. Our reductions are tight. In the proofs of Theorem 1 and Lemmas 1 and 2, we used Fano's, Jensen's, and triangle inequalities, respectively, all of which hold with equality if $p_{S_i|\boldsymbol{L}_i}(\cdot \mid \boldsymbol{l})$ is one-hot distribution for any i and $\boldsymbol{l}$. Moreover, the reduction from MI to SNL is also tight if $p_{S_i|\boldsymbol{L}_i}(\cdot \mid \boldsymbol{l}) = p_{S_i}$; that is, $p_{S_i|\boldsymbol{L}_i}(\cdot \mid \boldsymbol{l})$ is a uniform distribution over $\mathcal{S}$ for any i and $\boldsymbol{l}$.

5.4 Asymptotic and provable security of masking

Similar to [45], we derive the provable security of arithmetic masking (not limited to Boolean) in the LE, SNL, and MI notions, namely an exponential decay of SR (*i.e.*, side-channel adversarial advantage) by an increase of d, as follows:

Theorem 6 (Asymptotic security of masking in LE notion). *Let* SR_d *be the SR of* $(d+1)$*-th order SCA on* d*-th order masking represented in Figure 2. For any* d*, let* τ_{peak} *be a positive real number satisfying*

$$\max_i \max_{w \neq 0} \mathbb{E}_{\boldsymbol{L}}\left[\left|\hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L}_i)\right|^2\right] < \tau_{\mathrm{peak}}.$$

If $\tau_{\mathrm{peak}} < 1$*, it holds*

$$\left|\mathrm{SR}_d - \frac{1}{m}\right| < \sqrt{\frac{2n(m-1)\log \mathrm{e}}{\xi_c}}\tau_{\mathrm{peak}}^{d/2} = O(\tau_{\mathrm{peak}}^{d/2}) \quad (d \to \infty),$$

for a constant positive real number ξ_c.

[9] If we apply Lemma 2 to Theorem 3, the resulting bound is $\xi(\mathrm{SR}_d) \le n \log\left(1 + (m-1)^{d+1}\lambda^{d+1}\right)$, which is significantly looser than the bound in (10). This is because the proof of Theorem 3 uses Jensen's inequality to interchange summation in terms of w and exponentiation by d, which incurs a non-negligible loss. By contrast, the bound in (10) is derived directly from Lemma 1 without the interchange.

Proof. It is proven in the same manner as [45, Lemma 5.1 and Theorem 5.2] with Theorem 3, using the convexity of ξ and its second-order Taylor expansion. □

5.5 Provable security condition in NL and MI

Following from Theorem 6, we give the condition for provable security of Boolean and arithmetic maskings in the NL and MI notions, which states that the conjecture by Dziembowski *et al.* [36] in TCC 2016 is true[10]. Note again that δ is divided by two because of its definition in this paper.

Theorem 7 (A sufficient condition for asymptotic security of additive masking in NL and MI). *Consider the communication channel in Figure 2. Let* $\delta = \max_i \mathbb{E}_{\boldsymbol{L}}[\sum_s |p_{S_i|\boldsymbol{L}_i}(s \mid \boldsymbol{L}) - p_{S_i}(s)|]$, *and* $\epsilon = \max_i I(S_i; \boldsymbol{L}_i)$. *A sufficient condition for* $\mathrm{SR}_d - 1/m = O(\tau_{\mathrm{peak}}^{d/2})$ $(d \to \infty)$ *for some* $\max_{i,w\neq 0} \mathbb{E}_{\boldsymbol{L}}[|\hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L}_i)|^2] < \tau_{\mathrm{peak}} < 1$ *is given by*

$$\frac{\delta}{2} < 1 - \frac{|\mathbb{H}|}{|\mathbb{G}|} = 1 - \max_{1\leq w\leq m-1} \frac{g_w}{m}, \tag{11}$$

$$\epsilon < \min_{1\leq w\leq m-1} \log \frac{m}{g_w}, \tag{12}$$

where $\mathbb{G}$ *is the additive group of field/ring;* $\mathbb{H}$ *is the largest subgroup of* $\mathbb{G}$*;* m *is the cardinality of* $\mathbb{G}$*; and* g_w *is the greatest common divisor (GCD) of* w *and* m*.*

Satisfying either inequality (11) or (12) is sufficient. Let $\gamma = \min_w m/g_w$ be the non-trivial divisor of m (if m is a prime, $\gamma = m$). For arithmetic masking, the largest subgroup is $\mathbb{H} = \mathbb{G}/\gamma\mathbb{Z}$, where $|\mathbb{H}| \,/\, |\mathbb{G}| = 1/\gamma$ according to Lagrange's theorem. For Boolean masking, the largest subgroup is $\mathbb{H} = \mathbb{G}/\{s \mid \langle s \mid w\rangle = 0\}$ for all $w \neq 0$. In this case, Lagrange's theorem states that $|\mathbb{H}| \,/\, |\mathbb{G}| = 1/2$ holds for any $w \neq 0$. To prove Theorem 7, we first introduce Lemma 3.

Lemma 3 (Maximum of power spectral density). *Let* $p : \mathcal{Y} \to \mathbb{R}$ *be a probability mass function, where* $\mathcal{Y}$ *is a finite discrete set. Let* $m' = |\mathcal{Y}|$*. Its power spectral densities are upper-bounded by one; namely,* $|\hat{p}(w)|^2 \leq 1$ *holds for any* w*. When* m' *is prime, the equality holds for any* w *if and only if* p *is one-hot distribution, which is defined by*

$$p(y) = \begin{cases} 1 & \text{if } y = y' \text{ for a specific } y', \\ 0 & \text{otherwise.} \end{cases}$$

Proof. Owing to the definition of DFT, it holds

$$|\hat{p}(w)|^2 = \left| \sum_{s=0}^{m'-1} p(s) \mathrm{e}^{-\mathrm{i}s\Omega_w} \right|^2 \leq \left(\sum_{s=0}^{m'-1} |p(s)| \right)^2 = 1,$$

[10] The original open problem found in TCC 2016 was whether the noise requirement of $\delta/2 < 1 - |\mathbb{H}|/|\mathbb{G}|$ is sufficient for *arithmetic* masking with *non-prime* m.

according to the triangle inequality and the assumption of $\sum_{y\in\mathcal{Y}} p(y) = 1$ (*i.e.*, the law of total probability). The equality condition of the triangle inequality over $\mathbb{C}$ is that all non-zero numbers have an identical phase (*i.e.*, $\mathrm{e}^{-\mathrm{i}s\Omega_w}$). Hence, $|\hat{p}(w)|^2 = 1$ holds for co-prime pair of w and m' if and only if p is one-hot distribution, because all the twiddle factors are different when m' and w are co-prime. If m' is a prime, w and m' are always co-prime. This completes the proof. This is the same for WHT. □

We then prove Theorem 7 by decomposing the DFT/WHT into smaller ones with prime points to utilize Lemma 3.

Proof. We first prove that $\delta/2 < 1 - \min_{w\neq 0} g_w/m$ is sufficient for arithmetic masking. Let $\gamma_w = m/g_w$. Like FFT, at frequency w, the m-point DFT of $p_{S_i|\boldsymbol{L}_i}$ is decomposed to sum of g_w γ_w-point DFTs as

$$\begin{aligned}\hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L}) &= \sum_{s_1=0}^{g_w-1}\sum_{s_0=0}^{\gamma_w-1} p_{S_i|\boldsymbol{L}_i}(s_0+\gamma_w s_1 \mid \boldsymbol{L})\mathrm{e}^{-\mathrm{i}s_0\Omega_w} \\ &= \sum_{s_0=0}^{\gamma_w-1} \mathrm{e}^{-\mathrm{i}s_0\Omega'_w}\sum_{s_1=0}^{g_w-1} p_{S_i|\boldsymbol{L}_i}(s_0+\gamma_w s_1 \mid \boldsymbol{L}), \end{aligned} \tag{13}$$

where $\Omega'_w = 2\pi w/\gamma_w$ is the normalized frequency of the γ_w-point DFT. For a given w, let $Y_i^{(w)} = S_i \bmod \gamma_w$ be a random variable on $\mathbb{Z}_m/\gamma_w\mathbb{Z} = \mathbb{Z}_{\gamma_w}$, which implies that $p_{Y_i^{(w)}}(y) = \sum_{s_1=0}^{g_w-1} p_{S_i}(y+\gamma_w s_1)$. The DFT at w with g_w is expressed by a γ_w-point DFT as $\hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L}) = \hat{p}_{Y_i^{(w)}|\boldsymbol{L}_i}(w \mid \boldsymbol{L})$. According to Theorem 6, a sufficient condition for provably secure masking is

$$\max_{i,w\neq 0}\left|\hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L})\right|^2 = \max_{i,w\neq 0}\left|\hat{p}_{Y_i^{(w)}|\boldsymbol{L}_i}(w \mid \boldsymbol{L})\right|^2 < 1 \tag{14}$$

Because γ_w is co-prime to w, Lemma 3 states that the power spectral density reaches its maximum value of one if $p_{Y_i^{(w)}|\boldsymbol{L}_i}(\cdot \mid \boldsymbol{L})$ is a one-hot distribution for any i and $\boldsymbol{L}$, which is the farthest from uniform distribution. Thus, a necessary condition to meet Equation (14) is that δ should be smaller than the total variation distance between one-hot and uniform distributions for any $w \neq 0$. If $p_{Y_i^{(w)}|\boldsymbol{L}_i}$ is one-hot distribution for any $\boldsymbol{l}$, the total variation distance is

$$\begin{aligned}\sum_{y=0}^{\gamma_w-1}\mathbb{E}_{\boldsymbol{L}}\left[\left|p_{Y_i^{(w)}|\boldsymbol{L}_i}(y \mid \boldsymbol{L}) - p_{Y_i}(y)\right|\right] &= \left|1-\frac{1}{\gamma_w}\right| + (\gamma_w-1)\left|\frac{-1}{\gamma_w}\right| \\ &= 2-\frac{2g_w}{m}. \end{aligned} \tag{15}$$

Thus, the condition of δ for the asymptotic security is that δ is smaller than the maximum of Equation (15) in terms of $w \neq 0$ (as in Equation (14)); that is,

$$\delta < \max_{1\leq w\leq m-1}\left(2-\frac{2g_w}{m}\right) = 2 - \min_{1\leq w\leq m-1}\frac{2g_w}{m}.$$

Divide both sides by two. This completes the proof of condition (11).

Then, we consider the MI. As aforementioned, a necessary and sufficient condition of $\max_{i,w\neq 0} |\hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L})|^2 < 1$ is that $p_{Y_i^{(w)}|\boldsymbol{L}_i}$ is not a one-hot distribution for any $w \neq 0$. Recall that one-hot distribution maximizes the MI, because $I(Y_i^{(w)}; \boldsymbol{L}_i) = H(Y_i^{(w)}) - H(Y_i^{(w)} \mid \boldsymbol{L}_i)$ and $H(Y_i^{(w)} \mid \boldsymbol{L}_i)$ takes zero if and only if $Y_i^{(w)}$ is an injective function of $\boldsymbol{L}_i$ (*i.e.*, $p_{Y_i^{(w)}|\boldsymbol{L}_i}(y \mid \boldsymbol{l}) > 0$ only for a specific y). Hence, the condition for asymptotic security is $H(Y_i^{(w)} \mid \boldsymbol{L}_i) > 0$ for any w, which implies $I(Y_i^{(w)}; \boldsymbol{L}_i) < H(Y_i^{(w)}) = \log \gamma_w$.

Here, a Markov chain of $\boldsymbol{L}_i \leftrightarrow S_i \leftrightarrow Y_i^{(w)}$ is supposed to hold in the communication channel of Figure 2 because $\boldsymbol{L}_i$ depends on S_i and $Y_i^{(w)}$ is a function of S_i. Markov chain of $\boldsymbol{L}_i \leftrightarrow S_i \leftrightarrow Y_i^{(w)}$ states a data processing inequality of $I(Y_i^{(w)}; \boldsymbol{L}_i) \leq I(S_i; \boldsymbol{L}_i)$. Hence, it suffices for the asymptotic security to hold that, $I(S_i; \boldsymbol{L}_i) < \min_w \log \gamma_w$, which follows from the condition that $I(Y_i^{(w)}; \boldsymbol{L}_i) < \log \gamma_w$ for any w. This completes the proof of condition (12).

This proof holds for Boolean masking by replacing DFT to WHT, where we define $Y_i^{(w)} = \langle S_i \mid w \rangle$ according to the definition of its twiddle factor, as in [45]. Note that in this case, g_w is replaced with 2^{l-1} because any w represents a subgroup of $\mathbb{F}_2^l$ with period of two, resulting in the upper-bound requirement of $\delta/2 < 1 - 1/2 = 1/2$ and $\max_i I(S_i; \boldsymbol{L}_i) < 1$. □

The condition of (12) corresponds to $I(S_i; \boldsymbol{L}_i) < 1$ for Boolean and arithmetic maskings with even characteristics, while the bound is $I(S_i; \boldsymbol{L}_i) < \log m$ for prime-field masking. This condition is the strongest among existing masking proofs; that is, $I(S_i; \boldsymbol{L}) < 2^{-2l+1}$ using a reduction-based proof through RP [34], $I(S_i; \boldsymbol{L}_i) < (2 \ln 2)^{-1} \approx 0.72$ using communication-channel-based proofs [7, 45], and $I(S_i; \boldsymbol{L}_i) < 1$ in [47] limited to Boolean masking.

Meaning of $Y_i^{(w)}$ and subgroup leakage. The random variable $Y_i^{(w)}$ represents a residue class defined by the subgroup $\mathbb{Z}_{\gamma_w}$ where S_i belongs. The proof of Theorem 7 implies that if the attacker could estimate the subgroup to which each share belongs from leakage with probability 1 (*i.e.*, if the leakage contained no noise), the masking could not be secure, because attacker can uniquely determine the subgroup of Z from the leakages of shares and can trivially filter the key candidates. Otherwise (*i.e.*, if the leakage contains a non-zero noise no matter how small it is), the masking can be asymptotically secure. Generally, the security of masking heavily depends on such *subgroup leakage*—how accurately attacker can estimate the subgroup to which each share belong. Prime-field masking induces no non-trivial subgroup, exhibiting stronger security.

Relationship between subgroup and real-world leakages. If the subgroup leakage corresponds to the characteristics of real-world leakage, the masking would be less secure. Boolean masking induces the largest subgroup for any $w \neq 0$. For given $w \neq 0$, the corresponding subgroups consists of elements with $\langle S_i \mid w \rangle = 0$ or 1. An arithmetic masking induces the largest subgroup at

$w = m/\gamma$. Generally, it holds that

$$\hat{p}_{S_i|\boldsymbol{L}_i}(2^l - 1 \mid \boldsymbol{L}_i) = \Pr[\langle S_i \mid w \rangle = 0 \mid \boldsymbol{L}_i] - \Pr[\langle S_i \mid w \rangle = 1 \mid \boldsymbol{L}_i], \tag{16}$$

$$\hat{p}_{S_i|\boldsymbol{L}_i}(m/\gamma \mid \boldsymbol{L}_i) = \sum_{s=0}^{\gamma-1} \mathrm{e}^{-2s\pi\mathrm{i}/\gamma} \Pr[S_i \bmod \gamma = s \mid \boldsymbol{L}_i], \tag{17}$$

for Boolean and arithmetic masking, respectively. For the instances, it holds

$$\hat{p}_{S_i|\boldsymbol{L}_i}(2^l - 1 \mid \boldsymbol{L}_i) = \Pr[\mathrm{HW}(S_i) \text{ is even} \mid \boldsymbol{L}_i] - \Pr[\mathrm{HW}(S_i) \text{ is odd} \mid \boldsymbol{L}_i],$$
$$\hat{p}_{S_i|\boldsymbol{L}_i}(m/2 \mid \boldsymbol{L}_i) = \Pr[\mathrm{LSB}(S_i) = 0 \mid \boldsymbol{L}_i] - \Pr[\mathrm{LSB}(S_i) = 1 \mid \boldsymbol{L}_i],$$

for Boolean masking and for arithmetic masking with even characteristic, respectively, where HW is the Hamming weight and LSB is the least significant bit. The power spectrum for Boolean masking has an explicit relation to HW, so Boolean masking would be weak particularly under HW leakage. In contrast, the LSB is unlikely to leak without *algorithmic* noise (for devices with HW leakage), as cryptographic implementations do not usually process LSB of variable alone. Thus, arithmetic masking can be more secure in practice, as no subgroup leakage corresponds to HW. Numerical evaluation in Section 6 validates our claim.

5.6 Necessary and sufficient condition

An important question is whether the condition $\epsilon < \log \gamma$ (where $\gamma = \min_w(m/g_w)$) can be strengthened. The answer is no if no assumption about leakage is used. Lemma 4 states that $I(Z; \boldsymbol{X})$ does not converges to zero as $d \to \infty$ under subgroup leakages with probability 1 such that $I(Y_i^{(w)}; \boldsymbol{L}_i) = \log \gamma_w$ (equivalently, $\mathbb{E}_{\boldsymbol{L}}[|\hat{p}_{Y_i^{(w)}|\boldsymbol{L}_i}(w \mid \boldsymbol{L})|^2] = 1$) holds for some w for all i.

Lemma 4 (Limit of $I(Z; \boldsymbol{X})$ as $d \to \infty$). *Consider Figure 2. Let $\mathcal{W}_{\mathrm{peak}} = \{\, w \mid w \neq 0 \wedge \forall d, i, I(Y_i^{(w)}; \boldsymbol{L}_i) = \log \gamma_w \,\}$, where $\gamma_w = 2$ and m/g_w for Boolean and arithmetic masking (g_w is the GCD of m and w). It holds that*

$$I(Z; \boldsymbol{X}) \to \log\left(1 + \left|\mathcal{W}_{\mathrm{peak}}\right|\right),$$
$$I(Z; \boldsymbol{X}) \to \log\left(1 + \left|\mathcal{W}_{\mathrm{peak}}\right|\right) = \begin{cases} 0 & \text{if } \mathcal{W}_{\mathrm{peak}} = \emptyset, \\ \max\limits_{w \in \mathcal{W}_{\mathrm{peak}}} \log \gamma_w & \text{otherwise,} \end{cases}$$

as $d \to \infty$ for Boolean and arithmetic maskings, respectively.

Proof. We first prove the limit of both maskings. Recall that $\mathbb{E}_{\boldsymbol{L}}[|\hat{p}_{Y_i^{(w)}|\boldsymbol{L}_i}(w \mid \boldsymbol{L})|^2] < 1$ holds for $w \neq 0$ if $I(Y_i^{(w)}; \boldsymbol{L}_i) \neq \log \gamma_w$. According to Lemma 1, upper bound on $I(Z; \boldsymbol{X})$ converges as

$$I(Z; \boldsymbol{X}) \leq \log\left(1 + \sum_{w \neq 0} \prod_{i=0}^{d} \mathbb{E}_{\boldsymbol{L}}\left[\left|\hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L})\right|^2\right]\right)$$
$$\to \log\left(1 + \sum_{w \in \mathcal{W}_{\mathrm{peak}}} 1\right) = \log\left(1 + \left|\mathcal{W}_{\mathrm{peak}}\right|\right) \quad \text{as} \quad d \to \infty. \tag{18}$$

We then consider the limit of lower bound. For each i, let Q_i be a random variable over $\mathbb{R}$ such that, for all w,

$$\hat{p}_{S_i|Q_i}(w \mid Q) = \begin{cases} 1 & \text{if } \hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L}) = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Q_i is a leakage degraded from $\boldsymbol{L}_i$, which discards information of subgroup leakages with probability less than 1. By definition, $p_{Y_i^{(w)}|Q_i}$ is one-hot distribution. Let $\boldsymbol{X}' = (Q_i)_{i=0}^d$. As $\boldsymbol{X}'$ is derived from $\boldsymbol{X}$, a Markov chain of $Z \leftrightarrow \boldsymbol{X} \leftrightarrow \boldsymbol{X}'$ supposedly holds, and its data processing inequality states $I(Z;\boldsymbol{X}') \le I(Z;\boldsymbol{X})$. Lemma 1 also states an upper bound on $I(Z;\boldsymbol{X}')$ with equality, because $Y_i^{(w)}$ is a constant for given Q_i and $\bigoplus_i Y_i^{(w)} = \langle Z \mid w \rangle$ and $\sum_i Y_i^{(w)} = Z \bmod \gamma_w$ for Boolean and arithmetic masking, respectively, which are the equality condition of Jensen's inequality on $Y_i^{(w)}$. Then, it holds

$$I(Z;\boldsymbol{X}') = \log\left(1 + \sum_{w\neq 0}\prod_{i=0}^{d} \mathbb{E}_Q\left[\left|\hat{p}_{Y_i^{(w)}|Q_i}(w \mid Q)\right|^2\right]\right) \to \log\left(1 + \left|\mathcal{W}_{\text{peak}}\right|\right), \tag{19}$$

as $d \to \infty$ with equality. By combining Equations (18) and (19) with the data processing inequality, we have $I(Z;\boldsymbol{X}) \to \log(1 + \left|\mathcal{W}_{\text{peak}}\right|)$.

We then consider arithmetic masking. If $\mathcal{W}_{\text{peak}} = \emptyset$, the equality is trivially proven. Consider the case of $\mathcal{W}_{\text{peak}} \neq \emptyset$. If $I(Y_i^{(w)};\boldsymbol{L}_i) = \log \gamma_w$ for w, $\hat{p}_{Y_i^{(cg_w)}|\boldsymbol{L}_i}(w \mid \boldsymbol{L}) = 1$ holds for all $c \in \mathbb{N}$ such that $0 < cg_w < m$, owing to the periodicity of DFT. More precisely, if $I(Y_i^{(w)};\boldsymbol{L}_i) = \log \gamma_w$ holds for w, then $\hat{p}_{Y_i^{(w)}|\boldsymbol{L}_i}(w \mid \boldsymbol{L}) = 1$, meaning that all non-zero components ($p_{S_i|\boldsymbol{L}_i}(s \mid \boldsymbol{L}) > 0$) have an identical phase. In Equation (13), the term $\sum_{s_1=0}^{g_w-1} p_{S_i|\boldsymbol{L}_i}(s_0 + \gamma_w s_1 \mid \boldsymbol{L}) = p_{Y_i^{(w)}|\boldsymbol{L}_i}(s_0 \mid \boldsymbol{L})$ equals to 1 at a specific s_0 ($0 \le s_0 \le \gamma_w$) and equals to 0 at other s_0. Subsequently, $\hat{p}_{Y_i^{(cg_w)}|\boldsymbol{L}_i}(w \mid \boldsymbol{L}) = 1$ holds for all c, as $\hat{p}_{Y_i^{(cg_w)}|\boldsymbol{L}_i}$ has $p_{Y_i^{(w)}|\boldsymbol{L}_i}(s_0 \mid \boldsymbol{L})$ at a phase. The number of such c is $\max_{w\in\mathcal{W}_{\text{peak}}} \gamma_w - 1$, which is followed by $1 + \left|\mathcal{W}_{\text{peak}}\right| = \max_{w\in\mathcal{W}_{\text{peak}}} \gamma_w$. This completes the proof. □

Corollary 1 shows important instantiations of Lemma 4.

Corollary 1 (Examples of subgroup leakages with probability 1).

- *(Boolean masking) If* $\boldsymbol{L}_i = \mathrm{LSB}(\mathrm{HW}(S_i))$ *for all* i, $I(S_i;\boldsymbol{L}_i) = 1$ *and* $I(Z;\boldsymbol{X}) \to 1$ *as* $d \to \infty$, *where* LSB *corresponds to the parity.*
- *(Arithmetic masking with even characteristic) If* $\boldsymbol{L}_i = \mathrm{LSB}(S_i)$ *for all* i, $I(S_i;\boldsymbol{L}_i) = 1$ *and* $I(Z;\boldsymbol{X}) \to 1$ *as* $d \to \infty$.
- *(Arithmetic masking with arbitrary characteristic) If* $\boldsymbol{L}_i = S_i \bmod \gamma$ *for all* i, $I(S_i;\boldsymbol{L}_i) = \log \gamma$ *and* $I(Z;\boldsymbol{X}) \to \log \gamma$ *as* $d \to \infty$.

Proof. In these examples, $Y_i^{(w)} = \boldsymbol{L}_i$ holds when $w = 2^l - 1$ and $w = m/\gamma$ for Boolean and arithmetic maskings, respectively, because of Equations (16) and (17). This implies the equality condition of data processing inequality on Markov chain $Y_i^{(w)} \leftrightarrow S_i \leftrightarrow \boldsymbol{L}_i$ because $I(S_i; Y_i^{(w)} \mid \boldsymbol{L}_i) = H(Y_i^{(w)} \mid \boldsymbol{L}_i) - H(Y_i^{(w)} \mid S_i, \boldsymbol{L}_i) = 0$. With this w, we have $I(S_i; \boldsymbol{L}_i) = I(Y_i^{(w)}; \boldsymbol{L}_i) = \log \gamma$ if $Y_i^{(w)} = \boldsymbol{L}_i$. By instantiating Lemma 4 with $\mathcal{W}_{\mathrm{peak}} = \{2^l - 1\}$ or $\{m/\gamma\}$ under this leakage, the proof is completed. □

Corollary 1 and its proof show cases that masking is not asymptotically secure if $\epsilon = \log \gamma$, implying that our condition for asymptotic security is the strongest if no assumption about leakage. To clarify the meaning, we construct a matching key-recovery attack, assuming that w exists such that $I(Y_i^{(w)}; \boldsymbol{L}_i) = \log \gamma_w$ holds for all i. In this case, $p_{Y_i^{(w)} \boldsymbol{L}_i}$ is a one-hot distribution for all i. This means that the attacker can determine the subgroup to which Z belong from leakages of shares with probability 1. In other words, if attacker knows $\langle S_i \mid w \rangle$ and $S_i \bmod \gamma_w$ from leakage for all i, the attacker can compute $\langle Z \mid w \rangle = \langle S_0 + S_1 + \cdots + S_d \mid w \rangle$ and $Z \bmod \gamma_w = S_0 + S_1 + \cdots + S_d \bmod \gamma_w$ for Boolean and arithmetic maskings, respectively. Using $\langle Z \mid w \rangle$ or $Z \bmod \gamma_w$, this attacker can filter the secret key and reduce the key space by $1/\gamma_w$, no matter how large d is. Thus, if $I(Y_i^{(w)}; \boldsymbol{L}_i) = \log \gamma_w$ holds for all i, the implementation can achieve $(1/\gamma_w, 1, \log \gamma_w)$-SCA security (in the (SR, n, μ)-SCA notion) or $(1, \gamma_w, \log \gamma_w)$-SCA security for any large d, which is insecure and independent of d. Conversely, $\mathcal{W}_{\mathrm{peak}} = \emptyset$ is sufficient for asymptotic security. In consequence, using Lemma 4, we reveal a necessity and sufficient condition as Theorem 8.

Theorem 8 (Necessary and sufficient condition for asymptotic security of masking). *Both Boolean and arithmetic masking is asymptotically secure against SCA represented by Figure 2 if and only if $I(Y_i^{(w)}; \boldsymbol{L}_i) < \log \gamma_w$ holds for all w for infinite shares.*

Proof. According to Lemma 4, $I(Z; \boldsymbol{X})$ converges to zero for a subsequence of infinite shares satisfying $I(Y_i^{(w)}; \boldsymbol{L}_i) < \log \gamma_w$ for all w. For whole sequence, $I(Z; \boldsymbol{X})$ also converges to zero because the number of remaining shares satisfying $I(Y_i^{(w)}; \boldsymbol{L}_i) = \log \gamma_w$ for some w is finite, indicating the sufficiency.

For the necessity, we prove its inverse (*i.e.*, the contraposition of its converse). In this case, $I(Y_i^{(w)}; \boldsymbol{L}_i) < \log \gamma_w$ holds for all w only d' times ($d' < \infty$), whereas $I(Y_i^{(w)}; \boldsymbol{L}_i) = \log \gamma_w$ holds for some w for infinite shares. If we take subsequence in i satisfying $I(Y_i^{(w)}; \boldsymbol{L}_i) = \log \gamma_w$, Lemma 4 states that $I(Z; \boldsymbol{X}) \to \log\left(1 + \left|\mathcal{W}_{\mathrm{peak}}\right|\right)$ as $d \to \infty$ for this subsequence. Other finite d' shares contributes to the decay of $I(Z; \boldsymbol{X})$, which never becomes zero with finite d' (if $I(S_i; \boldsymbol{L}_i) > 0$ for all i), followed by $|\mathrm{SR}_d - 1/m| > 0$ even when $d \to \infty$. Therefore, under the assumption, masking is not asymptotically secure. □

Relationship between $I(S_i; \boldsymbol{L}_i)$ and $I(Y_i^{(w)}; \boldsymbol{L}_i)$. Even if $I(S_i; \boldsymbol{L}_i) \geq \log \gamma$, masking can be asymptotically secure, because $I(S_i; \boldsymbol{L}_i) \geq \log \gamma$ is not necessary

condition of $I(Y_i^{(w)}; \boldsymbol{L}_i) \geq \log \gamma$ for all w. In other words, $I(Y_i^{(w)}; \boldsymbol{L}_i) < \log \gamma_w$ can hold for all w even if $I(S_i; \boldsymbol{L}_i) \geq \log \gamma$ holds. For example, consider a probability distribution satisfying $p_{S_i|\boldsymbol{L}_i}(s \mid \boldsymbol{l}) = 1 - \nu m$ for specific s and $p_{S_i|\boldsymbol{L}_i}(s \mid \boldsymbol{l}) = \nu$ for some $\nu \in (0, 1/m)$. If ν is very small, $p_{S_i|\boldsymbol{L}_i}(s \mid \boldsymbol{l})$ becomes close to one-hot distribution and $I(S_i; \boldsymbol{L}_i) \to \log m$ holds as $\nu \to 0$. Yet, if $\nu > 0$, the leakage contains (even very tiny) noise, and attacker cannot estimate S_i with probability 1. No matter how small $\nu > 0$ is, $p_{Y_i^{(w)}|\boldsymbol{L}_i}$ is not a one-hot distribution and $I(Y_i^{(w)}; \boldsymbol{L}_i) < \log \gamma_w$ holds for any w. Thus, Corollary 2 states an important practical aspect—*non-zero* noise renders masking asymptotically secure.

Corollary 2 (A sufficient condition for asymptotic security of masking). *Both Boolean and arithmetic maskings are asymptotically secure if* $p_{S_i|\boldsymbol{L}_i}(s \mid \boldsymbol{l}) \neq 0$ *holds for any* i, s, *and* $\boldsymbol{l}$ *satisfying* $p_{\boldsymbol{L}_i}(\boldsymbol{l}) \neq 0$.

Proof. It is sufficient for $p_{Y_i^{(w)}|\boldsymbol{L}_i}$ not to be one-hot distribution for all w, which is a sufficient condition of Theorem 6. □

6 Numerical evaluation

6.1 Evaluation of leakage energy

In this section, we evaluate the log-energy of $p_{Z|\boldsymbol{X}}$ (*i.e.*, the right-hand side of Equation (5) in Lemma 1 as an upper bound on $I(Z; \boldsymbol{X})$) for Boolean and arithmetic maskings. We assumed an HW leakage without physical noise; namely, $\boldsymbol{L}_i = \mathrm{HW}(S_i)$. Although there was no physical noise, the HW function decreases the information of S_i like noise because it is not injective, which means $H(S_i \mid \boldsymbol{L}_i) > 0$. Thus, we observed the *noise amplification* (*i.e.*, leakage mitigation by increasing masking order) by this simulation.

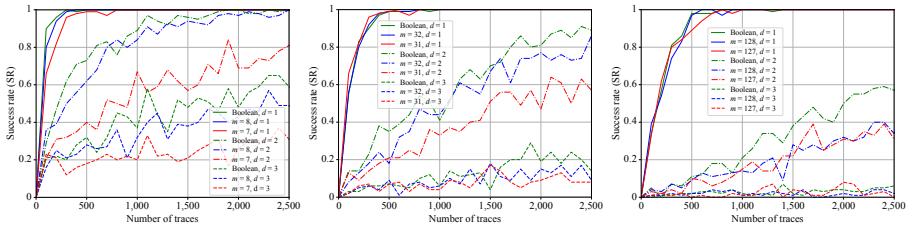
Table 2 reports the log-energy of $p_{Z|\boldsymbol{X}}$ for different bit length l and masking order d (where $d = 0$ means no masking), with field/ring characteristic of two (*i.e.*, Boolean), $2^l - 1$, and 2^l, respectively. See Section B for visualization of the power spectra of $p_{Z|\boldsymbol{X}}$. In Table 2, we found that arithmetic masking has significantly smaller leakage than Boolean masking. This is due to a subgroup leakage with $\mathcal{W}_{\mathrm{peak}} = \{2^l - 1\}$, as stated in Lemma 4. If the HW of shares leaks without physical noise, $\hat{p}_{Z|\boldsymbol{X}}(2^l - 1 \mid \boldsymbol{X}) = 1$ holds no matter how large d is (as confirmed in Section B), and $I(Z; \boldsymbol{X})$ converges to $\log 2 = 1$ as $d \to \infty$. Hence, Boolean masking is not secure against HW leakage without physical noise.

In contrast, even with no physical noise, the log-energy of arithmetic masking converges to zero as $d \to \infty$, because there is no explicit correspondence between subgroup of $\mathbb{Z}_m$ and the HW leakage. The power spectrum density never becomes one, meaning $\mathcal{W}_{\mathrm{peak}} = \emptyset$. Thus, we confirm the higher SCA security of arithmetic masking than Boolean masking. Comparing 2^l and odd characteristics, we found that the odd-characteristic masking has less leakage, especially when the field/ring size is small. However, if l is greater, their difference is small. This would be because the ratio of the number of divisors of 2^l decreases exponentially by increasing l. Therefore, if the bit length is greater, 2^l-characteristic masking would become closer to prime-field masking.

Table 2: Simulation result of log-energy of $p_{Z|\boldsymbol{X}}$

Order	$l=1$	$l=2$			$l=3$			$l=4$		
d	Any	Boolean	$m=3$	$m=4$	Boolean	$m=7$	$m=8$	Boolean	$m=15$	$m=16$
0	1	1.585	1.000	1.585	2.000	1.585	2.000	2.322	2.000	2.322
1	1	1.322	0.585	1.245	1.415	0.737	1.222	1.415	0.772	1.113
2	1	1.170	0.322	0.977	1.152	0.290	0.688	1.106	0.234	0.449
3	1	1.087	0.170	0.762	1.052	0.103	0.364	1.013	0.064	0.166

Order	$l=5$			$l=6$			$l=7$			$l=8$		
d	Boolean	$m=31$	$m=32$	Boolean	$m=63$	$m=64$	Boolean	$m=127$	$m=128$	Boolean	$m=255$	$m=256$
0	2.585	2.414	2.585	2.807	2.585	2.807	3.000	2.807	3.000	3.170	3.000	3.170
1	1.379	0.718	0.958	1.332	0.643	0.801	1.286	0.548	0.651	1.246	0.459	0.522
2	1.070	0.153	0.260	1.047	0.103	0.149	1.033	0.062	0.082	1.025	0.039	0.047
3	1.013	0.028	0.062	1.007	0.015	0.025	1.004	0.007	0.010	1.003	0.003	0.004



(a) $l=3$ (b) $l=5$ (c) $l=7$

Fig. 3: SR of template attack.

6.2 Evaluation of success rate

We demonstrate a simulation of template attack on masked implementations to evaluate its SR. In the simulation, the initial key addition of $Z = M + k$ was targeted, and we supposed that Z was masked and the attacker observes the leakages of shares. The leakage was given as $\boldsymbol{L}_i = \mathrm{HW}(S_i) + N$, where N is a zero-mean Gaussian noise with a variance $\sigma^2 = 1$. Note that all masking schemes satisfy the condition for asymptotic security in this condition, as leakages were given with non-zero noise. We performed higher-order template attack [24], which was proven optimal [42]. Figure 3 displays the relation between the number of traces and SR when $l = 3$, 5, and 7. For each number of trace, we evaluated the SR by repeating the attack on randomly generated datasets 100 times.

When $l = 3$, we found non-trivial differences among Boolean, 2^l-characteristic, and prime-field maskings. Arithmetic maskings are more secure than Boolean masking, and prime-field masking offers the highest level of security, as claimed in literature. In contrast, the difference gets smaller for larger l. In particular, when $l = 7$, we found little difference between 2^l-characteristic and prime. Recall that the difference of log-energy between arithmetic maskings are smaller when l is larger as observed in Table 2. These results would indicate that non-prime arithmetic masking can offer comparable security to prime-field masking when l is large. In consequence, we confirm the validity of our theory and claims.

7 Application to LR primitive design

7.1 Basic concept and construction

We discuss applications of our analysis to LR primitives. Some block cipher (BC) modes such as OCB [49] and XTS [35] employ *masked* BCs in their construction to build a tweakable BC (TBC) [50]. They are popular instances of BC-based TBC called XEX, proposed by Rogaway [63]. The structure of XEX is defined as $\mathsf{XEX}[E_K]((K,K'),M,T) \stackrel{\triangle}{=} E_K(M \oplus f_{K'}(T)) \oplus f_{K'}(T)$, where $T = (N,j)$ is a tweak consisting of nonce N and block index j, M is a plaintext, K and K' are the (possibly dependent) keys. For simplicity, we assume that they are independent. XEX is an essential building block for modes. By using XEX with tweak $T = (N,j)$ for encrypting j-th plaintext block with (non-repeating) nonce N, we expect the mask $f_{K'}(N,j)$ to inherently provide SCA protection without dedicated masking against SCAs, as claimed in [68]. However, since the mask addition in XEX is Boolean, it is less effective than arithmetic masking. To strengthen built-in SCA resilience of XEX, we propose a generalized variant of XEX, dubbed gXEX. Let $\mathbb{G}$ be an abelian group of size 2^b, where b is the block length, with an addition operator $\boxplus$ and a subtraction operator $\boxminus$. By interpreting a block as an element of $\mathbb{G}$, gXEX is defined as

$$\mathsf{gXEX}[E_K,\mathbb{G}]((K,K'),M,T) \stackrel{\triangle}{=} E_K(M \boxplus V) \boxminus V, \text{ where } V = f_{K'}(T).$$

We observe that $\mathsf{gXEX}[E_K,\mathbb{G}]$ is black-box secure if E is a secure block cipher and $f_{K'}$ is α-almost Delta universal with respect to $\mathbb{G}$ (α-AΔU($\mathbb{G}$), see below); the proof is a simple variation of the original proofs of LRW and XEX [50, 63].

Definition 9 (AΔU/AXU). *Let $H : \mathcal{K} \times \mathcal{T} \to \mathbb{G}$ be a function for an abelian group $\mathbb{G}$. We say H is α-AΔU($\mathbb{G}$) if $\max_{T,T' \in \mathcal{T}, T \neq T', \beta \in \mathbb{G}} \Pr[H_K(T) \boxplus H_K(T') = \beta] \leq \alpha$ holds for random K. An α-AΔU($\mathbb{F}_{2^b}$) hash function H, where $\boxplus$ is XOR, is also referred to α-almost XOR universal (AXU).*

XEX[E] is interpreted as $\mathsf{gXEX}[E_K,\mathbb{F}_{2^b}]$ with an α-AXU $f_{K'}$. To enhance SCA resilience, we realize $\mathsf{gXEX}[E_K,\mathbb{G}]$ for a certain non-Boolean $\mathbb{G}$ (*e.g.*, $\mathbb{Z}_{2^b}$ or $\mathbb{Z}^2_{2^{b/2}}$). For example, in [22], Chakraborty and Sarkar proposed several instantiations of $\mathsf{gXEX}[E_K,\mathbb{G}]$ for $\mathbb{G} = \mathbb{Z}_{2^b}$, with a dedicated A$\Delta$U($\mathbb{G}$) mask generation function $f_{K'}$. Still, since $\mathbb{Z}_{2^b}$ is not a field, the mask generation of [22] uses multiplication over a prime field slightly larger than 2^b. Such mask generation degrades efficiency when compared to those defined for XEX (*i.e.*, for $\mathbb{F}_{2^b}$).

When $\mathsf{gXEX}[E_K,\mathbb{G}]$ uses an AΔU($\mathbb{G}'$) mask generation function for some $\mathbb{G}' \neq \mathbb{G}$, we must conduct a dedicated security analysis. This can involve [46] for instance, where a mixture of addition over $\mathbb{F}_{2^b}$ and special 32-bit addition was studied for analyzing the security of GCM. This implies that, depending on the group for mask addition, we have to develop a dedicated mask generation function. Since a suitable mask addition group may change depending on platforms and environments, this situation is practically undesirable.

We show how to address this issue by reusing existing mask generation functions of XEX with a small overhead irrespective of the mask addition group. We first introduce α-ASU, a stronger security property than AXU and AΔU. Then, we prove that any AXU function can be elevated to ASU, and any ASU function is AΔU for any abelian group.

Definition 10 (ASU). *If* $\Pr[(H_K(T), H_K(T')) = (V, V')] \leq \alpha$ *holds for any distinct* $T, T' \in \mathcal{T}$ *and* $V, V' \in \mathbb{G}$, H *is called* α*-almost strongly universal (ASU).*

Lemma 5 (AXU to ASU). *For* $H : \mathcal{K} \times \mathcal{T} \to \mathbb{F}_{2^b}$, *let* $H^+ : \mathcal{K}' \times \mathcal{T} \to \mathbb{F}_{2^b}$ *for* $\mathcal{K}' = \mathcal{K} \times \mathbb{F}_{2^b}$ *be defined as* $H^+_{K'}(T) \stackrel{\triangle}{=} H_{K'_1}(T) \oplus K'_2$, *where* $K' = (K'_1, K'_2)$. *If* H *is* α*-AXU for some* $\alpha \in [0, 1]$, H^+ *is* $\alpha/2^b$*-ASU.*

Proof. Let $H : \mathcal{K} \times \mathcal{X} \to \{0,1\}^m$ be an α-AXU hash with m-bit output. Let $G : \mathcal{K}_G \times \mathcal{X} \to \{0,1\}^m$ for $\mathcal{K}_G = \mathcal{K} \times \{0,1\}^m$ be defined as $G((K_1, K_2), x) = H(K_1, x) \oplus K_2$. For any distinct $x, x' \in \mathcal{X}$, we have

$$\begin{aligned}
&\Pr[G((K_1, K_2), x) = y, G((K_1, K_2), x') = y'] \\
&= \Pr[G((K_1, K_2), x) \oplus G((K_1, K_2), x') = y \oplus y' \mid G((K_1, K_2), x') = y] \\
&\quad \cdot \Pr[G((K_1, K_2), x') = y']
\end{aligned}$$

The first probability of the right-hand side is

$$\begin{aligned}
&\Pr[H(K_1, x) \oplus H(K_1, x') = y \oplus y' \mid G((K_1, K_2), x') = y] \\
&\qquad = \Pr[H(K_1, x) \oplus H(K_1, x') = y \oplus y'] \leq \alpha
\end{aligned}$$

since K_2 cancels in the main clause and the conditional clause can be discarded because of the independence of K_2 from K_1. The second term is $1/2^m$ for the uniformity of K_2. Hence, $\Pr[G(K_1, K_2), x) = y, G((K_1, K_2), x') = y'] \leq \alpha/2^m$, showing G is $\alpha/2^m$-ASU. □

Lemma 6 (ASU implies AΔU). *If* $H : \mathcal{K} \times \mathcal{T} \to \mathbb{G}$ *is an* α*-ASU,* H *is an* $\alpha|\mathbb{G}'|$*-A*Δ*U*$(\mathbb{G}')$ *for any group* $\mathbb{G}'$ *of the same size as* $\mathbb{G}$, *assuming an implicit one-to-one mapping from* $\mathbb{G}$ *to* $\mathbb{G}'$.

Proof. Let $F : \mathcal{K} \times \mathcal{X} \to \mathbb{G}$ be α-ASU for an Abelian group $\mathbb{G}$. For any distinct $x, x' \in \mathcal{X}$, we have

$$\begin{aligned}
\Pr[F(K, x) - F(K, x') = d] &= \sum_{y, y'; y - y' = d} \Pr[F(K, x) = y, F(K, x') = y'] \\
&= \sum_{y \in \mathbb{G}, y' = y - d} \Pr[F(K, x) = y, F(K, x') = y'] \leq \alpha|\mathbb{G}|
\end{aligned}$$

since the additive inverse is unique on $\mathbb{G}$. Thus, F is $\alpha|\mathbb{G}|$-AΔU. □

Lemmas 5 and 6 suggest that, by XORing another independent key material to a mask generation function of known XEX instances such as [41, 49, 50, 55, 63],

we can securely choose the mask addition group, and the choice could be independent of the algorithm of the mask generation function, for its ASU property.

The above method enables a TBC with built-in first-order (arithmetic) masking. To achieve general d-th order (arithmetic) masking, we repeat the first-order mask generation procedure d times with independent keys, and add these masks by $\boxplus$ operator in a sequential manner. We name such a TBC with built-in d-th order arithmetic masking as $\mathsf{gXEX\text{-}A}^d$ (A for arithmetic); the construction could be optimized depending on the mask addition group $\mathbb{G}$. Thanks to Lemmas 5 and 6, the mask generation function of $\mathsf{gXEX\text{-}A}^d$ is AΔU. Using this fact with existing proofs [22, 50], with a minor extension to an AΔU mask generation instead of an AXU one, $\mathsf{gXEX\text{-}A}^d$ is proven to be a black-box secure TBC if the underlying block cipher is a strong pseudorandom permutation.

For convenience, we write $\mathsf{gXEX\text{-}X}^d$ to mean the counterpart of $\mathsf{gXEX\text{-}A}^d$ whose mask addition is XOR, corresponding to d-th order Boolean masking. Importantly, whereas standard d-th order masked implementations require $O(d^2)$ implementation cost [32], the overhead of $\mathsf{gXEX\text{-}A}^d$ is $O(d)$, breaking the barrier of $O(d^2)$ and enabling a highly efficient d-th order SCA-secure implementation.

7.2 Formulation of d-th order SCA security of $\mathsf{gXEX\text{-}A}^d$

One important question is whether our theory is applicable to $\mathsf{gXEX\text{-}A}^d$ because its encoding is different from standard masking. In [68], Ueno *et al.* showed that the encoding of XEX is generalization of Boolean masking with regard to the *XCCA attack* and XEX has inherent first-order SCA security, as the computation of intermediate value in BC requires a convolution of probability distributions (that is, an upper bound on $I(Z; \boldsymbol{X})$ holds similarly to Lemma 1). However, their formulation relies on the commutativity of key and mask additions in XEX, which does not hold for key XOR and arithmetic mask addition in $\mathsf{gXEX\text{-}A}^d$.

We show that our claims on masking hold for AES with $\mathsf{gXEX\text{-}A}^d$. Let S_i denote an l-bit partial mask corresponding to the i-th share and let $\boldsymbol{L}_i$ denote its leakage for $1 \le i \le d$. Addition and subtraction of $\mathsf{gXEX\text{-}A}^d$ are supposedly modulo 2^8, while AES key addition is XOR. Consider an SCA on the AES final round[11], where the encoded intermediate value is given as $Z = \phi((C - S_1 - \cdots - S_d) \oplus k)$ with a bijective selection function ϕ and an l-bit partial ciphertext C. Let $\boldsymbol{L}_0$ denote the leakage of Z. The probability distribution of $p_{Z|\boldsymbol{X}}$ with $\mathsf{gXEX\text{-}A}^d$ is expressed using convolutions as

$$\begin{aligned} p_{Z|\boldsymbol{X}}(z^{(k)} \mid \boldsymbol{X}) &= \sum_{z=\phi((c-s_1-\cdots-s_d)\oplus k)} p_{Z|\boldsymbol{L}_0}(z \mid \boldsymbol{L}_0) \prod_{i=1}^{d} p_{S_i|\boldsymbol{L}_i}(s_i \mid \boldsymbol{L}_i) \\ &= \sum_{c=\zeta+s_1+\cdots+s_d} p_{Z|\boldsymbol{L}_0}(\phi(\zeta \oplus k) \mid \boldsymbol{L}_0) \prod_{i=1}^{d} p_{S_i|\boldsymbol{L}_i}(s_i \mid \boldsymbol{L}_1) \\ &= \left(p^{(k)}_{Z|\boldsymbol{L}_0}(\phi(\cdot) \mid \boldsymbol{L}_0) * p_{S_1|\boldsymbol{L}_1}(\cdot \mid \boldsymbol{L}_1) * \cdots * p_{S_d|\boldsymbol{L}_d}(\cdot \mid \boldsymbol{L}_d)\right)(c), \end{aligned}$$

[11] Our claim holds for SCA targeting the AES first round by inverting the signs of S_i.

where $\zeta = \phi^{-1}(z) \oplus k$ denotes the BC output; function $p^{(k)}_{Z|\boldsymbol{L}_0}(\phi(\cdot) \mid \boldsymbol{L}_0) : \mathbb{Z}_m \to \mathbb{R}$ is a composition function of XOR with k followed by $p_{Z|\boldsymbol{L}_0}(\phi(\cdot) \mid \boldsymbol{L}_0)$; and $*$ is a convolution. Since $p_{Z|\boldsymbol{X}}$ is expressed as d-convolution of conditional probability distributions given leakages, it should have a d-th order SCA security, as stated and evaluated by XCCA attack in [68].

A difference between gXEX-A^d and XCCA attack in [68] is that a probability distribution is key-dependent due to the non-commutativity. However, in SCA on AES, major selection functions satisfy the *key-independence condition* [44], which implies that the probability distributions of leakage with different key values are that of permutation of each other. Hence, its quantity (*e.g.*, energy and MI) is independent of key values. This indicates that Lemma 1 is applicable to gXEX-A^d and has an intrinsic d-th order SCA security corresponding to mask addition. See [44, Section 3.1] for how to examine the key-independence condition.

Remark 4 (Order of mask additions). In gXEX-A^d (and gXEX-X^d), the d masks should be added *serially* to the input and output to ensure that the encoded intermediate value should be computed using d convolutions of probability distributions. If we compute sum of d masks in advance, the sum would leak and be exploited to avoid convolutions, which indicates a security order degradation.

7.3 Leakage-resilient AE based on gXEX-A^d

gXEX-A^d can serve as a building block of a class of efficient LR authenticated encryption (LR-AE). We present one such instantiation named AOCB in Figure 4. The design largely follows the blueprint proposed by Mennink [54].[12] This is a rate-one (*i.e.*, needing one BC call to process one input block) and parallelizable AE based on OCB using gXEX-A^d denoted by $\widetilde{E}$. It involves d independently-keyed instances of a mask generation function $f_L(\cdot,\cdot,\cdot)$, which enables efficient incremental computation, namely $f_L(N, i+1, j)$ can be easily and serially computed from $f_L(N, i, j)$, and is AΔU. As described above, we can instantiate f by using known incremental AXU functions [49, 63] with additional key whitening (XOR). To generate d instances, we assume KDF, which can be either instantiated by the internal BC or other functions. For simplicity, we assume that input messages are integral; other cases could be easily covered by a simple extension. A black-box security proof of AOCB follows those of OCB. In more detail, we can view gXEX-A^d as a variant of ΘCB [49], whose security is reduced to the underlying TBC $\widetilde{E}$. Assuming KDF is a PRF, each mask of gXEX-A^d is independent, and gXEX-A^d is a provably secure TBC built on BC.

Leakage resilience claim as leveled implementation. We leverage the core idea of Mennink towards LR-AE [54]; that is, building a TBC-based AE where every internal call of the TBC (either dedicated or built on a BC) takes a unique tweak during encryption. From this, AOCB claims leakage resilience as *leveled implementation*. Namely, AOCB achieves d-th order SCA security as AE

[12] We adopt a minor refinement suggested by [67] so that every associated data (AD) block is processed by the TBC taking a nonce. This is needed for leakage resilience.

Algorithm $\mathsf{AOCB}.\mathcal{E}_{\widetilde{E}_K}(N, A, M)$

1. $(M[1], \ldots, M[m]) \xleftarrow{n} M$
2. **for** $i \leftarrow 1$ **to** m
3. $\quad C[i] \leftarrow \widetilde{E}_K(M[i], (N, i, 0))$
4. $\Sigma \leftarrow M[1] \oplus \cdots \oplus M[m]$
5. $T \leftarrow \widetilde{E}_K(\Sigma, (N, m, 1))$
6. $T \leftarrow T \oplus \mathsf{PHASH}_{\widetilde{E}_K}(A)$
7. **return** (C, T)

Algorithm $\widetilde{E}_K(M, (N, i, j))$

1. $X \leftarrow M$
2. **for** $h = 1$ **to** d
3. $\quad L_h \leftarrow \mathsf{KDF}_K(h)$ //can be cached
4. $\quad S_h \leftarrow f_{L_h}(N, i, j)$
5. $\quad X \leftarrow X \boxplus S_h$
6. $C \leftarrow E_K(X)$
7. **for** $h = d$ **to** 1
8. $\quad C \leftarrow C \boxminus S_h$
9. **return** C

Algorithm $\mathsf{PHASH}_{\widetilde{E}_K}(A)$

1. $S \leftarrow 0^n$
2. $(A[1], \ldots, A[m]) \xleftarrow{n} A$
3. **for** $i \leftarrow 1$ **to** m
4. $\quad S \leftarrow S \oplus \widetilde{E}_K(A[i], (N, i, 2))$
5. **return** S

Algorithm $\mathsf{AOCB}.\mathcal{D}_{\widetilde{E}_K}(N, A, C, T)$

1. $(C[1], \ldots, C[m]) \xleftarrow{n} C$
2. **for** $i \leftarrow 1$ **to** m
3. $\quad M[i] \leftarrow \widetilde{E}_K^{-1}(C[i], (N, i, 0))$
4. $\Sigma \leftarrow M[1] \oplus \cdots \oplus M[m-1]$
5. $T^* \leftarrow \widetilde{E}_K(\Sigma, (N, m, 1))$
6. $T^* \leftarrow T^* \oplus \mathsf{PHASH}_{\widetilde{E}_K}(A)$
7. **if** $T = T^*$ **return** M
8. **else return** $\perp$

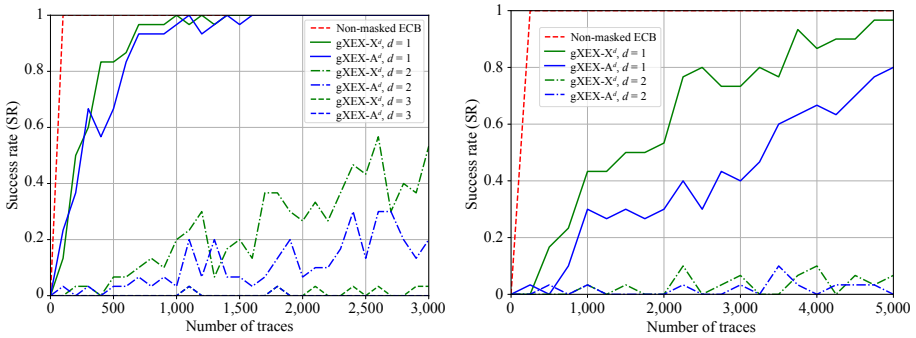
Algorithm $\widetilde{E}_K^{-1}(C, (N, i, j))$

1. $X \leftarrow C$
2. **for** $h = 1$ **to** d
3. $\quad L_h \leftarrow \mathsf{KDF}_K(h)$ //can be cached
4. $\quad S_h \leftarrow f_{L_h}(N, i, j)$
5. $\quad X \leftarrow X \boxplus S_h$
6. $M \leftarrow E_K^{-1}(X)$
7. **for** $h = d$ **to** 1
8. $\quad M \leftarrow M \boxminus S_h$
9. **return** M

Fig. 4: Algorithms of AOCB for integral input blocks. (N, A, M, C, T) denotes (Nonce, Associated data, Plaintext, Ciphertext, Tag). $X[1], \ldots, X[m] \xleftarrow{n} X$ denotes n-bit block parsing of bit string X. $C \leftarrow \widetilde{E}_K(M, T)$ denotes an encryption of n-bit plaintext block M with a tweak T which results in the ciphertext C. $M \leftarrow \widetilde{E}_K^{-1}(C, T)$ denotes the corresponding decryption.

by effectively combining a few DPA-resistant (*e.g.*, d-th order masked) modules for minor parts—only initialization and finalization—and non-masked $\mathsf{gXEX\text{-}A}^d$ modules for most (T)BC calls to process encryption and AD. Precisely, AOCB has two security orders. First, AOCB is d-th order secure against SCA on $\mathsf{gXEX\text{-}A}^d$ using the number of traces equal to $\mathsf{gXEX\text{-}A}^d$ calls. Second, let d' be the security order of initialization and finalization implementation. AOCB is d'-th order secure against SCA on them using the number of traces equal to AOCB calls.

Assumptions and limitations. The security claim of this type of LR-AE considers only encryption leakage in the nonce-respecting setting. If attacker can query repeated nonces, the attacker can perform first-order SCA on $\mathsf{gXEX\text{-}A}^d$ for encryption and AD processing parts. Nonce-disrespecting attacks and attacks on initialization and finalization are called *trivial* in [14]. As trivial attacks are not covered by the proof, their feasibility does not contradict the claim of LR-AE. In other words, AOCB implementation should protect initialization and finalization by SCA countermeasure and be nonce-respecting, as in existing LR-AEs based



(a) $\sigma^2 = 1$ (b) $\sigma^2 = 5$

Fig. 5: SR of SCA on AES with ECB, gXEX-X^d, and gXEX-A^d.

on leveled implementation. Another limitation of AOCB security would be that *blind SCAs*, which use neither BC input nor output, can be mounted with lower orders [61]. However, blind SCAs have difficulty in achieving high SRs even when compared to higher-order SCAs[13]. Thus, AOCB requires noise such that blind SCAs get difficult, as blind SCAs are highly noise sensitive than non-blind SCAs.

Comparison with existing (T)BC-based LR-AE schemes. Existing (T)BC-based LR-AEs of leveled implementation (*e.g.*, TEDT [16] and Triplex [65]) claim that all (T)BC modules except for initialization and finalization only requires SPA-resistance (*i.e.*, security against one-trace SCA). Yet, none achieved rate-one (i.e. needing one BC call for one input block). In contrast, the LR security claim of AOCB—inherent d-th order SCA security of gXEX-A^d—is weaker than them. AOCB requires BC implementation to be resistant against n-trace d-th order SCA, where n should be pre-determined as specification. However, it would be practically secure with small d as evaluated. In addition, AOCB is rate-one and parallelizable by following the OCB structure, hence, in terms of the number of BC calls, it achieves essentially the highest efficiency as an AE. Furthermore, AOCB uses only standard instruction sets on modern CPU/microcontrollers and can be implemented with little overhead. Thus, AOCB would offer a high practicality even when compared to existing counterparts.

7.4 Success rate evaluation

Figure 5 reports the SR of template attack on gXEX-A^d and gXEX-X^d using simulated leakages $\boldsymbol{L}_0, \boldsymbol{L}_1, \ldots, \boldsymbol{L}_d$ with different noise variance σ^2 as same as Section 6.2, whereas the number of trials was 30 due to its heavy computation.

[13] Even the state-of-the-art blind SCA did not achieve guessing entropy (GE) less than 1, which indicates the difficulty in achieving high SRs, as SR is upper- and lower-bounded by GE [44, Lemma 2]. Full-key recovery of AES (*i.e.*, simultaneous successes of 16 partial keys) requires significantly high SR even with key enumeration [69].

We used AES as the underlying BC, meaning that $l = 8$, and used an identity function as selection function ϕ (*i.e.*, we targeted the output of final-round Sbox). When no countermeasure was adopted (*i.e.*, non-masked ECB), the secret key was recovered within 100 and 250 traces under $\sigma^2 = 1$ and 5, respectively. However, gXEX-X^d and gXEX-A^d resisted the SCA with significantly more traces. Further, gXEX-A^d achieved lower SR and higher security than gXEX-X^d thanks to the stronger security of arithmetic masking. In addition, when the noise variance was large, arithmetic masking was more secure than Boolean with smaller d, indicating that arithmetic masking amplified noise more effectively. These findings empirically validated our claim on security by primitive design.

References

1. Aikata, A., Basso, A., Cassiers, G., Mert, A.C., Sinha Roy, S.: Kavach: Lightweight masking techniques for polynomial arithmetic in lattice-based cryptography. IACR Transactions on Cryptographic Hardware and Embedded Systems **2023**(3), 366–390 (2023). https://doi.org/10.46586/tches.v2023.i3.366-390, https://tches.iacr.org/index.php/TCHES/article/view/10967
2. Azouaoui, M., Bronchain, O., Cassiers, G., Hoffmann, C., Kuzovkova, Y., Renes, J., Schneider, T., Schönauer, M., Standaert, F.X., van Vredendaal, C.: Protecting Dilithium against leakage: Revisited sensitivity analysis and improved implementations. IACR Transactions on Cryptographic Hardware and Embedded Systems (4), 58–79 (2023). https://doi.org/10.46586/tches.v2023.i4.58-79, https://tches.iacr.org/index.php/TCHES/article/view/11158
3. Bache, F., Paglialong, C., Oder, T., Schneider, T., Güneysu, T.: High-speed masking for polynomial comparison in lattice-based KEMs. IACR Transactions on Cryptographic Hardware and Embedded Systems **2020**(3), 483–507 (2020)
4. Balon, B., Grassi, L., Méaux, P., Moos, T., Standaert, F.X., Steiner, M.J.: midpSquare: Leveraging the strong side-channel security of prime-field masking in software. IACR Transactions on Cryptographic Hardware and Embedded Systems **2025**(4), 486–519 (2025). https://doi.org/10.46586/tches.v2025.i4.486-519, https://tches.iacr.org/index.php/TCHES/article/view/12419
5. Barthe, G., Balaïd, S., Dupressoir, F., Fouque, P.A., Grégoire, B., Strub, P.Y., Zucchini, R.: Strong non-interference and type-directed higher-order masking. In: ACM SIGSAC Conference on Computer and Communications Security (CCS 2016). pp. 116–129 (2016)
6. Béguinot, J., Masure, L.: On the average random probing model. IACR Transactions on Cryptographic Hardware and Embedded Systems **2025**(3), 32–55 (2025). https://doi.org/10.46586/tches.v2025.i3.32-55
7. Béguinot, J., Cheng, W., Guilley, S., Liu, Y., Masure, L., Rioul, O., Standaert, F.X.: Removing the field size loss from Duc et al.'s conjectured bound for masked encodings. In: International Workshop on Constructive Side-Channel Analysis and Secure Design (COSADE). pp. 86–104 (2023). https://doi.org/10.1007/978-3-031-29497-6_5, https://doi.org/10.1007/978-3-031-29497-6_5
8. Béguinot, J., Cheng, W., Guilley, S., Rioul, O.: Formal security proofs via Doeblin coefficients: Optimal side-channel factorization from noisy leakage to random probing. In: Advances in Cryptology—CRYPTO 2024: 44th Annual International Cryptology Conference, Part VI. pp. 389–426 (2024). https://doi.org/10.1007/978-3-031-68391-6_12

9. Belaïd, S., Coron, J.S., Prouff, E., Rivain, M., Taleb, A.R.: Random probing security: Verification, composition, expansion and new constructions. In: Advances in Cryptology—CRYPTO 2020. pp. 339–368 (2020)
10. Belaïd, S., Rivain, M., Rossi, M.: On the power of expansion: More efficient constructions in the random probing model. In: Advances in Cryptology—EUROCRYPT 2021. pp. 313–343. Springer (2021)
11. Belaïd, S., Rivain, M., Rossi, M.: New techniques for random probing security and application to Raccoon signature scheme. In: Advances in Cryptology—EUROCRYPT 2025. pp. 94–123. Springer (2025). https://doi.org/10.1007/978-3-031-91101-9_4
12. Belaïd, S., Rivain, M., Taleb, A.R., Vergnaud, D.: Dynamic random probing expansion with quasi linear asymptotic complexity. In: Advances in Cryplology—ASIACRYPT 2021. pp. 157–188 (2021)
13. Bernstein, D.J.: ChaCha20, a variant of Salsa20 (2025), https://cr.yp.to/chacha/chacha-20080128.pdf
14. Berti, F., Bhasin, S., Breier, J., Hou, X., Poussier, R., Standaert, F.X., Udvarhelyi, B.: A finer-grain analysis of the leakage (non) resilience of OCB. IACR Transactions on Cryptographic Hardware and Embedded Systems **2022**, 461–481 (2021). https://doi.org/10.46586/tches.v2022.i1.461-481, https://tches.iacr.org/index.php/TCHES/article/view/9304
15. Berti, F., Faust, S., Orlt, M.: Provable secure parallel gadgets. IACR Transactions on Cryptographic Hardware and Embedded Systems **2023**(4), 420–459 (2023)
16. Berti, F., Guo, C., Pereira, O., Peters, T., Standaert, F.X.: TEDT, a leakage-resist AEAD mode for high physical security applications. IACR Transactions on Cryptographic Hardware and Embedded Systems **2020**(1), 256–320 (2019). https://doi.org/10.13154/tches.v2020.i1.256-320
17. Beyne, T., Verbauwhede, T.: Integral cryptanalysis in characteristic p. Asiacrypt 2025 (2025), https://eprint.iacr.org/2025/932
18. Bos, J.W., Gourjon, M., Renes, J., Schneider, T., van Vredendaal, C.: Masking Kyber: First- and higher-order implementations. IACR Transactions on Cryptographic Hardware and Embedded Systems **2021**(4), 173–214 (2021). https://doi.org/10.46586/tches.v2021.i4.173-214, https://tches.iacr.org/index.php/TCHES/article/view/9064
19. Brian, G., Dziembowski, S., Faust, S.: From random probing to noisy leakages without field-size dependence. In: Advances in Cryptology—EUROCRYPT 2024. pp. 345–374. Springer (2024)
20. Cassiers, G., Masure, L., Momin, C., Moos, T., Standaert, F.X.: Prime-field masking in hardware and its soundness against low-noise SCA attacks. IACR Transactions on Cryptographic Hardware and Embedded Systems **2023**(2), 482–518 (2023). https://doi.org/10.46586/tches.v2023.i2.482-518, https://tches.iacr.org/index.php/TCHES/article/view/10291
21. Cassiers, G., Standaert, F.X.: Trivially and efficiently composing masked gadgets with probe isolating non-interference. IEEE Transactions on Information Forensics and Security **15**, 2542–2555 (2020). https://doi.org/10.1109/TIFS.2020.2971153
22. Chakraborty, D., Sarkar, P.: A general construction of tweakable block ciphers and different modes of operations. IEEE Trans. Inf. Theory **54**(5), 1991–2006 (2008)
23. Chari, S., Jutla, C.S., Rao, J.R., Rohatgi, P.: Towards sound approaches to counteract power-analysis attacks. In: Advances in Cryptology—CRYPTO' 99. pp. 398–412 (1999)

24. Chari, S., Rao, J.R., Rohatgi, P.: Template attacks. In: International Workshop on Cryptographic Hardware and Embedded Systems. pp. 13–28. LNCS (2002)
25. de Chérisey, E., Guilley, S., Rioul, O., Piantanida, P.: Best information is most successful: Mutual information and success rate in side-channel analysis. IACR Transactions on Cryptographic Hardware and Embedded Systems **2019**(2), 49–79 (2019)
26. de Cnudde, T., Bilgin, B., Gierlichs, B., Nikov, V., Nikova, S., Rijmen, V.: Does coupling affect the security of masked implementations? In: International Workshop on Constructive Side-Channel Analysis and Secure Design (COSADE 2017). Lecture Notes in Computer Science, vol. 10348, pp. 1–18 (2017)
27. Coron, J.S., Gérard, F., Montoya, S., Zeitoun, R.: High-order polynomial comparison and masking lattice-based encryption. IACR Transactions on Cryptographic Hardware and Embedded Systems **2023**(1), 153–192 (2022). https://doi.org/10.46586/tches.v2023.i1.153-192, https://tches.iacr.org/index.php/TCHES/article/view/9950
28. Coron, J.S., Gérard, F., Montoya, S., Zeitoun, R.: High-order table-based conversion algorithms and masking lattice-based encryption. IACR Transactions on Cryptographic Hardware and Embedded Systems **2022**, 1–40 (2022). https://doi.org/10.46586/tches.v2022.i2.1-40, https://tches.iacr.org/index.php/TCHES/article/view/9479
29. Coron, J.S., Gérard, F., Trannoy, M., Zeitoun, R.: High-order masking of NTRU. IACR Transactions on Cryptographic Hardware and Embedded Systems **2023**(2), 180–211 (2023). https://doi.org/10.46586/tches.v2023.i2.180-211, https://tches.iacr.org/index.php/TCHES/article/view/10281
30. Coron, J.S., Gérard, F., Trannoy, M., Zeitoun, R.: Improved gadgets for the high-order masking of dilithium. IACR Transactions on Cryptographic Hardware and Embedded Systems **2023**(4), 110–145 (2023). https://doi.org/10.46586/tches.v2023.i4.110-145, https://tches.iacr.org/index.php/TCHES/article/view/11160
31. Cover, T.M., Thomas, J.A.: Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing). Wiley-Interscience, USA (2006)
32. Daemen, J.: Changing of the guards: A simple and efficient method for achieving uniformity in threshold sharing. In: International conference on Cryptographic Hardware and Embedded Systems. pp. 137–153. Springer International Publishing (2017)
33. Duc, A., Dziembowski, S., Faust, S.: Unifying leakage models: From probing attacks to noisy leakage. In: Advances in Cryptology—EUROCRYPT 2014. pp. 423–440 (2014)
34. Duc, A., Faust, S., Standaert, F.X.: Making masking security proofs concrete: Or how to evaluate the security of any leakage device. In: Advances in Cryptology—EUROCRYPT 2015. Lecture Notes in Computer Science, vol. 9056, pp. 401–429. Springer (2015)
35. Dworkin, M.: Recommendation for Block Cipher Modes of Operation: The XTS-AES Mode for Confidentiality on Storage Devices. Standard, National Institute of Standards and Technology. (2010)
36. Dziembowski, S., Faust, S., Skórski, M.: Optimal amplification of noisy leakages. In: Theory of Cryptography. pp. 291–318 (2016)
37. D'Anvers, J.P., Heinz, D., Pessl, P., Van Beirendonck, M., Verbauwhede, I.: Higher-order masked ciphertext comparison for lattice-based cryptography. IACR Transactions on Cryptographic Hardware and Embedded Systems **2022**(2), 115–139

(2022). https://doi.org/10.46586/tches.v2022.i2.115-139, https://tches.iacr.org/index.php/TCHES/article/view/9483
38. Faust, S., Grosso, V., Merino Del Pozo, S., Paglialonga, C., Standaert, F.X.: Composable masking schemes in the presence of physical defaults & the robust probing model. IACR Transactions on Cryptographic Hardware and Embedded Systems **2018**(3), 89–120 (2018)
39. Fedotov, A.A., Harremoës, P., Topsoe, F.: Refinements of Pinsker's inequality. IEEE Transactions on Information Theory **49**(6), 1491–1498 (2003). https://doi.org/10.1109/TIT.2003.811927
40. Gao, S., Marshall, B., Page, D., Oswald, E.: Share-slicing: Friend or foe? IACR Transactions on Cryptographic Hardware and Embedded Systems **2020**(1), 152–174 (2020)
41. Granger, R., Jovanovic, P., Mennink, B., Neves, S.: Improved masking for tweakable blockciphers with applications to authenticated encryption. In: EUROCRYPT (1). Lecture Notes in Computer Science, vol. 9665, pp. 263–293. Springer (2016)
42. Heuser, A., Rioul, O., Guilley, S.: Good is not good enough: Deriving optimal distinguishers from communication theory. In: International Workshop on Cryptographic Hardware and Embedded Systems. pp. 55–74 (2014)
43. Ishai, Y., Sahai, A., Wagner, D.: Private circuits: Securing hardware against probing attacks. In: Advances in Cryptology—CRYPTO 2003. pp. 463–481 (2003)
44. Ito, A., Ueno, R., Homma, N.: Toward optimal deep-learning based side-channel attacks: Probability concentration inequality loss and its usage. Cryptology ePrint Archive, Report 2021/1216 (2021), https://ia.cr/2021/1216
45. Ito, A., Ueno, R., Homma, N.: On the success rate of side-channel attacks on masked implementations: Information-theoretical bounds and their practical usage. In: ACM SIGSAC Conference on Computer and Communications Security (CCS 2022). pp. 1521–1535 (2022)
46. Iwata, T., Ohashi, K., Minematsu, K.: Breaking and repairing GCM security proofs. In: CRYPTO. Lecture Notes in Computer Science, vol. 7417, pp. 31–49. Springer (2012)
47. Jahandideh, V., Mennink, B., Batina, L.: A decomposition approach for evaluating security of masking. In: Advances in Cryptology—ASIACRYPT 2025. pp. 580–612 (2025)
48. Kocher, P., Jaffe, J., Jun, B.: Differential power analysis. In: Advances in Cryptology—CRYPTO' 99. Lecture Notes in Computer Science, vol. 1666, pp. 388–397 (1999)
49. Krovetz, T., Rogaway, P.: The software performance of authenticated-encryption modes. In: FSE. Lecture Notes in Computer Science, vol. 6733, pp. 306–327. Springer (2011)
50. Liskov, M.D., Rivest, R.L., Wagner, D.A.: Tweakable block ciphers. In: CRYPTO. Lecture Notes in Computer Science, vol. 2442, pp. 31–46. Springer (2002)
51. Mangard, S., Pramstaller, N., Oswald, E.: Successfully attacking masked AES hardware implementations. In: International Workshop on Cryptographic Hardware and Embedded Systems. pp. 157–171 (2005)
52. Masure, L., Méaux, P., Moos, T., Standaert, F.X.: Effective and efficient masking with low noise using small-Mersenne-prime ciphers. In: Advances in Cryptology—EUROCRYPT 2023. pp. 596–627 (2023)
53. Masure, L., Rioul, O., Standaert, F.X.: A nearly tight proof of Duc et al.'s conjectured security bound for masked implementations. In: Smart Card Research and Advanced Applications: 21st International Conference, CARDIS 2022. pp. 69–81 (2022). https://doi.org/10.1007/978-3-031-25319-5_4

54. Mennink, B.: Beyond birthday bound secure fresh rekeying: Application to authenticated encryption. In: ASIACRYPT (1). Lecture Notes in Computer Science, vol. 12491, pp. 630–661. Springer (2020)
55. Minematsu, K.: A short universal hash function from bit rotation, and applications to blockcipher modes. In: ProvSec. Lecture Notes in Computer Science, vol. 8209, pp. 221–238. Springer (2013)
56. Oder, T., Schneider, T., Pöppelmann, T., Güneysu, T.: Practical CCA2–secure and masked ring-LWE implementation. IACR Transactions on Cryptographic Hardware and Embedded Systems **2018**(1), 142–174 (2018)
57. del Pino, R., Prest, T., Rossi, M., Saarinen, M.J.O.: High-order masking of lattice signatures in quasilinear time. In: 2023 IEEE Symposium on Security and Privacy (SP). pp. 1168–1185 (2023). https://doi.org/10.1109/SP46215.2023.10179342
58. Pinsker, M.S., Feinstein, A.: Information and information stability of random variables and processes, vol. 12. Holden-Day San Francisco (1964)
59. Prouff, E., Rivain, M.: Masking against side-channel attacks: A formal security proof. In: Advances in Cryptology—EUROCRYPT 2013. pp. 142–159 (2013)
60. Reparaz, O., Bilgin, B., Nikova, S., Gierlichs, B., Verbauwhede, I.: Consolidating masking schemes. In: Advances in Cryptology—CRYPTO 2015. pp. 764–783 (2015)
61. Rezaeezade, A., Yap, T., Jap, D., Bhasin, S., Picek, S.: Breaking the blindfold: deep learning-based blind side-channel analysis. In: 34th USENIX Security Symposium (2025)
62. Rivain, M., Prouff, E.: Provably secure higher-order masking of AES. In: International Workshop on Cryptographic Hardware and Embedded Systems (CHES). pp. 413–427 (2010)
63. Rogaway, P.: Efficient instantiations of tweakable blockciphers and refinements to modes OCB and PMAC. In: ASIACRYPT. Lecture Notes in Computer Science, vol. 3329, pp. 16–31. Springer (2004)
64. Shamir, A.: How to share a secret. Communications of the ACM **22**(11), 612–613 (1979). https://doi.org/10.1145/359168.359176
65. Shen, Y., Peters, T., Standaert, F.X., Cassiers, G., Verhamme, C.: Triplex: an efficient and one-pass leakage-resistant mode of operation. IACR Transactions on Cryptographic Hardware and Embedded Systems **4**, 135–162 (2022)
66. Tibouchi, M.: Security Reduction, pp. 2333–2335. Springer Nature Switzerland (2025). https://doi.org/10.1007/978-3-030-71522-9_515
67. Ueno, R., Homma, N., Inoue, A., Minematsu, K.: Fallen sanctuary: A higher-order and leakage-resilient rekeying scheme. IACR Transactions on Cryptographic Hardware and Embedded Systems **1**(2024), 264–308 (2023)
68. Ueno, R., Ito, A., Todo, Y., Inoue, A., Minematsu, K., Ishikawa, H., Homma, N.: All you need is XOR-convolution: A generalized higher-order side-channel attack with application to XEX/XE-based encryptions. IACR Transactions on Cryptographic Hardware and Embedded Systems **2025**(3), 317–360 (2025). https://doi.org/10.46586/tches.v2025.i3.317-360, https://tches.iacr.org/index.php/TCHES/article/view/12219
69. Veyrat-Charvillon, N., Gérard, B., Renauld, M., Standaert, F.X.: An optimal key enumeration algorithm and its application to side-channel attacks. In: Selected Areas in Cryptography (2013)
70. Wang, B., Ji, F., Sun, Y., Wang, W.: Random probing security with precomputation. IACR Transactions on Cryptographic Hardware and Embedded Systems **2025**(1), 523–551 (2024)

A Discrete Fourier transform (DFT) and Walsh–Hadamard transform (WHT)

In this paper, we utilize the DFT and WHT to analyze masking schemes. They are defined as follows:

Definition 11 (Discrete Fourier transform (DFT)). *Let f be a function $\mathbb{Z}_m \to \mathbb{R}$. Its m-point DFT $\hat{f} : \mathbb{Z}_m \to \mathbb{C}$ is defined by*

$$\hat{f}(w) = \sum_{s=0}^{m-1} f(s) \mathrm{e}^{-\mathrm{i}s\Omega_w},$$

where $\Omega_w = 2\pi w/m$ is the normalized frequency.

Definition 12 (Walsh–Hadamard transform (WHT)). *Let f be a function $\mathbb{Z}_2^l \to \mathbb{R}$. Its WHT $\hat{f} : \mathbb{Z}_2^l \to \mathbb{R}$ is defined by*

$$\hat{f}(w) = \sum_{s=0}^{2^l-1} f(s)(-1)^{\langle s|w\rangle},$$

where $\langle \cdot \mid \cdot \rangle : \mathbb{Z}_2^l \times \mathbb{Z}_2^l \to \mathbb{Z}$ is an inner product.

In this paper, we exploit two important characteristics of them. First is that the cyclic convolution in the time representation corresponds to the Hadamard product in the frequency representation; that is, it holds

$$f(s) = \sum_{s=s_1+s_2} f_1(s_1) f_2(s_2) \Leftrightarrow \hat{f}(w) = \hat{f}_1(w)\hat{f}_2(w),$$

for any s and w, where the summation (*i.e.*, $s_1 + s_2$) is defined over the additive group of the input domain of DFT or WHT (*i.e.*, arithmetic addition modulo m for m-point DFT and XOR for WHT).

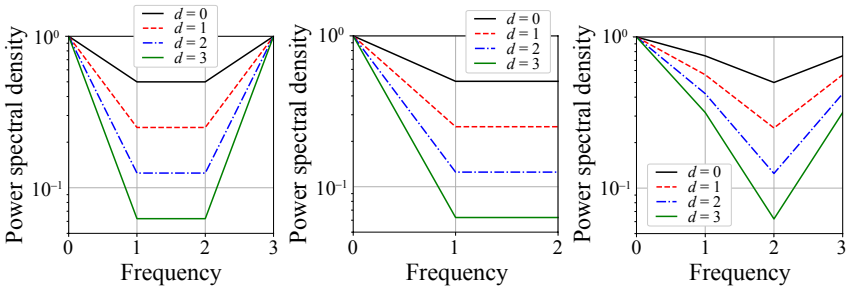
Second is the correspondence between the energy of a signal in the time and frequency representations, as stated by Parseval's theorem (or Plancherel's theorem).

Theorem 9 (Parseval (or Plancherel)). *Let f be a function and $\hat{f}$ be its DFT or WHT. It holds*

$$\sum_s \left(f(s)\right)^2 = \frac{1}{m} \sum_w \left|\hat{f}(w)\right|^2,$$

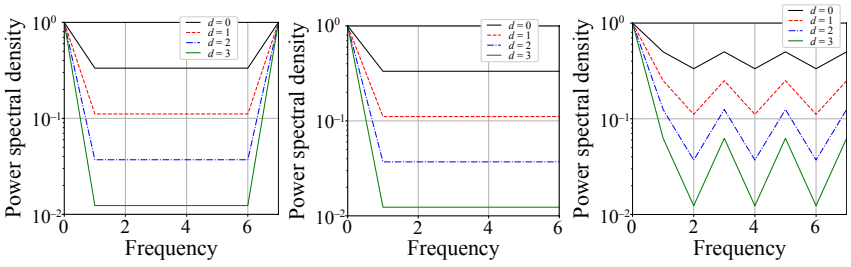
where m is the cardinality of the field/ring of the domain of f. Note that $|\hat{f}(w)|^2 = \hat{f}(w)\overline{\hat{f}(w)}$ for DFT ($\overline{c}$ denotes the complex conjugate of a complex number c), whereas $|\hat{f}(w)|^2 = (\hat{f}(w))^2$ for WHT.

Note that the normalization of DFT/WHT has no impact on our bounds if the result of convolutional product (*i.e.*, Hadamard product in frequency domain) and coefficient of Perseval's theorem are appropriately scaled and consistently normalized; that is, DC component at $w = 1$ of DFT/WHT of any probability distribution must equal to the normalization factor (in our definition, the normalization factor is one).



(a) Boolean (b) $m = 3$ (c) $m = 4$

Fig. 6: Power spectra of $p_{Z|\boldsymbol{X}}$ with $l = 2$.



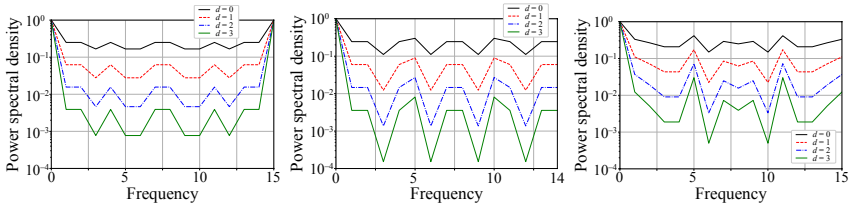
(a) Boolean (b) $m = 7$ (c) $m = 8$

Fig. 7: Power spectra of $p_{Z|\boldsymbol{X}}$ with $l = 3$.

B Power spectra of $p_{Z|X}$ under HW leakage

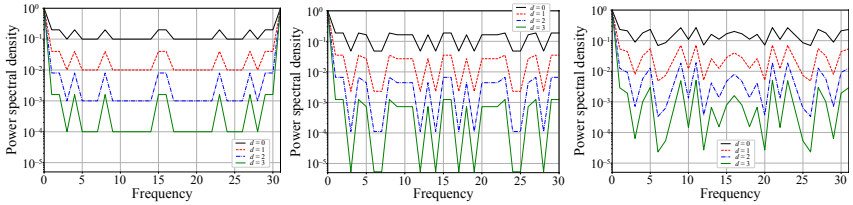
We simulated the power spectra of $p_{Z|\boldsymbol{X}}$ with different d, assuming that the leakage was given by HW without physical noise. Figures 6 to 12 illustrate the power spectra of $p_{Z|\boldsymbol{X}}$ for different bit length l and masking order d, where figures (a), (b), (c) are maskings with period of two (*i.e.*, Boolean), odd number, and 2^l, respectively. Note that we omit $l = 1$ as it is a trivial case. Here, $m = 3$, 7, 31, 127 (*i.e.*, figures (b) for $l = 2$, 3, 5, and 7, respectively) are prime, while others are composite.

The most important point would be that $\hat{p}_{Z|\boldsymbol{X}}(2^l - 1 \mid \boldsymbol{X}) = 1$ for Boolean masking, which indicates that the power spectrum density at $w = 2^l - 1$ is not decayed no matter how large d is if no physical noise, meaning that Boolean masking cannot be secure in this situation. In contrast, arithmetic masking has no peak of one, yielding an asymptotic security even without physical noise.



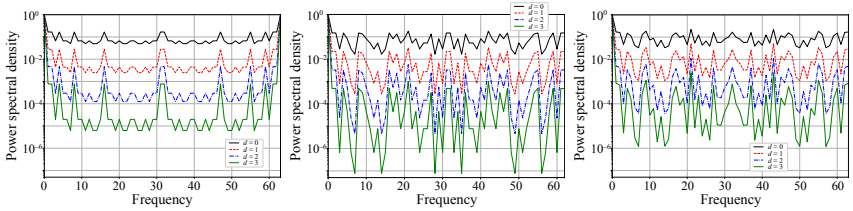
(a) Boolean (b) $m = 15$ (c) $m = 16$

Fig. 8: Power spectra of $p_{Z|\boldsymbol{X}}$ with $l = 4$.



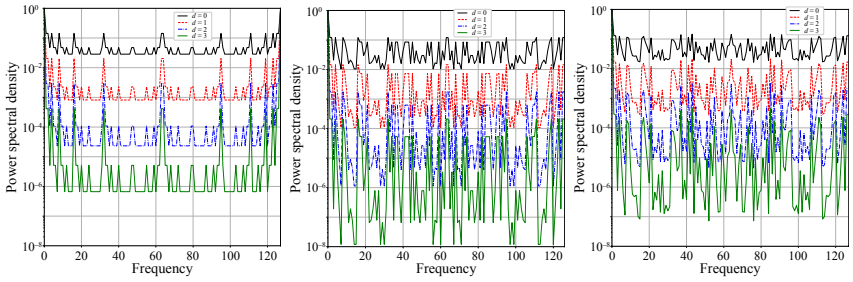
(a) Boolean (b) $m = 31$ (c) $m = 32$

Fig. 9: Power spectra of $p_{Z|\boldsymbol{X}}$ with $l = 5$.



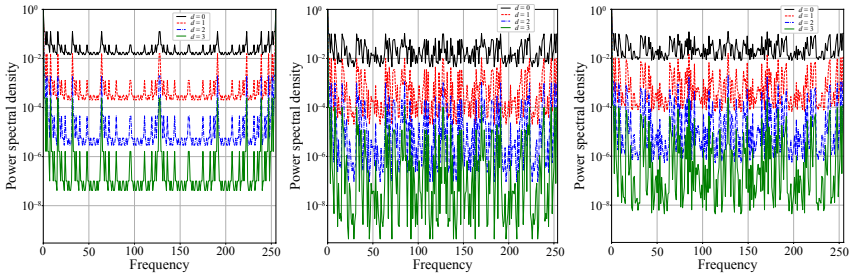
(a) Boolean (b) $m = 63$ (c) $m = 64$

Fig. 10: Power spectra of $p_{Z|\boldsymbol{X}}$ with $l = 6$.



(a) Boolean (b) $m = 127$ (c) $m = 128$

Fig. 11: Power spectra of $p_{Z|\boldsymbol{X}}$ with $l = 7$.



(a) Boolean (b) $m = 255$ (c) $m = 256$

Fig. 12: Power spectra of $p_{Z|\boldsymbol{X}}$ with $l = 8$.