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 over any finite abelian group—including Boolean and arithmetic maskings—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) [51]. In a d-th order masking, secret variable z is decomposed into $d+1$ random shares s_i $(0 \leq i \leq d)$ 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 the 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 [24], 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 [7,8,36,38,48,58,65]. 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 [7,36, 48, 58, 65], 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 [73].

Currently, the d-probing model (or d-threshold probing) [46] 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 [36], Duc *et al.* presented an SR upper bound by MI using a reduction from noisy leakage (NL) [35] to probing model through random probing (RP) model [35][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, 16, 20, 35, 76]. By contrast, the reduction *via* RP incurs a significant loss, resulting in a weak bound, as evaluated in comparison with actual SCAs in [27, 48]. 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 [48], Ito *et al.* gave another bound and proof based on communication-channel modeling of higher-order SCA with Fano's inequality [33, 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 [58], 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

[4] We use the term *reduction* for derivation of inequality among masking security/adversarial notions as in literature like [8, 36], while the meaning would be different from the provable security framework in cryptology [72]. In this paper, *reduction-based bounds/proofs* are defined as those based on reductions among (S)NL and RP.

circuit compiler and practical evaluations, have been conducted in the NL and RP, which cannot incorporate these stronger bounds in [7, 48, 58]. 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, 19, 29–32, 39, 61, 63]. More importantly, arithmetic masking, especially prime-field masking, is claimed to offer higher security than Boolean masking [38]. 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, 21, 57], 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 [18] 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 [14]), 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 Our contribution

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 [36]. Using it, we establish a new 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. 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 in [27,48]), 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 hold for masking over general abelian group including Boolean and arithmetic maskings. They are followed by a formal security proof, stating an exponential decay of SR 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

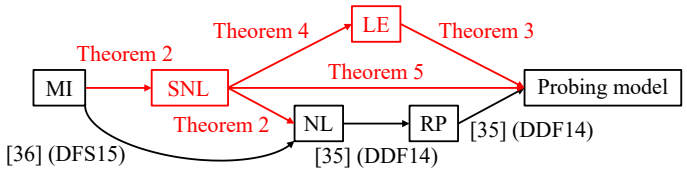


Fig. 1: Reductions among MI and masking security notions/models.

is prime, γ is the 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 $d \to \infty$. Our proof exhibits the importance of *subgroup leakage* with formal treatment; namely, the information of subgroup to which each share belongs 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) in the LE notion. Using $\mathsf{gXEX\text{-}A}^d$, we also present AOCB, an OCB-style authenticated encryption (AE) [69]. 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 Technical overview

In the reduction proofs, we leverage spectral analyses using discrete Fourier transform (DFT) in combination with Parseval's theorem. For the purpose, SNL is defined to enable a Fourier analysis, yielding a stronger bound than existing reduction-based ones. 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 [35], 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

[5] Our proof includes a resolution to open problem about noise requirement for asymptotic security of masking by Dziembowski *et al.* in TCC 2016 [38]. Jahandideh *et al.* claimed to prove the conjecture independently in Asiacrypt 2025 [50], while it is limited to Boolean masking and arithmetic masking over cyclic group. Our proof is based on a different approach. It enables us to discuss the noise requirement in MI beyond NL, its necessity and sufficiency, general finite abelian group including non-cyclic groups, asymptotic behavior, and the properties of leakage that render masking (in)secure in Section 5.6.

Table 1: Comparison of distance-based adversarial notions

Notion	Definition	Major purpose/usage
Bias [65]	Expected L_2 distance	Analysis of classes of noisy leakage function and leaking computation
NL [35]	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

[65, 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 analysis *via* Parseval's theorem. Importantly, our proofs have no specific assumption about leakage (*e.g.*, leakage is given from Hamming weight (HW) or least significant bit (LSB)), while some related analyses is devoted to specific leakage (*e.g.*, [41]).

To analyze the asymptotic behavior of masking and derive a necessary and sufficient condition of asymptotic security, we utilize a new concept of *degraded leakage*. Its mutual information with a corresponding share is defined to be a lower bound of $I(\boldsymbol{S}_i;\boldsymbol{L}_i)$, which preserves only a subgroup leakage with probability 1 but discards other information. The degraded leakage and proof emphasize the essential role of subgroup leakage in masking.

Our analysis and consequence are highly related to local leakage analysis of linear secret sharing in [13], which utilizes Fourier spectral analysis and Parseval's theorem and focuses on subgroup leakages. However, our technical contribution is distinct from the previous study. Our focus is devoted to a posteriori probability distribution, its power spectral density/energy, and SR through information-theoretic methodologies (*e.g.*, mutual information and Pinsker's, Fano's, and data processing inequalities), while the paper [13] focuses on a different aspect of Fourier coefficients and utilizes distinct methodologies of, for example, Poisson summation formula and combinatorics.

AOCB achieves a provable security of AE and leakage resilience of *leveled implementation* when combined with [59], assuming that the underlying BC (typically AES) is black-box secure. Its security claim is built on the LE notion and the line of existing LR-AE methodology, while we establish another concept of leveled implementation relating to masking order.

1.4 Paper organization

Section 2 introduces the mathematical notations and basics of masking. Section 3 reviews the existing security notions and proofs. 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 validates it experimentally. Appendices provide proofs of lemmas, theorems, and corollaries, a security proof of AOCB, and visualization of power spectra under HW leakage.

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 $X^n = (X_0, X_1, \dots, X_{n-1})$. 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 l-dimensional $(m_1, m_2, \dots, m_l)$-point DFT of a function $f : \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \cdots \times \mathbb{Z}_{m_l} \to \mathbb{R}$ is denoted by $\hat{f} : \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \cdots \times \mathbb{Z}_{m_l} \to \mathbb{C}$. Walsh–Hadamard transform (WHT) is a special case of high-dimensional DFT over binary ring with $m_j = 2$ for any j. 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 Appendix 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 $\mathbb{Z}_2^l$.

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 $s_0, s_1, \dots, s_d$ 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, 38, 57].

The security of masking is based on threshold secret sharing; 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 [46], 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 [46]. Although it was originally defined using a simulation of joint probability distribution, we introduce an intuitive definition that has been commonly used [68].

Definition 1 (Probing model [46, 68], informal). *A masked circuit is d-probing (or d-threshold probing) secure if any d combination of probes in the circuit never 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 [22]) 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 [35], 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 [35], 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) [65]). *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 [42, 64] 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 [36]. Another important characteristic of NL is that it is reduced to the probing model through the RP model [35].

Definition 3 (Random probing (RP) [35]). *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 [36], 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 [27, 48], 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 [38, 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 [50].

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 [33]. Subsequently, in 2022 [48], 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 [38], 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 $H_2 : [0,1] \to [0,1] : r \mapsto -r\log r - (1-r)\log(1-r)$ 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 [58], Masure *et al.* developed similar (slightly weaker) bound for additive masking not limited to Boolean by the *XOR lemma* [38]. Furthermore, in [7], Béguinot *et al.* derived another bound using the entropy power inequality [33, 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 [27, 48]. 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 [48] 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 ($0 \le i \le d$).
- $\mathbb{G}$: Additive group of field/ring of the masking. We consider arbitrary finite abelian group[9] of $\mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \cdots \times \mathbb{Z}_{m_j} \times \cdots \times \mathbb{Z}_{m_l}$. For instance, $\mathbb{G} = \mathbb{Z}_2^l$ (*i.e.*, $m_j = 2$ for all j) for l-bit Boolean masking and $\mathbb{G} = \mathbb{Z}_m$ (*i.e.*, $l = 1$) for arithmetic masking (we write $m = m_1$ when $l = 1$).
- $m \in \mathbb{N}$: The order of the additive group, which is equal to $\prod_{j=1}^{l} m_j$.
- $j \in \mathbb{N}$: Index of invariant factors of $\mathbb{G}$ ($1 \le j \le l$).
- $K \in \mathbb{G}$: Secret key, which is assumed to be sampled from a uniform distribution over $\mathbb{G}$, 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.
- $(S_0, S_1, \ldots, S_d)^n \in (\mathbb{G}^{d+1})^n$: Masking shares of Z. Note that $Z = S_0 + S_1 + \cdots + S_d$ holds and each S_i is uniformly distributed over $\mathcal{S} = \mathcal{Z} = \mathbb{G}$.
- $(\boldsymbol{L}_0, \boldsymbol{L}_1, \ldots, \boldsymbol{L}_d)^n \in ((\mathbb{R}^a)^{d+1})^n$: Leakages of S_j. Each $\boldsymbol{L}_j$ is defined as an a-dimensional vector over $\mathbb{R}$, where a is the number of sampling points. As in literature, $\boldsymbol{L}_0, \boldsymbol{L}_1, \ldots, \boldsymbol{L}_d$ are assumed to be mutually independent. Let $\boldsymbol{X} = (\boldsymbol{L}_0, \boldsymbol{L}_1, \ldots, \boldsymbol{L}_d)$ be a side-channel trace, which is a $(d+1)$-dimensional vector of all leakages (namely, $a \times (d+1)$-dimensional vector in total).

[9] According to Kronecker's structure theorem, any finite abelian group is isomorphic to $\mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \cdots \times \mathbb{Z}_{m_l}$ for some integers $m_1, m_2, \ldots, m_l$ and l.

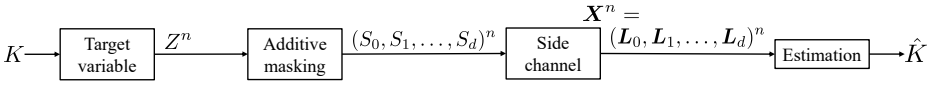


Fig. 2: Communication channel of n-trace SCA on d-th order additive masking.

– $\hat{K} \in \mathbb{G}$: 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 [27, 48].

Theorem 1 (Upper bound of SR by MI [27, 48]). *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 $H_2(r) = -r\log r - (1-r)\log(1-r)$ is the binary entropy function.

Proof. See Appendix B.

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. Intuitively, it represents the information amount of secret key bits required to achieve a given SR [48]. 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 [73] 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) [36]). *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.*, $I(Z; \boldsymbol{L}_0, \boldsymbol{L}_1, \dots, \boldsymbol{L}_{i-1}, \boldsymbol{L}_{i+1}, \dots, \boldsymbol{L}_d) = 0$ holds for any i) owing to the information-theoretic security of secret sharing [70]. 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 probes 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 obtain $\boldsymbol{L}_i$ 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 [36, 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 [65]) is an important future study.

Remark 2 (Physical defaults). Masking security order sometimes degrades due to share-interaction such as glitches [56], coupling [28], and microarchitectural interaction [43]. 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 [40,66]. 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. See Appendix C.

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 [48] (whereas it was proven only for Boolean masking in [48]). Here, Parseval'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*

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

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 [48, 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 the DFT of $p_{S_i|\boldsymbol{L}_i}(\cdot \mid \boldsymbol{L})$ over $\mathbb{G}$.

Proof. See Appendix D.

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. See Appendix E.

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*

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

Proof. See Appendix F.

Remark 3 (Meaning of Lemma 2 and its 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 $\lambda = \max_i \mathbb{E}_{\boldsymbol{L}}[(\sum_s |p_{S_i|\boldsymbol{L}_i}(s \mid \boldsymbol{L}) - p_{S_i}(s)|)^2]$. An λ-SNL adversary implies $(m-1)\lambda$-LE adversary. Namely, it holds*

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

Proof. See Appendix G.

5.3 Reduction from MI/SNL to d-probing

Theorem 5 (Reduction from MI/SNL to d-probing). *Consider Figure 2. Let $\epsilon = \max_i I(S_i; \boldsymbol{L}_i)$ and $\lambda = \max_i \mathbb{E}_{\boldsymbol{L}}[(\sum_s |p_{S_i|\boldsymbol{L}_i}(s \mid \boldsymbol{L}) - p_{S_i}(s)|)^2]$. 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*

$$\begin{aligned} \xi(\mathrm{SR}_d) &\le n \log\left(1 + (m-1)\lambda^{d+1}\right) \\ &\le n \log\left(1 + (m-1)(2\ln 2)^{d+1}\epsilon^{d+1}\right). \end{aligned} \tag{6}$$

Proof. See Appendix H.

Note that LE is also bounded as $\xi(\mathrm{SR}_d) \leq n \log\left(1 + (m-1)^{d+1}\lambda^{d+1}\right)$ by applying Lemma 2 to Theorem 3, which is looser than the direct bound in (6) according to loss incurred by Jensen's inequality. The derived SR bound by MI is the same as [48, Lemma 4.6] for Boolean masking; however, our new proof using Fourier analysis generalizes the result to general finite abelian group including arithmetic masking and clarifies the relationship between the communication channel in Figure 2 and masking security/adversarial notions including the NL through our security reductions.

5.4 Asymptotic and provable security of masking

Similar to [48], 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*.*

Proof. See Appendix I.

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.* [38] 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*

[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 any finite groups including *arithmetic* masking with *non-prime* m.

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*

$$\frac{\delta}{2} < 1 - \frac{|\mathbb{H}|}{|\mathbb{G}|} = 1 - \max_{1\le j\le l} \max_{1\le w_j\le m_j-1} \frac{g_{j,w_j}}{m_j}, \tag{7}$$

$$\epsilon < \min_{1\le j\le l} \min_{1\le w_j\le m_j-1} \log\left(\frac{m_j}{g_{j,w_j}}\right), \tag{8}$$

where $\mathbb{G}$ *is the additive group of field/ring;* $\mathbb{H}$ *is the largest proper subgroup of* $\mathbb{G}$*;* m_j *is the* j*-th invariant factor of* $\mathbb{G} = \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \cdots \times \mathbb{Z}_{m_l}$*; and* $g_{j,w_j} = \gcd(w_j, m_j)$ *is the greatest common divisor (gcd) of* w_j *and* m_j*.*

Proof. See Appendix J.

In Theorem 7, w_j represents the frequency at the j-th invariant factor. Satisfying either inequality (7) or (8) is sufficient. Let $\gamma_j = \min_{w_j} m_j/g_{j,w_j}$ be the smallest divisor of m_j (if m_j is a prime, $\gamma_j = m_j$) and let $\gamma = \min_j \gamma_j$. The largest proper subgroup $\mathbb{H}$ has order $|\mathbb{G}|/\gamma$, where $|\mathbb{H}|/|\mathbb{G}| = 1/\gamma$ according to Lagrange's theorem. For instance, in Boolean masking, the kernel $\mathbb{H}_w = \{\, s \mid \langle s \mid w\rangle = 0 \,\}$ is a largest proper subgroup for all $w \neq 0$, and Lagrange's theorem states that $|\mathbb{H}|/|\mathbb{G}| = 1/2$ holds for any $w \neq 0$. The proof exploits the quotient group by the kernel of the character (*i.e.*, phase-equivalence subgroup) that maximizes the power spectral density. We introduce Lemma 3, which plays an essential role in proof of Theorem 7 and explanation of subgroup leakage.

Lemma 3 (Maximum of power spectral density). *Let* $p : \mathbb{G} \to \mathbb{R}$ *be a probability mass function, where* $\mathbb{G} = \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \cdots \times \mathbb{Z}_{m_l}$ *is a finite discrete set. Its power spectral densities are upper-bounded by one; namely,* $|\hat{p}(w)|^2 \le 1$ *holds for all* w*. Let* $h = (h_1, h_2, \ldots, h_l) \in \mathbb{G}$*. Let* $\chi_w(h) = \exp(2\pi \mathrm{i} \sum_j h_j w_j/m_j)$ *denote the character. The equality holds at* $w = (w_1, w_2, \ldots, w_l)$ *if and only if* $\mathrm{supp}(p) \subset h' + \ker(\chi_w)$ *for some* h'*; namely, for some constant* $c \in \mathbb{Q}/\mathbb{Z}$*, it holds*

$$p(h) \begin{cases} \ge 0 & \text{if } \sum_{j=1}^{l} h_j w_j/m_j = c \bmod 1, \\ = 0 & \text{otherwise.} \end{cases}$$

When $\mathbb{G}$ *is isomorphic to* $\mathbb{Z}_m$ *with a prime* m*, a necessary and sufficient condition that the equality holds for any* w *is that* p *is a one-hot distribution.*

Proof. See Appendix K.

The equality condition means that $p(h)$ can take non-zero value only for input h that has a specific phase. Hence, when $l = 1$ and m is prime, the power spectral densities are maximized for all w if and only if p is a one-hot distribution.

The condition of (8) 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 [36], $I(S_i; \boldsymbol{L}_i) < (2 \ln 2)^{-1} \approx 0.72$ using communication-channel-based proofs [7, 48], and $I(S_i; \boldsymbol{L}_i) < 1$ in [50] limited to Boolean masking. Importantly, the minimum noise requirement is determined by $\min_j \gamma_j$ independent of l; the security of masking is highly relevant to the structure of each cyclic factor (coordinate group). This implies that we can analyze the masking over $\mathbb{G} = \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \cdots \times \mathbb{Z}_{m_l}$ in a coordinate-wise manner, especially in terms of *subgroup leakage*.

Hence, for the simplicity, we here elaborate the subgroup leakage of the arithmetic masking ($l = 1$) and boolean masking ($m_j = 2$ for any j) as important practical and smallest cases in l and m_j, respectively. Let m be the cardinality of $\mathbb{G}$ (equal to $\mathbb{Z}_m$ for arithmetic masking and $\mathbb{Z}_2^l$ for Boolean masking) and let w be the frequency of DFT or WHT over $\mathbb{G}$. For arithmetic masking, let $g_w = \gcd(w, m)$ and $\gamma_w = m/g_w$, and let $Y_i^{(w)} = S_i \bmod \gamma_w$ be a random variable on $\mathbb{Z}_{\gamma_w}$ that follows the probability mass function of $p_{Y_i^{(w)}}(y) = \sum_{s_1=0}^{g_w-1} p_{S_i}(y+\gamma_w s_1)$. For Boolean masking, let $Y_i^{(w)} = \langle S_i \mid w \rangle$ be a random variable over $\mathbb{Z}_2$. The random variable $Y_i^{(w)}$ plays an essential role in the proof of Theorem 7 as a coset associated with w to decompose the DFT/WHT into subgroups, and we here explain intuitively how it is important.

Meaning of $Y_i^{(w)}$ and subgroup leakage. The random variable $Y_i^{(w)}$ represents a coset of $\ker(\chi_w)$ 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 the 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 the attacker can estimate the subgroup to which each share belongs. 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. For example, Boolean masking induces the largest proper subgroup for any $w \neq 0$. For given $w \neq 0$, the corresponding subgroups consist of elements with $\langle S_i \mid w \rangle = 0$ or 1. An arithmetic masking induces the largest proper subgroup at $w = m/\gamma$. Generally, it holds that

$$\hat{p}_{S_i|\boldsymbol{L}_i}(2^l - 1 \mid \boldsymbol{L}_i) = \Pr[\langle S_i \mid w \rangle = 0 \mid \boldsymbol{L}_i] - \Pr[\langle S_i \mid w \rangle = 1 \mid \boldsymbol{L}_i], \tag{9}$$

$$\hat{p}_{S_i|\boldsymbol{L}_i}(m/\gamma \mid \boldsymbol{L}_i) = \sum_{s=0}^{\gamma-1} \mathrm{e}^{-2s\pi\mathrm{i}/\gamma} \Pr[S_i \bmod \gamma = s \mid \boldsymbol{L}_i], \tag{10}$$

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 $\epsilon < \min_j \log \gamma_j$ (where $\gamma_j = \min_{w_j}(m_j/g_{j,w_j})$) can be strengthened. The answer is no if no assumption about leakage is used. Lemma 4 states that $I(Z;\boldsymbol{X})$ does not converge 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* γ_w *is the order of quotient group associated with* w *(e.g.,* $\gamma_w = 2$ *and* m/g_w *for Boolean and arithmetic masking, respectively). It holds that*

$$I(Z;\boldsymbol{X}) \to \log\left(1 + \left|\mathcal{W}_{\mathrm{peak}}\right|\right),$$

as $d \to \infty$*. In particular, for arithmetic masking, the limit is given by*

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

Proof. See Appendix L.

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. See Appendix M.

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.[11] 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 belongs from leakages of shares with probability 1. In other words, if the 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 necessary and sufficient condition as Theorem 8.

Theorem 8 (Necessary and sufficient condition for asymptotic security of masking). *Any 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, where γ_w is the order of quotient group associated with w.*

Proof. See Appendix N.

Relationship between $I(S_i; \boldsymbol{L}_i)$ and $I(Y_i^{(w)}; \boldsymbol{L}_i)$. Even if $I(S_i; \boldsymbol{L}_i) \geq \log \gamma_w$ for some w, masking can be asymptotically secure, because $I(S_i; \boldsymbol{L}_i) \geq \log \gamma_w$ is not a 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 $I(S_i; \boldsymbol{L}_i) \geq \log \gamma_w$ holds for some w. For example, consider a probability distribution satisfying $p_{S_i|\boldsymbol{L}_i}(s \mid \boldsymbol{l}) = 1 - \nu(m-1)$ 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 the 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). *Any masking over finite abelian group is 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. See Appendix O.

[11] This is called *homomorphic attack* in [38] and parity-of-parity attack in [55] as local leakage of linear secret sharing schemes. In [41], a formal and quantitative analysis of LSB and HW leakages for prime-field masking is shown.

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), $2^\ell - 1$, and 2^ℓ, respectively, where ℓ is the bit length for arithmetic masking. See Appendix S. 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, 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.

6.2 Evaluation of success rate

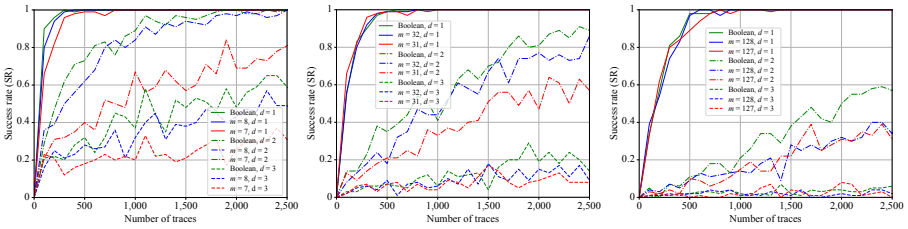
We demonstrate a simulation of template attack on masked implementations to evaluate its SR. In the simulation, the initial key addition of $Z = M + k$ was targeted, and we supposed that Z was masked and the attacker observes the leakages of shares. The leakage was given as $\boldsymbol{L}_i = \mathrm{HW}(S_i) + N$, 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 [25], which was proven optimal [45]. 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 3-bit, we found non-trivial differences among Boolean, 2^l-characteristic, and prime-field maskings. Arithmetic maskings are more secure than Boolean

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

Order	1-bit	2-bit			3-bit			4-bit		
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	5-bit			6-bit			7-bit			8-bit		
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) 3-bit (b) 5-bit (c) 7-bit

Fig. 3: SR of template attack.

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 $\ell = 7$, we found little difference between 2^ℓ-characteristic and prime. Recall that the difference of log-energy between arithmetic maskings are smaller when the bit length 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 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 [52] and XTS [37] employ *masked* BCs in their construction to build a tweakable BC (TBC) [54], which is an extension of classical BC. A TBC has a public value called a tweak, and a tuple of a secret key and a tweak determines a permutation over the message space. A popular instance of BC-based TBC is called XEX, proposed by Rogaway [69]. The structure of XEX is defined as $\mathsf{XEX}[E_K]((K,K'),M,T) \triangleq E_K(M \oplus f_{K'}(T)) \oplus f_{K'}(T)$, 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. The value $f_{K'}(T)$ works as the mask inside XEX. 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 [74]. 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

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

We observe that $\mathsf{gXEX}[E_K, \mathbb{G}]$ is a (black-box) secure TBC if E is a secure BC 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 [54, 69]. For completeness, we present the proof in the full version.

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 $\max_{T,T' \in \mathcal{T}, T \neq T', \beta \in \mathbb{G}} \Pr[H_K(T) \boxminus H_K(T') = \beta] \leq \alpha$ holds for random K. An α-AΔU($\mathbb{F}_{2^b}$) hash function H, where $\boxminus$ 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 [23], 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 [23] 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 [49] 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 $H^+_{K'}(T) \triangleq H_{K'_1}(T) \oplus K'_2$, where $K' = (K'_1, K'_2)$. If H is α-AXU for some $\alpha \in [0, 1]$, H^+ is $\alpha/2^b$-ASU.*

Proof. See Appendix P.

Lemma 6 (ASU implies AΔU). *If $H : \mathcal{K} \times \mathcal{T} \to \mathbb{G}$ is an α-ASU, H is an $\alpha|\mathbb{G}'|$-AΔU$(\mathbb{G}')$ 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. See Appendix Q.

Lemmas 5 and 6 suggest that, by XORing another independent key material to a mask generation function of known XEX instances such as [44,52,54,60,69], 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 gXEX-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 gXEX-A^d is AΔU. As aforementioned, using this fact with existing proofs [23,54], with a minor extension to an AΔU mask generation instead of an AXU one, gXEX-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 gXEX-X^d to mean the counterpart of gXEX-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 [34], the overhead of gXEX-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 gXEX-A^d

One important question is whether our theory is applicable to gXEX-A^d because its encoding is different from standard masking. In [74], 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 gXEX-A^d.

We show that our claims on masking hold for AES with gXEX-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 $1 \leq i \leq d$. Addition and subtraction of gXEX-A^d are supposedly modulo 2^8, while AES key addition is XOR. Consider an SCA on the AES final round[12], 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

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

C. Let $\boldsymbol{L}_0$ denote the leakage of Z. The probability distribution of $p_{Z|\boldsymbol{X}}$ with gXEX-A^d is expressed using convolutions as

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

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 in the LE model, as the LE bound of $I(Z; \boldsymbol{X}) \leq \log(1 + \tau^{d+1})$ holds for gXEX-A^d by renaming $Z = S_0$. This is consistent with the claim and evaluation results by XCCA attack in [74]. Thus, LE is useful to evaluate the SCA resistance.

A difference between gXEX-A^d and XCCA attack in [74] 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* [47], 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 [47, 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 [59].[13] 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

[13] We adopt a minor refinement suggested by [73] 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. $T \leftarrow T \oplus \mathsf{PHASH}_{\widetilde{E}_K}(N, A)$
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 $\mathsf{PHASH}_{\widetilde{E}_K}(N, A)$

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. $\Sigma \leftarrow M[1] \oplus \cdots \oplus M[m]$
5. $T^* \leftarrow \widetilde{E}_K(\Sigma, (N, m, 1))$
6. $T^* \leftarrow T^* \oplus \mathsf{PHASH}_{\widetilde{E}_K}(N, A)$
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.

using known incremental AXU functions [52, 69] 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 as a variant of ΘCB [52], whose security is reduced to the underlying TBC $\widetilde{E}$. See Section R.6. presents the proof of this claim (mostly identical to that of ΘCB). Assuming KDF is a PRF, each mask of $\mathsf{gXEX\text{-}A}^d$ is independent, and $\mathsf{gXEX\text{-}A}^d$ is a provably secure TBC built on BC.

Leakage resilience claim as leveled implementation. We leverage the core idea of Mennink towards LR-AE [59]; 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 by effectively combining a few DPA-resistant (*e.g.*, d-th order masked) mod-

ules for minor parts—only initialization—and non-masked gXEX-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 gXEX-A^d using the number of traces equal to gXEX-A^d calls. Second, let d' be the security order of the initialization 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 gXEX-A^d for encryption and AD processing parts. Nonce-disrespecting attacks and attacks on initialization are called *trivial* in [15]. 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 by SCA countermeasure and be nonce-respecting, as in existing LR-AEs based 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 [67]. However, blind SCAs have difficulty in achieving high SRs even when compared to higher-order SCAs[14]. Thus, AOCB requires noise such that blind SCAs get difficult, as blind SCAs are more noise-sensitive than non-blind SCAs.

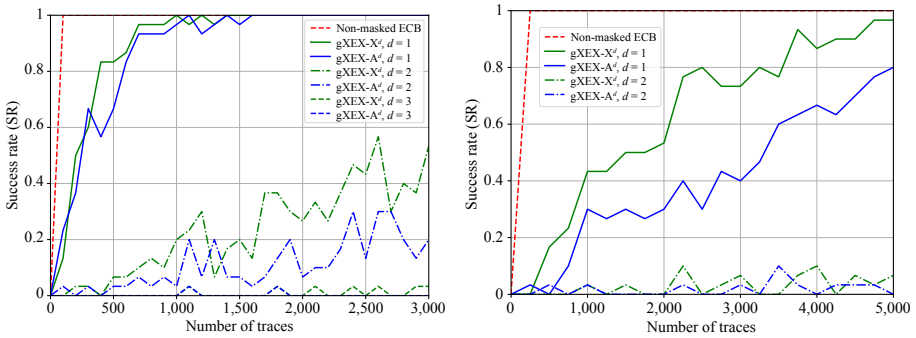
Comparison with existing (T)BC-based LR-AE schemes. Existing (T)BC-based LR-AEs of leveled implementation (*e.g.*, TEDT [17] and Triplex [71]) claim that all (T)BC modules except for initialization only require SPA-resistance (*i.e.*, security against one-trace SCA). Yet, none achieved rate-one (*i.e.*, needing one BC call for one input block).[15] 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 in Section 6.2,

[14] 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 [47, Lemma 2]. Full-key recovery of AES (*i.e.*, simultaneous successes of 16 partial keys) requires significantly high SR even with key enumeration [75].

[15] The OCB-like LR-AE scheme presented at [59] is rate-one in terms of TBC; however, inside the TBC, the scheme needs multiple calls of a block cipher and a universal hash function. Mennink [59] briefly mentioned potential efficiency improvement by optimizing the design including the internal universal hash without further details.



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

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

whereas the number of trials was 30 due to its heavy computation. 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.

8 Conclusion

We presented a formal security analysis of masking security without relying on the RP. Our analysis introduced new notions—SNL and LE—and derived stronger SR upper bounds for both Boolean and arithmetic maskings than existing reduction-based ones in the NL–RP framework, enabling a proof of asymptotic security of masking along with revealing a necessary and sufficient condition. A numerical evaluation validated our theory and claims. As a practical application, we presented gXEX-A^d, an LR primitive useful to realize OCB-style AE with intrinsic SCA security of arithmetic masking by design. It is important to extend our modeling and analyses to circuits beyond encoding in future. Formal proof of AE security of AOCB *with leakage oracles* is also a future study.

Acknowledgments

We sincerely thank all the anonymous reviewers for their feedback. This work has been supported by JSPS Kakenhi Grant No. 25K03117, JST CRONOS Grant No. JPMJCS25N1, and JST K Program Grant No. JPMJKP24C1, Japan.

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 Cryptology—ASIACRYPT 2021. pp. 157–188 (2021)

13. Benhamouda, F., Degwekar, A., Ishai, Y., Rabin, T.: On the local leakage resilience of linear secret sharing schemes. In: Advances in Cryptology—CRYPTO 2018. pp. 531–561. Springer International Publishing (2018)
14. Bernstein, D.J.: ChaCha20, a variant of Salsa20 (2025), https://cr.yp.to/chacha/chacha-20080128.pdf
15. 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
16. Berti, F., Faust, S., Orlt, M.: Provable secure parallel gadgets. IACR Transactions on Cryptographic Hardware and Embedded Systems **2023**(4), 420–459 (2023)
17. 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
18. Beyne, T., Verbauwhede, T.: Integral cryptanalysis in characteristic p. Asiacrypt 2025 (2025), https://eprint.iacr.org/2025/932
19. 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
20. 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)
21. 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
22. 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
23. 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)
24. 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)
25. Chari, S., Rao, J.R., Rohatgi, P.: Template attacks. In: International Workshop on Cryptographic Hardware and Embedded Systems. pp. 13–28. LNCS (2002)
26. Chen, S., Steinberger, J.P.: Tight security bounds for key-alternating ciphers. pp. 327–350 (2014). https://doi.org/10.1007/978-3-642-55220-5_19
27. 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)
28. de Cnudde, T., Bilgin, B., Gierlichs, B., Nikov, V., Nikova, S., Rijmen, V.: Does coupling affect the security of masked implementations? In: International Work-

shop on Constructive Side-Channel Analysis and Secure Design (COSADE 2017). Lecture Notes in Computer Science, vol. 10348, pp. 1–18 (2017)
29. 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
30. 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
31. 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
32. 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
33. Cover, T.M., Thomas, J.A.: Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing). Wiley-Interscience, USA (2006)
34. 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)
35. 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)
36. 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)
37. 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)
38. Dziembowski, S., Faust, S., Skórski, M.: Optimal amplification of noisy leakages. In: Theory of Cryptography. pp. 291–318 (2016)
39. 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
40. 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)
41. Faust, S., Masure, L., Micheli, E., Orlt, M., Standaert, F.: Connecting leakage-resilient secret sharing to practice: Scaling trends and physical dependencies of prime field masking. In: Joye, M., Leander, G. (eds.) Advances in

Cryptology—EUROCRYPT. pp. 316–344 (2024). https://doi.org/10.1007/978-3-031-58737-5_12
42. 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
43. 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)
44. 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)
45. 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)
46. Ishai, Y., Sahai, A., Wagner, D.: Private circuits: Securing hardware against probing attacks. In: Advances in Cryptology—CRYPTO 2003. pp. 463–481 (2003)
47. 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
48. 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)
49. 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)
50. Jahandideh, V., Mennink, B., Batina, L.: A decomposition approach for evaluating security of masking. In: Advances in Cryptology—ASIACRYPT 2025. pp. 580–612 (2025)
51. 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)
52. 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)
53. Liskov, M., Rivest, R.L., Wagner, D.: Tweakable block ciphers. pp. 31–46 (2002). https://doi.org/10.1007/3-540-45708-9_3
54. 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)
55. Maji, H.K., Nguyen, H.H., Paskin-Cherniavsky, A., Suad, T., Wang, M., Ye, X., Yu, A.: Tight estimate of the local leakage resilience of the additive secret-sharing scheme & its consequences. In: Information-Theoretic Cryptography (ITC) (2022)
56. 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)
57. 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)
58. 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
59. 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)
60. 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)
61. 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)
62. Patarin, J.: The "coefficients H" technique (invited talk). pp. 328–345 (2009). https://doi.org/10.1007/978-3-642-04159-4_21
63. 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
64. Pinsker, M.S., Feinstein, A.: Information and information stability of random variables and processes, vol. 12. Holden-Day San Francisco (1964)
65. Prouff, E., Rivain, M.: Masking against side-channel attacks: A formal security proof. In: Advances in Cryptology—EUROCRYPT 2013. pp. 142–159 (2013)
66. Reparaz, O., Bilgin, B., Nikova, S., Gierlichs, B., Verbauwhede, I.: Consolidating masking schemes. In: Advances in Cryptology—CRYPTO 2015. pp. 764–783 (2015)
67. 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)
68. 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)
69. 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)
70. Shamir, A.: How to share a secret. Communications of the ACM **22**(11), 612–613 (1979). https://doi.org/10.1145/359168.359176
71. 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)
72. Tibouchi, M.: Security Reduction, pp. 2333–2335. Springer Nature Switzerland (2025). https://doi.org/10.1007/978-3-030-71522-9_515
73. 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)
74. 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
75. 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)

76. 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 (High-dimensional DFT). *Let f be a function $\mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \cdots \times \mathbb{Z}_{m_l} \to \mathbb{R}$. Its l-dimensional $(m_1, m_2, \ldots, m_j, \ldots, m_l)$-point DFT $\hat{f} : \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \cdots \times \mathbb{Z}_{m_l} \to \mathbb{C}$ is defined by*

$$\hat{f}(w) = \sum_{s_1=0}^{m_1-1} \sum_{s_2=0}^{m_2-1} \cdots \sum_{s_l=0}^{m_l-1} f(s_1, s_2, \ldots, s_l) \exp\left(-\mathrm{i} \sum_j s_j \Omega_{w_j}^{(j)}\right),$$

where $w = (w_1, w_2, \ldots, w_j, \ldots, w_l)$ is an l-dimensional frequency and $\Omega_{w_j}^{(j)} = 2\pi w_j/m_j$ is the normalized frequency at j-th invariant factor.

Definition 13 (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 $\langle \cdot \mid \cdot \rangle : \mathbb{Z}_2^l \times \mathbb{Z}_2^l \to \mathbb{Z}_2$ is an inner product.

Note that WHT is a special case of DFT over l-dimensional binary ring $\mathbb{Z}_2^l$ with $m_1 = m_2 = \cdots = m_l = 2$. 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*

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

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).

B Proof of Theorem 1

Proof. Despite the difference of communication channel model, Theorem 1 is proven in the same manner as [27, 48]. 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 [33, Theorem 2.8.1] on this Markov chain states $I(K; \hat{K}) \leq I(Z^n; \boldsymbol{X}^n)$. Then, Fano's inequality [33, 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. □

C Proof of Theorem 2

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

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

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

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

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 $\delta^2 \leq \lambda$ is proven by Jensen's inequality [33, 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 $\boldsymbol{l}$. □

D Proof of Lemma 1

Proof. For Boolean masking, Lemma 1 is proven in [48, Lemma 4.4], where $\hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L})$ is a WHT. We prove the case of general abelian group of $\mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \cdots \times \mathbb{Z}_{m_l}$ by replacing the WHT to DFT. As in the proof of [48, 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

$$\begin{aligned} I(Z; \boldsymbol{X}) &= H(Z) - H(Z \mid \boldsymbol{X}) \\ &= \log m + \mathbb{E}_{Z,\boldsymbol{X}} \log p_{Z|\boldsymbol{X}}(Z \mid \boldsymbol{X}) \\ &\leq \log \left(m \mathbb{E}_{Z,\boldsymbol{X}} p_{Z|\boldsymbol{X}}(Z \mid \boldsymbol{X}) \right) \\ &= \log \left(m \mathbb{E}_{\boldsymbol{X}} \left[\sum_z \left(p_{Z|\boldsymbol{X}}(z \mid \boldsymbol{X}) \right)^2 \right] \right). \end{aligned} \tag{11}$$

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

$$\sum_z \left(p_{Z|\boldsymbol{X}}(z \mid \boldsymbol{X}) \right)^2 = \frac{1}{m} \sum_w \left| \hat{p}_{Z|\boldsymbol{X}}(w \mid \boldsymbol{X}) \right|^2. \tag{12}$$

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

$$p_{Z|\boldsymbol{X}}(z \mid \boldsymbol{X}) = \sum_{z = s_0 + s_1 + \cdots + s_d} \prod_{i=0}^{d} p_{S_i|\boldsymbol{L}_i}(s_i \mid \boldsymbol{L}_i). \tag{13}$$

The right-hand side is cyclic convolution with a period of m. Using l-dimensional $(m_1, m_2, \ldots, m_l)$-point DFT, in the frequency domain, Equation (13) is equivalently expressed by

$$\hat{p}_{Z|\boldsymbol{X}}(w \mid \boldsymbol{X}) = \prod_{i=0}^{d} \hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L}_i). \tag{14}$$

By combining Equations (11), (12) and (14), we conclude Equation (5). □

E Proof of Theorem 3

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

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

as required. □

F Proof of Lemma 2

Proof. The proof does not depend on i. For simplicity, we first consider an arithmetic masking with a period of m over $\mathbb{Z}_m$ with $l = 1$. 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. For general abelian group of $\mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \cdots \times \mathbb{Z}_{m_j} \times \cdots \times \mathbb{Z}_{m_l}$, the proof is done by replacing the twiddle factor $\mathrm{e}^{-\mathrm{i}s\Omega_w}$ in one-dimension to that in l-dimension (*i.e.*, $\exp\left(-\mathrm{i}\sum_j s_j \Omega_{w_j}^{(j)}\right)$ where $\Omega_{w_j}^{(j)} = 2\pi w_j/m_j$). □

G Proof of Theorem 4

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. □

H Proof of Theorem 5

Proof. According to Theorems 1 and 2 and Lemmas 1 and 2, it holds that

$$\begin{aligned}\xi(\mathrm{SR}_d) &\le n \log \left(\sum_w \prod_{i=0}^{d} \mathbb{E}_{\boldsymbol{L}} \left[\left| \hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L}_i) \right|^2 \right] \right) \\ &\le n \log \left(1 + (m-1)\lambda^{d+1}\right) \\ &\le n \log \left(1 + (m-1)(2\ln 2)^{d+1}\epsilon^{d+1}\right),\end{aligned}$$

as required. □

I Proof of Theorem 6

Proof. It is proven in the same manner as [48, Lemma 5.1 and Theorem 5.2] with Theorem 3, using the convexity of ξ and its second-order Taylor expansion. □

J Proof of Theorem 7

Proof. For an intuitive proof, we first prove the case of $l = 1$ (*i.e.*, arithmetic masking), which is followed by general cases including Boolean masking.

We firstly prove that $\delta/2 < 1 - \min_{w \neq 0} g_w/m$ is sufficient. 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

$$\begin{aligned}\hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L}) &= \sum_{s_1=0}^{g_w-1} \sum_{s_0=0}^{\gamma_w-1} p_{S_i|\boldsymbol{L}_i}(s_0 + \gamma_w s_1 \mid \boldsymbol{L}) \mathrm{e}^{-\mathrm{i}s_0\Omega_w} \\ &= \sum_{s_0=0}^{\gamma_w-1} \mathrm{e}^{-\mathrm{i}s_0\Omega'_w} \sum_{s_1=0}^{g_w-1} p_{S_i|\boldsymbol{L}_i}(s_0 + \gamma_w s_1 \mid \boldsymbol{L}), \end{aligned} \tag{15}$$

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

$$\max_{i,w\neq 0} \left| \hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L}) \right|^2 = \max_{i,w\neq 0} \left| \hat{p}_{Y_i^{(w)}|\boldsymbol{L}_i}(w \mid \boldsymbol{L}) \right|^2 < 1 \tag{16}$$

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 (16) 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

$$\sum_{y=0}^{\gamma_w - 1} \mathbb{E}_{\boldsymbol{L}} \left[\left| p_{Y_i^{(w)}|\boldsymbol{L}_i}(y \mid \boldsymbol{L}) - p_{Y_i}(y) \right| \right] = \left| 1 - \frac{1}{\gamma_w} \right| + (\gamma_w - 1) \left| \frac{-1}{\gamma_w} \right| = 2 - \frac{2g_w}{m}. \tag{17}$$

Thus, the condition of δ for the asymptotic security is that δ is smaller than the minimum of Equation (18) in terms of $w \neq 0$ (as in Equation (16)); that is,

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

Divide both sides by two. This proves the condition (7) for arithmetic masking.

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 $\boldsymbol{L}_i \leftrightarrow S_i \leftrightarrow Y_i^{(w)}$ 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 $\boldsymbol{L}_i \leftrightarrow S_i \leftrightarrow Y_i^{(w)}$ states a data processing inequality $I(Y_i^{(w)}; \boldsymbol{L}_i) \leq I(S_i; \boldsymbol{L}_i)$. 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 for arithmetic masking.

General case of $\mathbb{G} = \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \cdots \times \mathbb{Z}_{m_l}$. We can prove Theorem 7 by generalizing the definition of $Y_i^{(w)}$ using the quotient group of $\mathbb{G}$ by the kernel of character. Let Let $s_i = (h_1, h_2, \ldots, h_l) \in \mathbb{G}$ and $w = (w_1, w_2, \ldots, w_l)$ denote the input and frequency at the l-th invariant factor, respectively (note that this argumentation is independent of i). Let $\chi_w(s_i) = \exp(2\pi \mathrm{i} \sum_j h_j w_j / m_j)$ denote the character. At frequency w, the DFT decomposition can be done using the subgroup $\mathbb{H}_w = \ker(\chi_w)$ and the quotient group $\mathbb{Q}_w = \mathbb{G}/\mathbb{H}_w$ (note that $\ker(\chi_w) = \{\, y \mid y = 0 \bmod \gamma_w \,\} = \gamma_w \mathbb{Z}$ when $l = 1$). Let $Y_i^{(w)} = S_i + \ker(\chi_w)$ be a random variable of $\mathbb{G}/\mathbb{H}_w$. With this, a necessary condition for asymptotic security is that Equation (16) holds for any w. As same as the case of arithmetic masking, for any $w \neq 0$, the total variation distance is maximized when $p_{Y_i^{(w)}|\boldsymbol{L}_i}$

is a one-hot distribution. The total variation distance is given by

$$\sum_{y\in\mathbb{Q}_w}\mathbb{E}_{\boldsymbol{L}}\left[\left|p_{Y_i^{(w)}|\boldsymbol{L}_i}(y\mid\boldsymbol{L})-p_{Y_i}(y)\right|\right]=\left|1-\frac{1}{|\mathbb{Q}_w|}\right|+(|\mathbb{Q}_w|-1)\left|\frac{-1}{|\mathbb{Q}_w|}\right|$$
$$=2-\frac{2\,|\mathbb{H}_w|}{|\mathbb{G}|}. \tag{18}$$

Thus, the condition should be $\delta/2<1-\max_w(|\mathbb{H}_w|\,/\,|\mathbb{G}|)$, meaning that the noise requirement is determined by the largest proper subgroup.

According to the first isomorphism theorem, $\mathbb{G}/\ker(\chi_w)$ is isomorphic to $\mathrm{Im}(\chi_w)$. Because $\mathrm{Im}(\chi_w)\subset\mathbb{C}^\times$ (as $\mathrm{Im}(\chi_w)$ is a finite set of unit complex numbers), the image $\mathrm{Im}(\chi_w)$ (and hence $\mathbb{G}/\ker(\chi_w)$) are isormorphic to a finite cyclic group, that is, $\mathbb{Z}_{m'}$ for some integer m'. Hence, the cardinality of $\mathbb{Q}_w$ is

$$|\mathbb{G}/\mathbb{H}_w|=\mathrm{ord}(\chi_w)=\mathrm{lcm}\left(\frac{m_1}{g_{1,w_1}},\frac{m_2}{g_{2,w_2}},\ldots,\frac{m_l}{g_{l,w_l}}\right),$$

where ord is the order, lcm is the least common multiple, and $g_{j,w_j}=\gcd(m_j,w_j)$, because $\mathrm{ord}(\chi_x)$ is the least positive integer t such that $\chi_w^t=1$ (which indicates that tw_j must be divisible by m_j for any j). Here, $|\mathbb{H}_w|$ is maximized at frequency w when $\mathbb{Q}_w=\mathbb{G}/\mathbb{H}_w$ is minimized. Define $g_{j,0}=m_j$ for any j, because $m_j/g_{j,w_j}$ represents the order of character. Excluding the trivial case that $m_j/g_{j,w_j}=1$ for all j (*i.e.*, $w=0$), it holds that

$$\min_{w\neq0}|\mathbb{G}/\mathbb{H}_w|=\min_{w\neq0}\mathrm{lcm}\left(\frac{m_1}{g_{1,w_1}},\frac{m_2}{g_{2,w_2}},\ldots,\frac{m_l}{g_{l,w_l}}\right)=\min_{1\le j\le l}\min_{1\le w_j\le m_j}\frac{m_j}{g_{j,w_j}}.$$

Such w minimizing $|\mathbb{H}_w|$ is given by, for example, $w_j=0$ for all j except for one invariant factor. Thus, according to Lagrange's theorem, the minimum noise requirement is $\delta/2<1-\max_w(|\mathbb{H}_w|\,/\,|\mathbb{G}|)=1-\max_j\max_{w_j}(g_{j,w_j}/m_j)$.

The MI condition can be also proven as in the case of arithmetic masking with the definition of $Y_i^{(w)}$ using the quotient group $\mathbb{Q}_w$, one-hot distribution (Lemma 3), and frequency w that maximizes $|\mathbb{H}_w|$. □

K Proof of Lemma 3

Proof. We first prove the case of $l=1$ (*i.e.*, arithmetic masking) for an easy explanation, which is followed by general cases.

Owing to the definition of DFT, it holds

$$|\hat{p}(w)|^2=\left|\sum_{h=0}^{h-1}p(h)\mathrm{e}^{-\mathrm{i}h\Omega_w}\right|^2\le\left(\sum_{h=0}^{h-1}|p(h)|\right)^2=1, \tag{19}$$

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 for $l = 1$.

General case. The twiddle factor is replaced with $\exp(-\mathrm{i}\sum_j h_j \Omega_{w_j}^{(j)}/m_j)$ and h takes values of $\mathbb{G} = \mathbb{Z}_{m_1} \times \mathbb{Z}_{m_2} \times \cdots \times \mathbb{Z}_{m_l}$. Obviously, the inequality of (19) also holds for this case. If $\mathbb{G}$ is isormorphic to $\mathbb{Z}_m$ with prime m, all coefficients of $\sum_h p(h) \exp(-\mathrm{i}\sum_j h_j \Omega_{w_j}^{(j)}/m_j) = \sum_h p(h)\mathrm{e}^{-\mathrm{i}h\Omega_w}$ have different phase when $w = 0$. For the equality condition of triangle inequality, all non-zero values should have an identical phase. Thus, $p(h)$ can have non-zero value only at specific h, which indicates that p is a one-hot distribution. □

L Proof of Lemma 4

Proof. We first prove the limit of general 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

$$\begin{aligned} I(Z; \boldsymbol{X}) &\leq \log\left(1 + \sum_{w\neq 0}\prod_{i=0}^{d} \mathbb{E}_{\boldsymbol{L}}\left[\left|\hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L})\right|^2\right]\right) \\ &\to \log\left(1 + \sum_{w\in\mathcal{W}_{\mathrm{peak}}} 1\right) = \log\left(1 + \left|\mathcal{W}_{\mathrm{peak}}\right|\right) \quad \text{as} \quad d \to \infty. \end{aligned} \tag{20}$$

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 for w satisfying $\hat{p}_{S_i|\boldsymbol{L}_i}(w \mid \boldsymbol{L}) = 1$. Consider the former case. 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 $I(Z; \boldsymbol{X}') \leq I(Z; \boldsymbol{X})$. 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

$$I(Z; \boldsymbol{X}') = \log\left(1 + \sum_{w\neq 0}\prod_{i=0}^{d} \mathbb{E}_{Q}\left[\left|\hat{p}_{Y_i^{(w)}|Q_i}(w \mid Q)\right|^2\right]\right) \to \log\left(1 + \left|\mathcal{W}_{\mathrm{peak}}\right|\right), \tag{21}$$

as $d \to \infty$ with equality. By combining Equations (20) and (21) with the above data processing inequality, we have $I(Z; \boldsymbol{X}) \to \log(1 + |\mathcal{W}_{\text{peak}}|)$.

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 (15), 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 ($0 \leq s_0 \leq \gamma_w$) 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 $1 + |\mathcal{W}_{\text{peak}}| = \max_{w \in \mathcal{W}_{\text{peak}}} \gamma_w$. This completes the proof. □

M Proof of Corollary 1

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 (9) and (10). 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 $\mathcal{W}_{\text{peak}} = \{2^l - 1\}$ or $\{m/\gamma\}$ under this leakage, the proof is completed. □

N Proof of Theorem 8

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, and this constant is multiplied with the factor of the above exponential decay. This indicates 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 $I(Z; \boldsymbol{X}) \to \log\left(1 + |\mathcal{W}_{\text{peak}}|\right)$ 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. □

O Proof of Corollary 2

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

P Proof of Lemma 5

Proof. Let $H : \mathcal{K} \times \mathcal{X} \to \{0,1\}^b$ be an α-AXU hash with b-bit output. Let $G : \mathcal{K}_G \times \mathcal{X} \to \{0,1\}^b$ for $\mathcal{K}_G = \mathcal{K} \times \{0,1\}^b$ 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

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

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 $1/2^b$ for the uniformity of K_2. Hence, $\Pr[G(K_1,K_2),x) = y, G((K_1,K_2),x') = y'] \le \alpha/2^b$, showing G is $\alpha/2^b$-ASU. □

Q Proof of Lemma 6

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

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

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

R Black-box security of gXEX and AOCB

R.1 Notations

For non-negative integers i and j with $i \le j$, let $[i..j] \stackrel{\triangle}{=} \{i, i+1, i+2, \ldots, j\}$, and let $(j)_i \stackrel{\triangle}{=} j(j-1)(j-2) \cdots (j-i+1)$. If X is uniformly chosen from the set $\mathcal{X}$, we write $X \stackrel{\$}{\leftarrow} \mathcal{X}$. Suppose $\mathcal{A}$ is an adversary in a game. $\mathcal{A}$ can query to s

oracles, $\mathcal{O}_1, \ldots, \mathcal{O}_s$, in any order. In the case of a distinguishing game, $\mathcal{A}$ outputs a bit after the queries, which is a random variable whose probability space is defined under $\mathcal{A}$ and $\{\mathcal{O}_i\}_{i\in[1..s]}$. Let $[\mathcal{A}^{\mathcal{O}_1,\ldots,\mathcal{O}_s} = 1]$ denote the event that this bit is 1. By writing $\mathbf{Adv}^{\mathtt{xxx}}_{\Pi}(\mathcal{A})$, we mean *advantage* of the adversary $\mathcal{A}$ in the game $\mathtt{xxx}$ involving Π.

R.2 (Tweakable) Block cipher and random primitives

Any keyed function $F : \mathcal{K} \times \mathcal{X} \to \mathcal{Y}$ with key space $\mathcal{K}$, we write $F_K(x)$ to denote $F(K, x)$. Unless otherwise specified, we assume $K \xleftarrow{\$} \mathcal{K}$ in evaluation of $F_K(x)$. A block cipher $E : \mathcal{K} \times \mathcal{M} \to \mathcal{M}$ is a keyed function such that for any $K \in \mathcal{K}$, $E(K, \cdot)$ is a permutation over the message space $\mathcal{M}$. We write $E_K^{-1}(\cdot)$ to denote the decryption function. A tweakable block cipher (TBC) [53] $\widetilde{E} : \mathcal{K} \times \mathcal{T}_w \times \mathcal{M} \to \mathcal{M}$ is a keyed function such that for any $K \in \mathcal{K}$ and *tweak* $W \in \mathcal{T}_w$, $\widetilde{E}(K, W, \cdot)$ is a permutation over $\mathcal{M}$. Tweak is typically used as a public/chosen value in contrast to the secret key. Let $\mathrm{Perm}(b)$ be the set of all permutations of b bits. A uniform random permutation (URP) $\mathsf{P} : \{0,1\}^b \to \{0,1\}^b$ is a random permutation uniformly chosen over $\mathrm{Perm}(n)$. A tweakable uniform random permutation (TURP) $\tilde{\mathsf{P}} : \mathcal{T}_w \times \mathcal{M} \to \mathcal{M}$ consists of $|\mathcal{T}_w|$ independent instances of P over $\mathcal{M}$, and the first argument (tweak) specifies the URP used. The security of TBC $\widetilde{E}$ is defined as the following distinguishing probability.

$$\mathbf{Adv}^{\mathtt{TPRP}}_{\widetilde{E}}(\mathcal{A}) \triangleq |\Pr[\mathcal{A}^{\widetilde{E}_K} = 1] - \Pr[\mathcal{A}^{\tilde{\mathsf{P}}} = 1]|,$$
$$\mathbf{Adv}^{\mathtt{TSPRP}}_{\widetilde{E}}(\mathcal{B}) \triangleq |\Pr[\mathcal{A}^{\widetilde{E}_K, \widetilde{E}_K^{-1}} = 1] - \Pr[\mathcal{A}^{\tilde{\mathsf{P}}, \tilde{\mathsf{P}}^{-1}} = 1]|,$$

where $\tilde{\mathsf{P}}$ has the same domain and range as $\widetilde{E}_K$ and $\tilde{\mathsf{P}}^{-1}$ is its inverse. When $\mathcal{T}_w$ is a singleton, the security of block cipher E_K is obtained by the above security, denoted by $\mathbf{Adv}^{\mathtt{PRP}}_{E}(\mathcal{A})$ and $\mathbf{Adv}^{\mathtt{SPRP}}_{E}(\mathcal{B})$.

R.3 Authenticated encryption

A nonce-based AE scheme $\Pi = (\Pi.\mathcal{E}, \Pi.\mathcal{D})$ is defined over a key space $\mathcal{K}$, a nonce space $\mathcal{N}$, an associated data (AD) space $\mathcal{AD}$, a message space $\mathcal{M}$, and a tag space $\mathcal{T}$. An AD is information that is not to be kept confidential, but its integrity must be ensured. For example, network encryption will treat the protocol header as an AD. In this paper, we assume that the tag length is b bits for simplicity. Formally, AE is a tuple of an encryption function $\Pi.\mathcal{E} \colon \mathcal{K} \times \mathcal{N} \times \mathcal{AD} \times \mathcal{M} \to \mathcal{M} \times \mathcal{T}$, and a decryption function $\Pi.\mathcal{D} : \mathcal{K} \times \mathcal{N} \times \mathcal{AD} \times \mathcal{M} \times \mathcal{T} \to \mathcal{M} \cup \{\perp\}$, where symbol $\perp$ indicates a verification failure. We require $\Pi.\mathcal{D}_K(N, A, C, T) = M$ if $\Pi.\mathcal{E}_K(N, A, M) = (C, T)$ for soundness.

Let $\Pi = (\Pi.\mathcal{E}, \Pi.\mathcal{D})$ be an AE scheme. Let \$ be the random-bit oracle that takes (N, A, M) and returns random $|M| + b$ bits. The standard security notions of Π are Priv (confidentiality/privacy) and Auth (authenticity/integrity), defined below.

Definition 14. *Let $\mathcal{A}$ be a* Priv *adversary against* Π *and let $\mathcal{B}$ be an* Auth *adversary against* Π.

$$\mathbf{Adv}^{\mathtt{Priv}}_{\Pi}(\mathcal{A}) \triangleq |\Pr[\mathcal{A}^{\Pi.\mathcal{E}_K} = 1] - \Pr[\mathcal{A}^{\$} = 1]|,$$
$$\mathbf{Adv}^{\mathtt{Auth}}_{\Pi}(\mathcal{B}) \triangleq \Pr[\mathcal{B}^{\Pi.\mathcal{E}_K, \Pi.\mathcal{D}_K} \text{ forges}],$$

where $\mathcal{A}$ and $\mathcal{B}$ are nonce-respecting; i.e., they do not repeat the same nonce in their encryption queries. The event $[\mathcal{B}^{\Pi.\mathcal{E}_K, \Pi.\mathcal{D}_K} \text{ forges}]$ *means that $\mathcal{B}$ receives some $M' \neq \perp$ from $\Pi.\mathcal{D}_K$. $\mathcal{B}$ does not forward queries from $\Pi.\mathcal{E}_K$ to $\Pi.\mathcal{D}_K$, namely, $\mathcal{B}$ does not query (N, A, C, T) to $\Pi.\mathcal{D}_K$ after obtaining (N, A, M, C, T) as a result of query-response to $\Pi.\mathcal{E}_K$.*

R.4 H-Coefficient technique

We assume that an adversary $\mathcal{A}$ queries the two worlds, real and ideal, denoted by $\mathcal{O}_{\mathsf{re}}$ and $\mathcal{O}_{\mathsf{id}}$, and tries to distinguish them. The H-coefficient [26, 62] is a general technique to evaluate the distinguishing probability of $\mathcal{A}$. We define a *transcript* as a set of input/output values that $\mathcal{A}$ obtains during the interaction with the world. Let T_{re} (*resp.* T_{id}) denote the probability distribution of the transcript induced by the real world (*resp.* the ideal world). By extension, we also use the same notation to refer to a random variable distributed according to each distribution. We say that a transcript θ is *attainable* if $\Pr[\mathsf{T}_{\mathsf{id}} = \theta] > 0$ holds with respect to $\mathcal{A}$. Let Θ denote the set of attainable transcripts. The following is the fundamental lemma of the H-coefficient technique. See *e.g.* [26] for the proof.

Lemma 7. *Let $\Theta = \Theta_{\text{good}} \sqcup \Theta_{\text{bad}}$ be a partition of the set of attainable transcripts. Assume that there exists $\varepsilon_1 \geq 0$ such that for any $\theta \in \Theta_{\text{good}}$, one has*

$$\frac{\Pr[\mathsf{T}_{\mathsf{re}} = \theta]}{\Pr[\mathsf{T}_{\mathsf{id}} = \theta]} \geq 1 - \varepsilon_1,$$

and that there exists $\varepsilon_2 \geq 0$ such that $\Pr[\mathsf{T}_{\mathsf{id}} \in \Theta_{\text{bad}}] \leq \varepsilon_2$. Then $|\Pr[\mathcal{A}^{\mathcal{O}_{\mathsf{re}}} = 1] - \Pr[\mathcal{A}^{\mathcal{O}_{\mathsf{id}}} = 1]| \leq \varepsilon_1 + \varepsilon_2$.

R.5 Security bound of gXEX and its proof

Theorem 10. *Let $\mathcal{A}$ be a TSPRP adversary with time complexity t. We have*

$$\mathbf{Adv}^{\mathtt{TSPRP}}_{\mathsf{gXEX}[E_K, \mathbb{G}]}(\mathcal{A}) \leq \mathbf{Adv}^{\mathtt{SPRP}}_{E}(\mathcal{A}') + \alpha q^2,$$

where q is the number of queries of $\mathcal{A}$ in encryption and decryption query, and $\mathcal{A}'$ is an SPRP adversary performing q queries with time complexity $t + O(q)$. From the above inequation, we also have $\mathbf{Adv}^{\mathtt{TPRP}}_{\mathsf{gXEX}[E_K, \mathbb{G}]}(\mathcal{B}) \leq \mathbf{Adv}^{\mathtt{PRP}}_{E}(\mathcal{B}') + \alpha q_e^2$ for TPRP adversary $\mathcal{B}$ with time complexity t, where q_e is the number of $\mathcal{B}$'s queries, and $\mathcal{B}'$ is a PRP adversary performing q_e queries with time complexity $t + O(q_e)$.

Proof. For the evaluations of $\mathbf{Adv}^{\mathtt{TSPRP}}_{\mathsf{gXEX}[E_K,\mathbb{G}]}(\mathcal{A})$ and $\mathbf{Adv}^{\mathtt{TPRP}}_{\mathsf{gXEX}[E_K,\mathbb{G}]}(\mathcal{B})$, we first replace E_K with P, which yields the terms $\mathbf{Adv}^{\mathtt{SPRP}}_{E}(\mathcal{A}')$ and $\mathbf{Adv}^{\mathtt{PRP}}_{E}(\mathcal{B}')$, respectively. All that remains is the evaluation of $\mathbf{Adv}^{\mathtt{TSPRP}}_{\mathsf{gXEX}[\mathsf{P},\mathbb{G}]}(\mathcal{A})$ because $\mathbf{Adv}^{\mathtt{TPRP}}_{\mathsf{gXEX}[\mathsf{P},\mathbb{G}]}(\mathcal{B}) \leq \mathbf{Adv}^{\mathtt{TSPRP}}_{\mathsf{gXEX}[\mathsf{P},\mathbb{G}]}(\mathcal{A})$. For $i \in [1..q]$, let (T_i, M_i, C_i) be an input-output tuple of $\mathcal{A}$'s i-th query. For the sake of ease, we assume $\mathcal{A}$ obtains the secret key K' after all queries but before it determines an output bit. In the real world, $\mathcal{A}$ obtains the real key K', and in the ideal world, it obtains a dummy key $K' \xleftarrow{\$} \{0,1\}^b$. We define a transcript of $\mathcal{A}$ as $\theta \stackrel{\triangle}{=} \{\{(T_i, M_i, C_i)\}_{i\in[1..q]}, K'\}$, and define bad events as follows.

- Bad1: There exists i, $j \in [1..q]$, $i \neq j$, *s.t.* $M_i \boxplus f_{K'}(T_i) = M_j \boxplus f_{K'}(T_j)$.
- Bad2: There exists i, $j \in [1..q]$, $i \neq j$, *s.t.* $C_i \boxplus f_{K'}(T_i) = C_j \boxplus f_{K'}(T_j)$.

We obtain $\Pr[\mathsf{T}_{\mathsf{id}} \in \Theta_{\mathrm{bad}}] = \Pr[\mathrm{Bad1} \cup \mathrm{Bad2}] \leq 2\binom{q}{2} \Pr[f_{K'}(T_i) \boxminus f_{K'}(T_j) = M_j \boxminus M_i] \leq \alpha q^2$ since $f_{K'}$ is α-AΔU($\mathbb{G}$).

Let $\mathsf{T}^{IO}_{\mathsf{id}}$ and $\mathsf{T}^{K'}_{\mathsf{id}}$ (*resp.* $\mathsf{T}^{IO}_{\mathsf{re}}$ and $\mathsf{T}^{K'}_{\mathsf{re}}$) denote the random variables of $\{(T_i, M_i, C_i)\}_{i\in[1..q]}$ and K' in the ideal (*resp.* real) world. For a good transcript $\theta \in \Theta_{\mathrm{good}}$, we obtain

$$\Pr[\mathsf{T}_{\mathsf{re}} = \theta] = \Pr[\mathsf{T}^{IO}_{\mathsf{re}} = \{(T_i, M_i, C_i)\}_{i\in[1..q]} \mid \mathsf{T}^{K'}_{\mathsf{re}} = K'] \Pr[\mathsf{T}^{K'}_{\mathsf{re}} = K'] = \frac{1}{(2^b)_q} \cdot \frac{1}{2^b} \tag{22}$$

since θ is good, *i.e.*, there is no collision on inputs/outputs of P in θ. For the evaluation $\Pr[\mathsf{T}_{\mathsf{id}} = \theta]$, let t be the number of distinct tweaks in θ, and for $i \in [1..t]$, let q_i be the number of queries using the i-th tweak *i.e.*, $\sum_{i=i}^{t} q_i = q$. We obtain

$$\Pr[\mathsf{T}_{\mathsf{id}} = \theta] = \prod_{i=1}^{t} \frac{1}{(2^b)_{q_i}} \cdot \frac{1}{2^b}. \tag{23}$$

From (22) and (23), we obtain $\Pr[\mathsf{T}_{\mathsf{re}} = \theta] / \Pr[\mathsf{T}_{\mathsf{id}} = \theta] \geq 1$, and thus $\mathbf{Adv}^{\mathtt{TSPRP}}_{\mathsf{gXEX}[\mathsf{P},\mathbb{G}]}(\mathcal{A}) \leq \alpha q^2$. This concludes the proof.

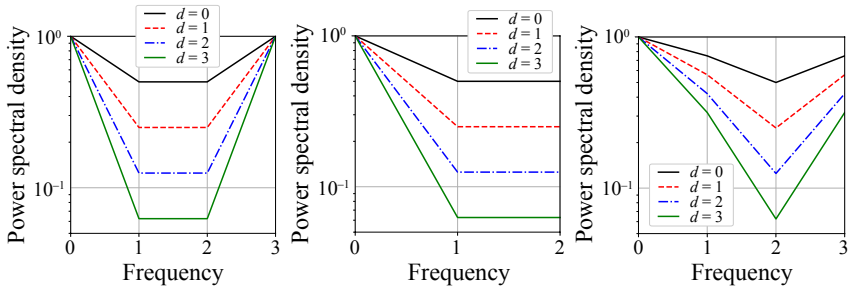
R.6 Security bounds of AOCB

Theorem 11. *Let $\mathcal{A}$ (resp. $\mathcal{B}$) be a Priv (resp. Auth) adversary with time complexity t. We have*

$$\mathbf{Adv}^{\mathtt{Priv}}_{\mathsf{AOCB}_{\widetilde{E}_K}}(\mathcal{A}) \leq \mathbf{Adv}^{\mathtt{TPRP}}_{\widetilde{E}_K}(\mathcal{A}'),$$

$$\mathbf{Adv}^{\mathtt{Auth}}_{\mathsf{AOCB}_{\widetilde{E}_K}}(\mathcal{B}) \leq \mathbf{Adv}^{\mathtt{TSPRP}}_{\widetilde{E}_K}(\mathcal{B}') + \frac{2q_d}{2^b},$$

where q_d is the number of decryption queries of $\mathcal{B}$. The adversary $\mathcal{A}'$ is a PRP adversary performing σ_e queries with time complexity $t + O(\sigma_e)$, where σ_e is the total number of queried blocks made by $\mathcal{A}$, and $\mathcal{B}'$ is an SPRP adversary performing σ queries with time complexity $t + O(\sigma)$, where σ is the total number of queried blocks made by $\mathcal{B}$.



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

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

The proof is almost identical to that of OCB, and thus, we provide only an overview. We adopt the following hybrid argument.

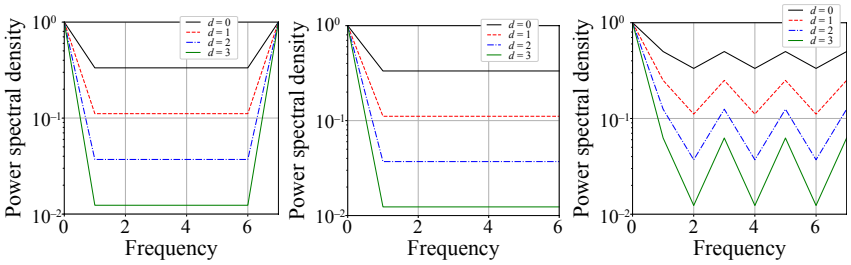
$$\mathbf{Adv}^{\mathtt{Priv}}_{\mathsf{AOCB}_{\widetilde{E}_K}}(\mathcal{A}) \le \mathbf{Adv}^{\mathtt{TPRP}}_{\widetilde{E}_K}(\mathcal{A}') + \mathbf{Adv}^{\mathtt{Priv}}_{\mathsf{AOCB}_{\widetilde{\mathsf{P}}}}(\mathcal{A}),$$
$$\mathbf{Adv}^{\mathtt{Auth}}_{\mathsf{AOCB}_{\widetilde{E}_K}}(\mathcal{B}) \le \mathbf{Adv}^{\mathtt{TSPRP}}_{\widetilde{E}_K}(\mathcal{B}') + \mathbf{Adv}^{\mathtt{Auth}}_{\mathsf{AOCB}_{\widetilde{\mathsf{P}}}}(\mathcal{B}),$$

where $\mathcal{A}'$ and $\mathcal{B}'$ are TPRP and TSPRP adversaries against $\widetilde{E}_K$ invoked by $\mathcal{A}$ and $\mathcal{B}$, respectively. The adversary $\mathcal{A}'$ performs σ_e encryption queries, and $\mathcal{B}'$ performs σ queries in total. In the same manner as the proof of ΘCB, we obtain $\mathbf{Adv}^{\mathtt{Priv}}_{\mathsf{AOCB}_{\widetilde{\mathsf{P}}}}(\mathcal{A}) = 0$ and $\mathbf{Adv}^{\mathtt{Auth}}_{\mathsf{AOCB}_{\widetilde{\mathsf{P}}}}(\mathcal{B}) \le q_d/(2^b - 1) \le 2q_d/2^b$.

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

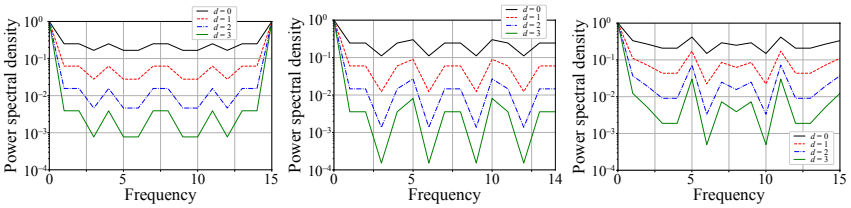
We simulated the power spectra of $p_{Z|\boldsymbol{X}}$ with different d, assuming that the leakage was given by HW without physical noise. Figures 6 to 12 illustrate the power spectra of $p_{Z|\boldsymbol{X}}$ for different bit length l and masking order d, where figures (a), (b), (c) are maskings 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.



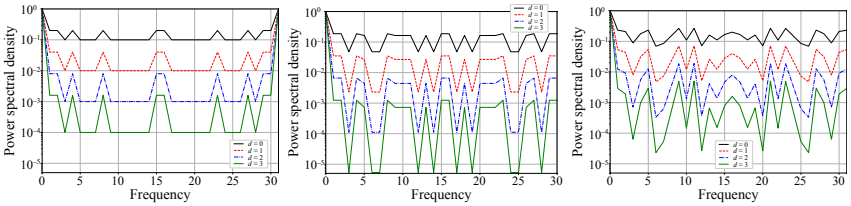
(a) Boolean (b) $m = 7$ (c) $m = 8$

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



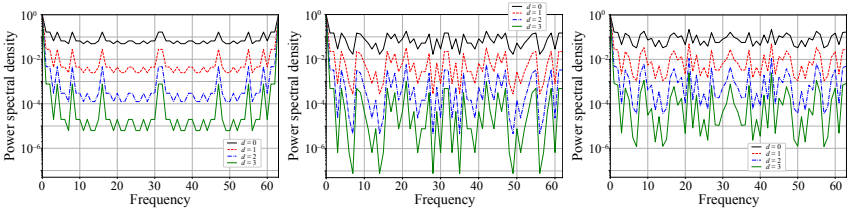
(a) Boolean (b) $m = 15$ (c) $m = 16$

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



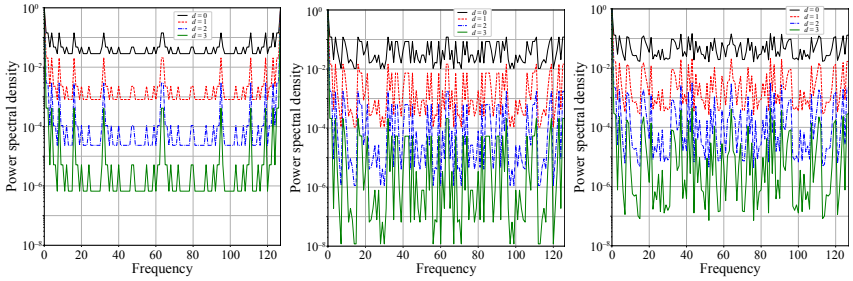
(a) Boolean (b) $m = 31$ (c) $m = 32$

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



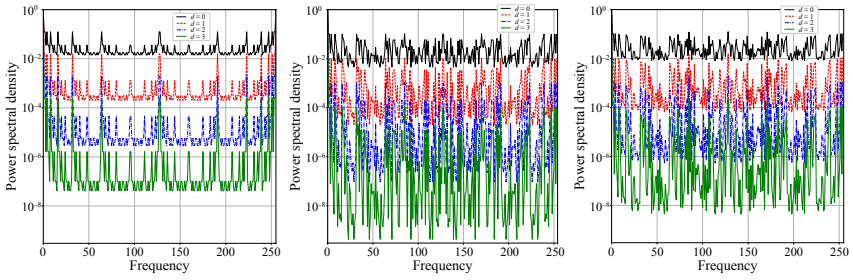
(a) Boolean (b) $m = 63$ (c) $m = 64$

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 $l = 8$.