An ANOVA Method to Rapidly Assess Information Leakage Near Cryptographic Modules

Vishnuvardhan V. Iyer, *Student Member, IEEE*, and Ali E. Yılmaz, *Senior Member, IEEE*

***Abstract*—A measurement method based on the analysis of variance (ANOVA) F-statistic is presented to rapidly evaluate cryptographic modules' vulnerability to fine-grained EM side-channel analysis (SCA) attacks. The proposed method assumes that evaluators can control the device under test to set carefully chosen inputs to computations of interest and to repeat measurements as many times as needed. It identifies optimal measurement configurations—that minimize the marginal cost for repeated attacks to extract the data of interest—in three stages. In the first two stages, the variances in observed fields are analyzed using specially designed test cases and low F-value measurement configurations susceptible to noise are eliminated. In the third stage, the data of interest are extracted via a correlation-analysis attack using the remaining, high F-value, configurations. The method is used to evaluate 9 Advanced Encryption Standard (AES) implementations, 7 of which were hardened against EM SCA attacks. The test cases for the first two stages are constructed by generating extreme AES encryption keys and input plaintexts. The least/most effective countermeasures are found to increase the marginal cost of EM SCA attacks by ~1.1×/>30×; the proposed method could evaluate the vulnerabilities of hardened AES modules using ~1.5-37× fewer measurements than alternatives.**

***Index Terms*—Analysis of variance, electromagnetic measurements, measurement techniques, measurement uncertainty, side-channel attacks, cryptography.**

I. Introduction

Electromagnetic side-channel analysis (EM SCA) attacks exploit unintentionally emanated fields to break the security of computing systems [1]-[11]. These non-invasive attacks are particularly potent when used to disclose encryption keys of cryptographic modules [5]-[20]. For example, using the measurement setup in Fig. 1, the authors could extract from the probed fields near an FGPA the key it uses to encrypt data with the advanced encryption standard (AES) [10], [18], [20], thus exposing all ciphertexts secured with that key. Numerous such attacks that exploit chip emanations to break cryptography [5]-[11] and countermeasures that increase resilience against such attacks [12]-[20] have been developed. These EM SCA attacks deduce critical data by correlating observed fields—sums of signals from exploitable sources, noise from other system processes, and measurement noise—to on-chip computations, while countermeasures against them degrade the correlation of observed fields and computations of interest.

The vulnerability of a cryptographic module to EM SCA attacks can be evaluated empirically by performing a correlation-analysis attack on its baseline or hardened implementations and observing the cost/number of measurements needed to disclose the data of interest [9], [10], [12], [16], [20], i.e., by emulating actual EM SCA attacks. Such correlation-analysis attacks can be categorized as:

- *Coarse-grained EM SCA attacks* [8], [9], [12], [16] use relatively large probes (comparable to chip size) that aggregate fields from a multitude of on-/off-chip sources, including those uncorrelated to the computations of interest [6]. As a result, they typically require many measurements to establish sufficient correlation and recover data. Furthermore, these are memoryless attacks: previous attacks do not impact future evaluations.
- *Fine-grained EM SCA attacks* use relatively small probes (smaller than chip size) to scan for and isolate vulnerable regions [5]-[7], [9]-[11]. These attacks first search for optimal configurations, e.g., locations and orientations, of probes that are most sensitive to target signals/least sensitive to noise; they then use these configurations to perform correlation analysis and recover data.

While coarse-grained EM SCA attacks are simpler to implement, fine-grained EM SCA attacks can require far fewer measurements when used with optimal probe configurations, making them more potent than the conventional power/coarse-grained EM SCA attack methods [9],[11],[38],[39]; moreover, once identified, these configurations can be reused to minimize the cost of future attacks on similar chips. Fine-grained EM SCA attacks' initial search for optimal probe configurations, however, can be rather costly [10] because of the large number of probe configurations that must be evaluated.

Emulating actual correlation-based attacks to empirically evaluate side-channel security of a cryptographic module is often infeasible against fine-grained EM SCA attacks, especially for modules hardened with countermeasures. This is because empirical verification requires security evaluators to test many possible probe configurations, including ineffective ones, to ensure that they do not miss any effective

Manuscript received November 29, 2021; revised January 11,2022 and February 9,2022. This work was supported in part by NSF Grant CCF-1901446. (*Corresponding author: V. Iyer*)

The authors are with the Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712 USA (e-mail: vishnuv.iyer@utexas.edu).

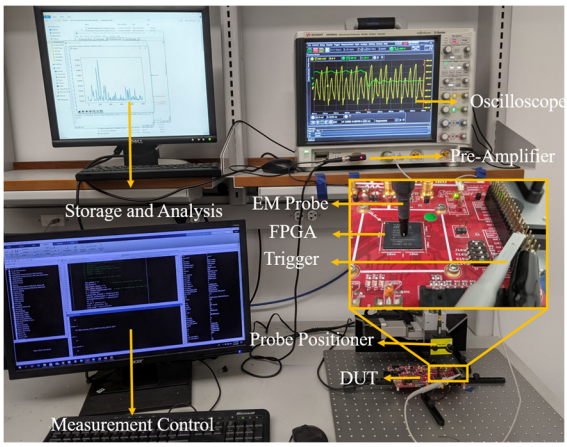


Fig. 1. Near-field measurement setup to perform fine-grained EM SCA attacks on an FPGA running the AES algorithm [6].

configurations. In contrast, actual attackers may be able to find effective probe configurations (by chance) within a few configurations they test. Thus, there is an inherent asymmetry between evaluators, who must ensure the module is sufficiently secure against all probe configurations, and actual attackers, who must ensure it is sufficiently vulnerable to only one probe configuration. The asymmetry is amplified when evaluators and actual attackers are subject to different constraints; in particular, on their ability to observe or control the module's inputs, outputs, or keys. These constraints are formalized in threat models: Actual attackers are often restricted to a "black-box threat model" [9], where the module's output and EM fields can be observed for a potentially unlimited number of encryptions but its input or key cannot be accessed. In contrast, security evaluators may also be granted partial/full control over the input (a "gray-/white-box threat model" [9], [11], [19]) and the key [21] (a "gold-box threat model" [9]). Thus, evaluators may observe the output and EM fields for specific encryptions with specially designed inputs or keys [21]. When evaluators face fewer restrictions, they can accelerate the security evaluation by implementing targeted tests and obtaining statistical indicators of information leakage, e.g., via test vector leakage assessment (TVLA) [22], [23] or analysis of variance (ANOVA) [4], [6], [19], [24], prior to performing correlation-analysis attacks. This paper presents a novel method for empirical evaluation of cryptographic modules' vulnerability to fine-grained EM SCA attacks, including for modules hardened with countermeasures.

In a preliminary study [6], the authors proposed a method using the ANOVA F-statistic to eliminate configurations most impacted by measurement noise, as a precursor to fine-grained EM correlation-analysis attacks. This paper expands the work in [6], which did not address obfuscation due to uncorrelated system processes and used only time-domain fields. It presents an ANOVA method to accelerate security evaluations in the presence of both measurement noise and algorithmic noise (fields generated by other system processes), using time- or frequency-domain fields. Unlike [6], where evaluators were assumed to have partial control over the input, here they are assumed to have full control over the input *and* the encryption key of the device under test (DUT), which corresponds to a less restrictive (gold-box instead of gray-box) threat model. Thus, evaluators can generate extreme variations in target signals and rapidly obtain statistical indicators using a small set of targeted tests. Once these indicators are obtained, correlation analysis is performed as a confirmatory step to validate the presence of information leakage predicted by the statistical metrics [44]. In this paper, EM side-channel security is evaluated in three stages: Stage I eliminates probe configurations most affected by measurement noise using an ANOVA indicator. Stage II eliminates from the remaining configurations those most affected by algorithmic noise using a second ANOVA indicator. Stage III emulates a correlation-analysis attack only with the remaining configurations. Therefore, Stages I and II condense the set of potential optimal configurations with a series of low-cost scans, and Stage III performs expensive correlation analyses only within this condensed set and actually extracts the data of interest: the AES key. Specifically, in Stages I and II, targeted tests are constructed systematically according to the leakage model used in Stage III. The proposed methodology is used to evaluate the EM side-channel security of AES implementations with three types of countermeasures:

- Repeatability countermeasures, e.g., random scaling of supply voltage and/or clock frequency [20];
- Algorithmic countermeasures, e.g., masking or byte order randomization [19], [26], [27]; and
- Physical design strategies, e.g., shielding [17] or changing power-grid layout [18].

The rest of this paper is organized as follows. Section II presents background on EM SCA attacks on AES, followed by an overview of existing methodologies, and the role of noise in such attacks. Section III describes the proposed method as applied to AES. Section IV details the evaluated baseline and hardened AES implementations. Section V presents the measurement setup and results for baseline AES implementations. Section VI demonstrates the proposed method's suitability for evaluating the effectiveness of countermeasures detailed in Section IV. Section VII concludes the paper.

II. EM Side-Channel Analysis Attacks on AES

This section presents an overview of the vulnerability of AES to SCA attacks, how correlation-analysis attacks exploit it, the exhaustive correlation analysis attacks vs. alternatives, the impact of noise on these attacks, and the ANOVA method to quantify this impact. All probed fields and their correlation analysis shown in this section were obtained from an attack on the first key byte of AES-128, using the Artix-7 FPGA, operated at 20 MHz clock and 1 V supply voltage, and the optimal measurement configuration in [10], [20]: a 1-mm diameter H-field probe, oriented in the x direction, and located at (9.7, 8, 0.5) mm from the bottom left corner of the chip. This attack is detailed in Section V.A.

A. The AES Algorithm and Its SCA Vulnerability

AES, a commonly adopted standard for processor and wireless security, specifies a symmetric-key algorithm [25] that uses the same key for encryption and decryption. It is a block cipher that groups inputs into fixed 16-byte blocks and can use keys of size 128, 192, or 256 bits; the 128-bit implementation

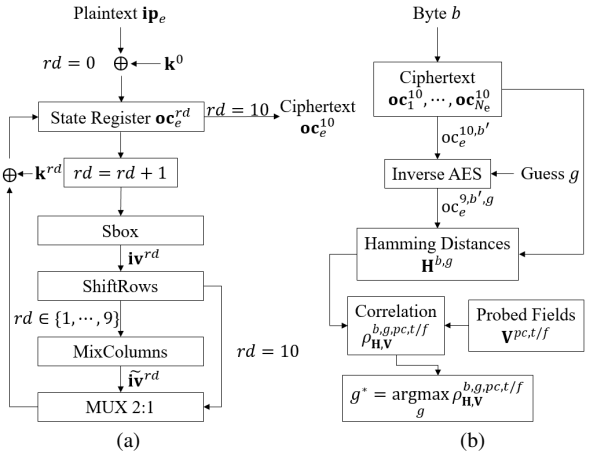


Fig. 2. Flowcharts of (a) AES-128 [25] and (b) correlation analysis [42].

is used in this paper (Fig. 2(a)). Each encryption e by AES-128 requires 10 rounds of operations to transform the 16-byte input plaintext $\mathbf{ip}_e$ to the output ciphertext $\mathbf{oc}_e^{10}$ using the key $\mathbf{k}^0$ (Fig. 2(a)). In each round $rd \in \{1, \cdots, 10\}$, a round key $\mathbf{k}^{rd}$ (generated from the key $\mathbf{k}^0$ via a key-expansion algorithm [25]) is used to update the 16-byte output to $\mathbf{oc}_e^{rd} = [oc_e^{rd,0}, \cdots, oc_e^{rd,15}]$. All AES operations are performed byte wise: In each round rd, first, each byte $b' \in \{0, \cdots, 15\}$ of the previous round's output $oc_e^{rd-1,b'}$ is replaced by an intermediate value $iv_e^{rd,b'}$ using a substitution box (*Sbox*). The *Sbox* transform replaces a byte's value using a one-to-one non-linear map defined by Rijndael's finite field [25]. Then, the byte order of $iv_e^{rd,b'}$ is shuffled using the *ShiftRows* and *MixColumns* transforms to generate $\tilde{iv}_e^{rd,b}$, where $b \in \{0, \cdots, 15\}$ is the new position of the byte in the updated 16-byte array. Finally, the intermediate value is *XOR*ed with the key byte $k^{rd,b}$ to generate the output byte $oc_e^{rd,b}$. The *MixColumns* operation is skipped in the last round; thus, the last round of AES can be represented as

$$oc_e^{10,b} = ShiftRows\left(Sbox\left(oc_e^{9,b'}\right)\right) \oplus k^{10,b} \quad (1)$$

If attackers have access to the output and if they know/correctly guess the 10th round key $\mathbf{k}^{10}$—the data of interest for SCA attacks on AES—they can invert (1) as

$$oc_e^{9,b'} = Sbox^{-1}\left(ShiftRows^{-1}\left(k^{10,b} \oplus oc_e^{10,b}\right)\right). \quad (2)$$

B. Correlation Analysis Attacks

The fields emanated in the final round of AES depend on the key, which causes an EM side-channel vulnerability [5], [6], [10], [19]. EM SCA attacks on AES use hypothetical leakage models [28] to correlate observed fields to the computations/ processes during the final round of AES. These models abstract the sources of emanations in the DUT, such as transistor switching, currents on clock and power traces, EM coupling, etc., using simplified quantities. This paper employs a byte-wise SCA attack (Fig. 2(b)), which adopts a Hamming distance (HD) leakage model [6]. The attack correlates the observed fields with the HD between $oc_e^{9,b'}$ and $oc_e^{10,b'}$ to disclose $k^{10,b}$. Byte-wise analysis significantly reduces the complexity of key search [39]. In this attack, the attackers observe N_e encryptions and for each encryption $e \in \{1, \cdots, N_e\}$, they use the observed

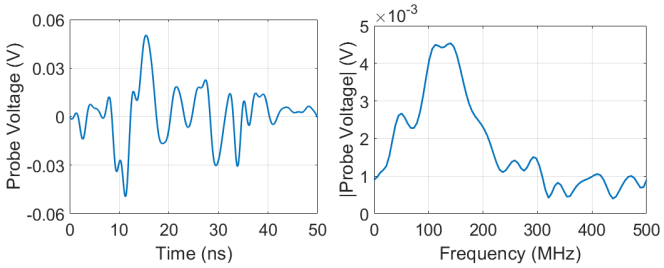


Fig. 3. Time-domain (left) and frequency-domain (right) fields observed during the last round of AES when the key and the input plaintext are set to $\mathbf{k}^0 = \mathbf{k}_1$ and $\mathbf{ip}_1$. The fields were captured using the optimal probe configuration $pc^{0,\text{opt}}$ identified in [10] to disclose byte 0 of the 10th round key ($k^{10,0}$).

$\mathbf{oc}_e^{10}$ together with every possible guess $g \in \{0, \cdots, 255\}$ for the key byte $k^{10,b}$ in (2) to compute the corresponding penultimate round value $oc_e^{9,b',g}$ for each byte $b \in \{0, \cdots, 15\}$. Let $H_e^{b,g}$ denote the HD between $oc_e^{9,b',g}$ and $oc_e^{10,b'}$ and let the integer array $\mathbf{H}^{b,g} = [H_1^{b,g}, \cdots, H_{N_e}^{b,g}]$ store the HDs for all encryptions; there are 16×256 such arrays. The attackers also observe the probed fields $V_e^{pc,t/f}$ at times t or frequencies f during the last round of AES using a multitude of probe configurations pc—referring to the probe's transverse location l, height h, and orientation o above the DUT. Let the real array $\mathbf{V}^{pc,t/f} = [V_1^{pc,t/f}, \cdots, V_{N_e}^{pc,t/f}]$ store the probed fields (only their magnitudes in frequency domain) for all encryptions; there are $N_l \times N_h \times N_o \times N_{t/f}$ such arrays. Attackers compute the Pearson correlation coefficient $\rho_{\mathbf{H},\mathbf{V}}^{b,g,pc,t/f}$ between the arrays $\mathbf{H}^{b,g}$ and $\mathbf{V}^{pc,t/f}$ for each key byte b, guess g, configuration pc, and time/frequency sample t/f [6], [20],[42]:

$$\rho_{\mathbf{H},\mathbf{V}}^{b,g,pc,t/f} = \frac{\text{Cov}(\mathbf{H}^{b,g}, \mathbf{V}^{pc,t/f})}{\sqrt{\text{Var}(\mathbf{H}^{b,g})\text{Var}(\mathbf{V}^{pc,t/f})}} \quad (3)$$

Attackers can compute the correlation coefficients in (3) using time or frequency samples; e.g., the probed fields $V_1^{pc,t/f}$ are shown in Fig. 3 for $\mathbf{k}_1 = [0x00, 0x01, \cdots, 0x0F]$ and $\mathbf{ip}_1 = [0x00, 0x00, \cdots, 0x00]$.

The largest correlation coefficient will correspond to the correct guess $g^* = k^{10,b}$ for byte b if the leakage model accurately categorizes the underlying sources of emanations (after observing a sufficient number of encryptions); e.g., the coefficients that result from observing $N_e = 4000$ encryptions with randomly generated input plaintexts are shown in Fig. 4. While the correlation coefficient corresponding to the correct guess stands out in Fig. 4, it is important to ask if $k^{10,0}$ could be disclosed by observing fewer encryptions. Indeed, to evaluate side-channel security, *the minimum number of measurements necessary to disclose all key bytes must be quantified*. Let $MTD^{b,pc}$ denote the minimum number of measurements to disclose key byte b when using the probe configuration pc [10], i.e., when $N_e \geq MTD^{b,pc}$, the correlation coefficient corresponding to the correct guess g^* is sufficiently larger than those corresponding to the incorrect guesses. In this paper, a correlation coefficient $\rho_{\mathbf{H},\mathbf{V}}^{b,g,pc,t/f}$ is considered sufficiently large if its maximum value over all time/frequency samples crosses the null hypothesis threshold derived from the inverse t-distribution for a confidence interval of 99.99% [10],[33]. Let

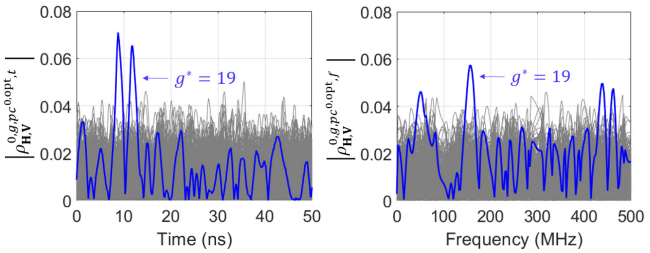


Fig. 4. Time-domain (left) and frequency-domain (right) correlation coefficients for all 256 guesses for $k^{10,0}$, when $N_e = 4000$ encryptions are observed. The coefficient corresponding to the correct guess $g^* = 19$ is shown in blue. The fields were captured using the optimal probe configuration $pc^{0,\text{opt}}$ to disclose byte 0 of the 10th round key [10].

$$pc^{b,\text{opt}} = \underset{pc}{\text{argmin}}\, MTD^{b,pc}\,;\, mMTD^{b} = MTD^{b,pc^{b,\text{opt}}} \quad (4)$$

denote the optimal probe configuration to disclose $k^{10,b}$ and the minimum number of measurements to do so. For the example in Fig. 4, the correct guess for $k^{10,0}$ could be identified by observing time- or frequency-domain fields only for $mMTD^0 \approx$ 600 or 800 encryptions when using $pc^{0,\text{opt}}$ (Fig. 5).

Once all 16 bytes of $\mathbf{k}^{10}$ are disclosed, the AES key-expansion algorithm is inverted to disclose the key $\mathbf{k}^0$, which can then be used to decrypt any ciphertext $\mathbf{oc}_e^{10}$ and recover the corresponding plaintext $\mathbf{ip}_e$ from any past or future encryption.

C. Exhaustive and Optimized Correlation Analysis Attacks

Performing the attack in Section II.B with all $N_\text{l} \times N_\text{h} \times N_\text{o}$ possible probe configurations in the search space to identify $pc^{b,\text{opt}}$, i.e., an exhaustive search for the optimal probe configurations, is infeasible when the search space or $mMTD^b$ are large, e.g., when high-resolution scans are used or the module is secured with countermeasures. Several recently proposed protocols for fine-grained EM SCA attacks can accelerate the search significantly [5], [7], [10], [11], [37].

Adaptive scan algorithms such as greedy [10] or gradient search [11] select probe configurations over multiple scans by introducing constraints on the resolution, search area, or the number of measurements and discarding non-optimal configurations in each scan. These search algorithms may zero in on local minima and cannot guarantee the best probe configuration will be identified, unlike the exhaustive search.

Measurement costs can also be reduced by pre-supposing that information leakage is limited to certain time/frequency samples or locations [5],[7],[37]. In [5], the information-leaking frequency was constant across the search space and a small set of initial guess configurations were used to rapidly isolate leakage to near decoupling capacitors over a test board implementing AES. Similarly, in [37], both the time window and frequencies of information leakage were identified, potentially reducing future measurement costs. Such methods are contingent on the invariance of information-leaking times/frequencies/locations. Repeatability countermeasures, however, can change signal profiles from encryption to encryption (see Section IV.B). Pre-supposing narrow time/frequency/spatial windows to reduce the search space in the presence of such countermeasures can erroneously indicate that a system is resilient. Thus, these methods have limited utility for evaluating EM SCA attack vulnerabilities of hardened implementations. In [7], information-leaking locations were assumed to show maximum peak-to-peak field variation as the module operated in idle and active phases. The intensity of EM fields associated with information leakage, however, are generally not directly related to the intensity of the overall EM fields [5]; indeed, measurement noise and algorithmic noise also contribute to variations in observed signals [4], [6], [9].

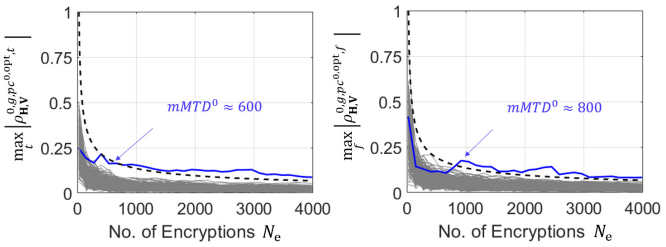


Fig. 5. Maximum value of the time-domain (left) and frequency-domain (right) correlation coefficients for all 256 guesses for $k^{10,0}$ as the number of encryptions increases. The fields were captured using $pc^{0,\text{opt}}$ [10]. The value corresponding to the correct guess $g^* = 19$ (blue) crosses the null hypothesis threshold (dashed) after $mMTD^0$ measurements.

D. Effect of Noise on Correlation Analysis Attacks

The correlation analysis is degraded and EM SCA attacks fail when noise obfuscates the target signals—originating from the computation of byte b of the output ciphertext $oc_e^{10,b}$ in (1)—in the probed fields $\mathbf{V}^{pc,t/f}$ [6]. The noise can be categorized as *measurement noise*, which arises from the environment—temperature variations, vibrations, equipment sensitivity, drift, variability of supply voltage, input clock jitter, etc. [6], [29]—and *algorithmic noise*, which arises from uncorrelated background computations/processes in the DUT [4], [19]. Measurement noise exhibits as variations in observed fields when the exact same encryption is repeated [30]-[32]. For AES-128, the algorithmic noise for the byte b computation includes fields that originate from the computation of the 15 bytes other than byte b of the output ciphertext [19], [24].

To analyze the effect of noise, let's decompose the observed fields in the arrays $\mathbf{V}^{pc,t/f}$ into the independent and hypothetical quantities listed in the arrays $\mathbf{T}^{b,pc,t/f}$, $\mathbf{B}^{b,pc,t/f}$, $\mathbf{N}^{pc,t/f}$ [4], [6]. Here, target signals in $\mathbf{T}$, algorithmic noise in $\mathbf{B}$, and measurement noise in $\mathbf{N}$ arise from computations involving the data of interest ($k^{10,b}$), background computations in the DUT, and other EM sources, respectively. Then, the time-domain correlation coefficient in (3) can be expressed as [6]:

$$\rho_{\mathbf{H},\mathbf{V}}^{b,g,pc,t} = \underbrace{\frac{\text{Cov}\left(\mathbf{H}^{b,g},\mathbf{T}^{b,pc,t}\right)}{\sqrt{\text{Var}\left(\mathbf{H}^{b,g}\right)\text{Var}\left(\mathbf{T}^{b,pc,t}\right)}}}_{\rho_{\mathbf{H},\mathbf{T}}^{b,g,pc,t}} \frac{1}{\sqrt{1+\frac{\text{Var}\left(\mathbf{B}^{b,pc,t}\right)}{\text{Var}\left(\mathbf{T}^{b,pc,t}\right)}+\frac{\text{Var}\left(\mathbf{N}^{pc,t}\right)}{\text{Var}\left(\mathbf{T}^{b,pc,t}\right)}}} \quad (5)$$

The corresponding frequency-domain expression is obtained by replacing the superscripts t with f. In this representation, the noise-free correlation coefficient $\rho_{\mathbf{H},\mathbf{T}}^{b,g,pc,t/f}$ is degraded by the variance terms. Probe configurations that have larger ratios $\text{Var}(\mathbf{T}^{b,pc,t})/\text{Var}(\mathbf{B}^{b,pc,t})$ and $\text{Var}(\mathbf{T}^{b,pc,t})/\text{Var}(\mathbf{N}^{pc,t})$ will yield correlation coefficients $\rho_{\mathbf{H},\mathbf{V}}^{b,g,pc,t/f}$ closer to the noise-free value. The variance ratios in (5) are often combined and represented as signal-to-noise ratio in SCA attacks [24],[42].

Because the entries in the arrays $\mathbf{T}, \mathbf{N}$, and $\mathbf{B}$ are unmeasurable hypothetical quantities, the ratios of their variances cannot be found exactly. They can be estimated,

however, from measured fields via ANOVA [4], [6], [19], [24]. The ANOVA F-statistic, defined as a ratio of variances, is used for hypothesis testing to determine if a dataset is sensitive to variations in a target process. The methodology groups data based on different versions of a target process, and compares variance between groups and variance within groups, to quantify the dependence of the dataset on the target. Here, the F-statistics are used to estimate the two ratios in (5) as [4],[6]:

$$\frac{\mathrm{Var}(\mathbf{T}^{b,pc,t})}{\mathrm{Var}(\mathbf{N}^{pc,t})} \approx F_N^{b,pc,t} \quad \frac{\mathrm{Var}(\mathbf{T}^{b,pc,t})}{\mathrm{Var}(\mathbf{B}^{b,pc,t})} \approx F_B^{b,pc,t} \tag{6}$$

The most accurate estimates in (6) require observing all possible variants in the relevant computations; e.g., to obtain $F_B^{b,pc,t}$, fields can be measured for up to 256×256 possible variants in the switch from $oc_e^{9,b'}$ to $oc_e^{10,b'}$ and $256^{15} \times 256^{15}$ possible variants in background computations. Typically, far fewer samples are sufficient; e.g., the $F_B^{b,pc,t}$ statistic was previously obtained using $oc_e^{10,b'} = \{0,1,\dots,255\}$, ignoring $oc_e^{9,b'}$ values, and 4-40 variants in background computations [19],[24]. The proposed ANOVA method ranks probe configurations according to $F_N^{b,pc,t}$ and $F_B^{b,pc,t}$. If the F-statistics are sufficiently accurate, configurations with the largest F-statistics will include the optimal probe configurations, and those with the smallest ones, which are too sensitive to noise, can be eliminated to accelerate the security evaluation.

III. Measurement Protocol

This section presents the proposed 3-stage measurement protocol that uses ANOVA indicators to evaluate side-channel security. The $F_N^{b,pc,t/f}$ and $F_B^{b,pc,t/f}$ metrics are computed and used to reduce the search space in Stages I and II, respectively. The remaining configurations are used to perform correlation analysis in Stage III and acquire optimal probe configurations. The acquisition cost and measurement time of the proposed protocol are quantified and contrasted to the TVLA indictor. All analyses shown in this section were performed using probed fields obtained with the same measurement setup as in Section II, detailed in Section V.A.

A. Threat Model

As mentioned in the Introduction, the proposed method assumes side-channel security evaluators are less constrained than actual attackers, i.e., they have full control over the input and the encryption key of the DUT (a "gold-box threat model" [9]). This permits evaluators to not just emulate but enhance correlation-analysis attacks, which are applicable even under the highly restrictive black-box threat model [9] but quickly become infeasible for fine-grained EM SCA evaluation (see Section II.C). In particular, fewer restrictions permit evaluators to design targeted tests, estimate the impact of noise, and rapidly identify ineffective probe configurations.

B. Choosing Test Cases to Compute F-statistics

To compute each F statistic, a set of test cases is constructed. Because evaluators are permitted to modify the AES encryption key as well as the input plaintext, each encryption e in the set can use a potentially different plaintext $\mathbf{ip}_e$ and key $\mathbf{k}_e^0$. To construct the test cases, all 16 bytes of the ciphertext in the penultimate round are enforced to be constant and set to zero for simplicity, i.e., $\mathbf{oc}_e^9 = [0\mathrm{x}00, \cdots, 0\mathrm{x}00]$. Thus, the HD between $oc_e^{9,b'}$ and $oc_e^{10,b'}$ is the Hamming weight of $oc_e^{10,b'}$; e.g., $oc_e^{10,b'} = 0\mathrm{x}00$ gives HD_0 and $oc_e^{10,b'} = 0\mathrm{xFF}$ gives HD_8. As a result, evaluators can specify test cases (set each plaintext $\mathbf{ip}_e$ and key $\mathbf{k}_e^0$) by only setting the output ciphertext $\mathbf{oc}_e^{10}$. Once $\mathbf{oc}_e^{10}$ is set, the last round key $\mathbf{k}_e^{10}$ is found from (1) as:

$$k_e^{10,b'} = 0\mathrm{x}63 \oplus oc_e^{10,b'} \tag{7}$$

This is because each byte in the specified $\mathbf{oc}_e^9$ (0x00) is always mapped to 0x63 by AES. Once all 16 bytes of $\mathbf{k}_e^{10}$ are deduced, the key $\mathbf{k}_e^0$ and plaintext $\mathbf{ip}_e$ corresponding to $\mathbf{oc}_e^{10}$ are extracted as detailed in Section II.B. The first two stages of the proposed protocol use test cases constructed with this approach.

The test cases should be chosen based on the leakage model used in the correlation analysis; thus, in this paper, they are chosen using the HD leakage model, where the data of interest $k^{10,b}$ is disclosed by targeting the switching in the last AES round from $oc_e^{9,b'}$ to $oc_e^{10,b'}$. Other leakage models may be more suitable depending on the implementation and algorithm; e.g., test cases were constructed using Hamming weights in [4] to model the fields emanated during data transfer on a processor bus. The HD leakage model used in this paper assumes that the target signals arising from computations involving $k^{10,b}$ have only 9 instead of 256 possible variants $\{\mathrm{HD}_0, \cdots, \mathrm{HD}_8\}$ corresponding to the HD between $oc_e^{9,b'}$ and $oc_e^{10,b'}$, all test cases with the same HD yield indistinguishable target signals, and test cases that correspond to HD_0 and HD_8 are extreme variants, whose target signals differ the most.

C. Stage I: Measurement-Noise-Based Leakage Indicator

In Stage I, the $F_N^{b,pc,t/f}$ statistic is evaluated by using test cases that correspond to extreme variants for the computations of interest and minimize algorithmic noise; i.e., test cases consist of the 2 extreme variants for each byte b—corresponding to HD_0 and HD_8 between $oc_e^{9,b'}$ and $oc_e^{10,b'}$—while the other 15 bytes of $\mathbf{oc}_e^{10}$ are kept constant and set to 0x00 (HD_0). Because the test case corresponding to $\mathbf{oc}_e^{10} = \mathbf{0}$ can be reused as one of the extreme variants for each byte and because the remaining test cases are generated by changing only one of 16 bytes of $\mathbf{oc}_e^{10}$ to 0xFF (HD_8), a total of $N_{e,I} = 17$ plaintext-key pairs are used as test cases in Stage I. The HDs for these 17 test cases can be stored in a 17×16 integer array :

$$\mathbf{H}_I = \begin{bmatrix} \mathrm{HD}_0 & \mathrm{HD}_0 & & \mathrm{HD}_0 \\ \mathrm{HD}_8 & \mathrm{HD}_0 & \cdots & \mathrm{HD}_0 \\ \mathrm{HD}_0 & \mathrm{HD}_8 & & \mathrm{HD}_0 \\ \vdots & & \ddots & \vdots \\ \mathrm{HD}_0 & \mathrm{HD}_0 & \cdots & \mathrm{HD}_8 \end{bmatrix} \tag{8}$$

These $N_{e,I}$ encryptions are repeated $N_{r,I}$ times for each possible probe configuration and the F-statistic is evaluated as:

$$F_N^{b,pc,t/f} = \frac{2N_{r,I} \times \mathrm{Var}(\bar{x}_{\mathrm{HD}_0}^{b,pc,t/f}, \bar{x}_{\mathrm{HD}_8}^{b,pc,t/f})}{\mathrm{Mean}(s_{\mathrm{HD}_0}^{b,pc,t/f}, s_{\mathrm{HD}_8}^{b,pc,t/f})} \tag{9}$$

Here, sample means $\bar{x}_{\mathrm{HD}_{0/8}}^{b,pc,t/f}$ and variances $s_{\mathrm{HD}_{0/8}}^{b,pc,t/f}$ of the probed fields are computed across the $N_{r,I}$ samples. The fields for $\bar{x}_{\mathrm{HD}_0}^{b,pc,t/f}$ and $s_{\mathrm{HD}_0}^{b,pc,t/f}$ ($\bar{x}_{\mathrm{HD}_8}^{b,pc,t/f}$ and $s_{\mathrm{HD}_8}^{b,pc,t/f}$) are observed

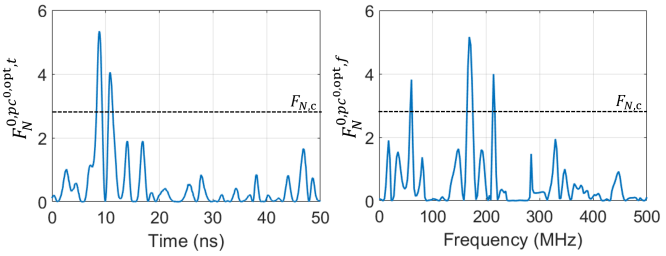


Fig. 6. Time-domain (left) and frequency-domain (right) $F_N^{0,pc,t/f}$ metric, evaluated with the probe configuration $pc^{0,\mathrm{opt}}$.

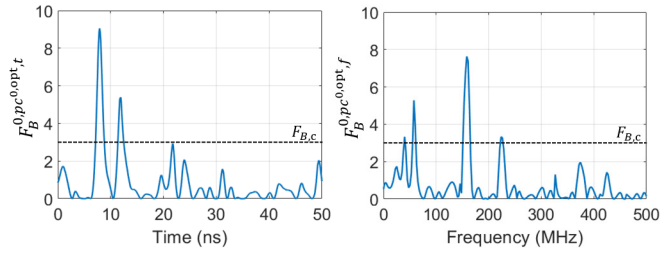


Fig. 7. Time-domain (left) and frequency-domain (right) $F_B^{0,pc,t/f}$ metric, evaluated with the probe configuration $pc^{0,\mathrm{opt}}$.

using the test case generated by setting $oc_e^{10,b'}$ to 0x00 (0xFF) and all other bytes of $\mathbf{oc}_\mathrm{e}^{10}$ to 0x00, i.e., the test case in row 1 $(b'+2)$ of $\mathbf{H}_\mathrm{I}$.

An example of the F-statistic computed using $N_{\mathrm{r,I}} = 30$ repetitions is shown in Fig. 6. Comparing the data to Fig. 4 shows that, $F_N^{b,pc,t/f}$ is large whenever $\rho_{\mathbf{H},\mathbf{V}}^{b,g,pc,t/f}$ is large but the converse is not true, i.e., the indicator captures the information leakage but also overestimates it. The computed F-values are compared to a threshold $F_{N,c}$ to generate a leakage indicator:

$$Indicator_\mathrm{I}^{b,pc} = \begin{cases} 1 & \text{if } \max\limits_{t/f} F_N^{b,pc,t/f} \geq F_{N,c} \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

Only configurations with indicator value 1 are selected for measurements in Stage II, i.e., only $N_{\mathrm{pc,II}}^b = \sum_{pc} Indicator_\mathrm{I}^{b,pc}$ probe configurations are used.

D. Stage II: Algorithmic-Noise-based Leakage Indicator

In Stage II, the $F_B^{b,pc,t/f}$ statistic is evaluated by using test cases that correspond to extreme variants for both the computations of interest and background computations. Test cases consist of the 2 extreme variants for each byte b, while 14 of the remaining 15 bytes of $\mathbf{oc}_e^{10}$ are kept constant at 0x00 (HD_0) and 1 other byte is set to the 2 extreme variants. Consider the 32 test cases for byte $b = 0$: In half of these cases, $oc_e^{10,0}$ (byte 0 is not impacted by ShiftRows, so $b' = b$) is 0x00 (HD_0) or 0xFF (HD_8); for each half, $N_\mathrm{B} = 16$ background process variants are generated by setting all or all but one of the remaining bytes of $\mathbf{oc}_\mathrm{e}^{10}$ to HD_0. The HDs for these 32 test cases can be stored in an integer array of size 32×16:

$$\mathbf{H}_\mathrm{II}^0 = \begin{bmatrix} \mathrm{HD}_0 & \mathrm{HD}_0 & \mathrm{HD}_0 & \cdots & \mathrm{HD}_0 \\ \mathrm{HD}_0 & \mathrm{HD}_8 & \mathrm{HD}_0 & \cdots & \mathrm{HD}_0 \\ \mathrm{HD}_0 & \mathrm{HD}_0 & \mathrm{HD}_8 & \cdots & \mathrm{HD}_0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathrm{HD}_0 & \mathrm{HD}_0 & \mathrm{HD}_0 & \cdots & \mathrm{HD}_8 \\ \mathrm{HD}_8 & \mathrm{HD}_0 & \mathrm{HD}_0 & \cdots & \mathrm{HD}_0 \\ \mathrm{HD}_8 & \mathrm{HD}_8 & \mathrm{HD}_0 & \cdots & \mathrm{HD}_0 \\ \mathrm{HD}_8 & \mathrm{HD}_0 & \mathrm{HD}_8 & \cdots & \mathrm{HD}_0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathrm{HD}_8 & \mathrm{HD}_0 & \mathrm{HD}_0 & \cdots & \mathrm{HD}_8 \end{bmatrix} \quad (11)$$

Similar test cases and their HD arrays $\mathbf{H}_\mathrm{II}^b$ are constructed for all bytes b. The first 17 rows of each $\mathbf{H}_\mathrm{II}^b$ is a reordering of the 17 test cases in $\mathbf{H}_\mathrm{I}$; thus, only $N_{\mathrm{e,II}}^b = 15$ new plaintext-encryption key pairs are needed for each byte in Stage II.

Using these test cases, the F-statistic is evaluated as

$$F_B^{b,pc,t/f} = \frac{2N_\mathrm{B} \times \mathrm{Var}(\bar{\bar{x}}_{\mathrm{HD}_0}^{b,pc,t/f}, \bar{\bar{x}}_{\mathrm{HD}_8}^{b,pc,t/f})}{\mathrm{Mean}(\bar{s}_{\mathrm{HD}_0}^{b,pc,t/f}, \bar{s}_{\mathrm{HD}_8}^{b,pc,t/f})} \quad (12)$$

Here, the sample means $\bar{\bar{x}}_{\mathrm{HD}_{0/8}}^{b,pc,t/f}$ and variances $\bar{s}_{\mathrm{HD}_{0/8}}^{b,pc,t/f}$ are computed across the N_B samples. The fields for $\bar{\bar{x}}_{\mathrm{HD}_0}^{b,pc,t/f}$ and $\bar{s}_{\mathrm{HD}_0}^{b,pc,t/f}$ ($\bar{\bar{x}}_{\mathrm{HD}_8}^{b,pc,t/f}$ and $\bar{s}_{\mathrm{HD}_8}^{b,pc,t/f}$) are observed using the test cases in rows 1-16 (17-32) of $\mathbf{H}_\mathrm{II}^b$. Extra bars are used above the sample means and variances because the tests are repeated $N_{\mathrm{r,II}}$ times and the probed fields are first averaged over them. The number of repetitions per test case in Stage II can be lower than that in Stage I, i.e., $N_{\mathrm{r,II}} < N_{\mathrm{r,I}}$, in part because configurations most sensitive to measurement noise are discarded in Stage I and in part because the goal is to reduce noise rather than accurately capture variations in repeated measurements.

An example of the F-statistic computed with $N_{\mathrm{r,II}} = 10$ repetitions is shown in Fig. 7. Comparing Figs. 4, 6, and 7 it can be observed that both F-statistics must be maximized to successfully disclose the encryption key. Similar to Stage I, the computed F-values are compared to a threshold $F_{B,c}$ to generate a leakage indicator:

$$Indicator_\mathrm{II}^{b,pc} = \begin{cases} 1 & \text{if } \max\limits_{t/f} F_B^{b,pc,t/f} \geq F_{B,c} \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

Configurations with indicator value 0 are eliminated at the end of Stage II, i.e., only $N_{\mathrm{pc,III}}^b = \sum_{pc} Indicator_\mathrm{II}^{b,pc}$ probe configurations are used in Stage III . The thresholds $F_{N,c}$ and $F_{B,c}$ are derived from F-distributions for a 90% confidence level.

E. Stage III: ANOVA-Informed Correlation Analysis

In Stage III, correlation analysis is performed to identify $pc^{b,\mathrm{opt}}$ by using only the probe configurations not eliminated at the end of Stage II. One potential approach, after collecting N_e measurements, is to repeatedly compute the correlation coefficient in (3), starting with N_e encryptions, followed by $N_\mathrm{e} - 1$ encryptions, and so on, until $MTD^{b,pc}$ is identified, i.e., where the coefficient for the correct guess drops below the null hypothesis threshold (Fig. 5). This requires $\mathrm{O}(N_\mathrm{e})$ to $\mathrm{O}(N_e^2)$ operations; alternatively, a binary search algorithm can be implemented to identify $MTD^{b,pc}$ in $\mathrm{O}(N_\mathrm{e} \log N_\mathrm{e})$ operations [33]. Stage III ends by identifying $mMTD^b$ and $pc^{b,\mathrm{opt}}$ for each byte b.

A naïve approach to ensure $mMTD^b$ is identified in Stage III is to set $N_\mathrm{e} = N_\mathrm{e}^{\max}$, a large number of encryptions that ensures all key bytes are disclosed. Alternatively, the F-values found in Stage II can be used to inform the search and potentially reduce the measurement costs of Stage III: In this approach, $N_{\mathrm{scan,III}}^b$

scans are performed for each byte b, using all $N^b_{\text{pc,III}}$ probe configurations. Before these scans, the probe configurations are arranged in descending order of their F-values found in Stage II as $\{pc^{b,1}, pc^{b,2}, \dots, pc^{b,N^b_{\text{pc,III}}}\}$. In each scan $s = 1, \dots, N^b_{\text{scan,III}}$, an initial estimate of $mMTD^b$ is chosen as $mMTD^{b,\text{est}}_s$ and $mMTD^{b,\text{est}}_s$ encryptions are observed using each configuration $pc^{b,i}$ for $i = 1, \dots, N^b_{\text{pc,III}}$. If $MTD^{pc^{b,i}}_b < mMTD^{b,\text{est}}_s$ for any configuration, the remaining configurations are evaluated by reducing $mMTD^{b,\text{est}}_s$ to $MTD^{pc^{b,i}}_b$. The estimate is so updated throughout the scan and this process continues until all $N^b_{\text{pc,III}}$ probe configurations are tested. The scans are terminated if $MTD^{pc^{b,\text{i}}}_b < mMTD^{b,\text{est}}_s$ for any probe configuration and the key was disclosed. Otherwise, $mMTD^{b,\text{est}}_s$ is increased to $mMTD^{b,\text{est}}_{s+1}$ and the process is repeated until the key is disclosed; e.g., in this paper, each scan incremented the estimate by 500 encryptions. If the number of encryptions is increased, only the additional $mMTD^{b,\text{est}}_{s+1} - mMTD^{b,\text{est}}_s$ encryptions have to be observed because the observations from the previous scan can be reused when computing the correlation coefficient. In the best-case/ minimum-cost scenario, the first configuration tested in the first scan reveals $pc^{b,\text{opt}}$ and $mMTD^b$, while in the worst-case/ maximum-cost scenario, the final configuration at the end of the final scan reveals the optimal configuration. Therefore, $mMTD^b \le N^b_{\text{e,III}} \le mMTD^{b,\text{est}}_{N^b_{\text{scan,III}}}$ encryptions are observed with each probe configuration in Stage III.

F. Measurement Costs

The acquisition cost of the proposed protocol is the total number of measurements in each stage, which is the product of the number of encryptions observed per configuration, number of repetitions, and number of configurations probed, i.e.,

$$\begin{aligned} Acquis.Cost \quad &= N_{\text{e,I}} N_{\text{r,I}} N_{\text{l}} N_{\text{h}} N_{\text{o}} + \\ &\sum_{b=1}^{16} N^b_{\text{e,II}} N_{\text{r,II}} N^b_{\text{pc,II}} + \\ &\sum_{b=1}^{16} N^b_{\text{e,III}} N^b_{\text{pc,III}} \quad \text{(ANOVA)} \end{aligned} \tag{14}$$

Once the acquisition cost is accrued, the *marginal cost* of future evaluations can be reduced by reusing probe configurations $pc^{b,\text{opt}}$ and performing only the minimum number of measurements $mMTD^b$ for each byte. The marginal cost of future evaluations is [10],[33],

$$Marginal\ Cost = \sum_{b=1}^{16} mMTD^b \tag{15}$$

The marginal cost of evaluating a module employing a countermeasure is compared to that of a baseline module to quantify the improvement in the EM side-channel security of hardened AES modules in this paper. In the most resilient modules, some key bytes may potentially not be disclosed [20]. In these cases, to limit the measurement costs of the evaluation, the number of encryptions performed per configuration is restricted to be no more than $N^{\max}_{\text{e}}$.

G. Alternatives to Proposed Approach

The proposed protocol is compared to several alternatives in Sections V and VI. The exhaustive search method (Section II. C) [10], [34] is one potential alternative. It performs correlation analysis by observing $N^{\max}_{\text{e}}$ encryptions across the entire search space of probe configurations in a single, high-resolution scan. As a result, the exhaustive approach requires [10], [33]

$$Acquis.Cost = N^{\max}_{\text{e}} N_{\text{l}} N_{\text{h}} N_{\text{o}} \text{ (exhaustive search)} \tag{16}$$

measurements to be observed.

A more viable correlation-analysis approach is the greedy-search adaptive scan protocol implemented in [10], [20], [33] and briefly described next. Phase I of the protocol identifies $pc^{b,\text{opt}}_0$ and $mMTD^b_0$ for each byte b, using $N_{\text{scan,I}}$ progressively more expensive low-resolution scans, performed over the entire chip area. Phase I only terminates once all key bytes are disclosed, increasing the number of locations $N_{\text{l},s,\text{I}}$ and encryptions $N_{\text{e},s,\text{I}}$ in each scan s, until this goal is achieved. These configurations are further optimized in Phase II, using $N_{\text{scan,II}}$ progressively cheaper byte-wise scans, constraining the area around $pc^{b,\text{opt}}_{s-1}$ and the number of measurements to $mMTD^b_{s-1}$ in each subsequent scan s. The initial search space can be reduced with a cheap pre-characterization stage consistent with the black-box threat model: Observe $N^{\text{pre}}_{\text{e}}$ encryptions with all configurations and discard those with small $\text{Var}\left(\mathbf{V}^{pc,t/f}\right)$, i.e., with little observed field variations. This greedy-search protocol requires [10],

$$\begin{aligned} Acquis.Cost = N^{\text{pre}}_{\text{e}} N_{\text{h}} N_{\text{o}} N_{\text{l}} + \sum_{s=1}^{N_{\text{scan,I}}} N_{\text{e},s,\text{I}} N_{\text{h}} N_{\text{o}} N_{\text{l},s,\text{I}} + \\ \sum_{b=1}^{16} \sum_{s=1}^{N_{\text{scan,II}}} mMTD^b_{s-1} N_{\text{l},s,\text{II}} \text{ (Greedy Search)} \end{aligned} \tag{17}$$

measurements. Note that this approach can have unlimited cost, e.g., for hardened modules, if the number of scans is not bounded. In practice, the acquisition cost of this protocol should be bounded by that of the exhaustive search method in (16) by limiting its phase I to at most $N^{\max}_{\text{e}}$ encryptions and its phase II to have at most the same resolution as the exhaustive search.

Another alternative is the TVLA method, a commonly used leakage indicator, including in the ISO/IEC 17825 standard [43],[44], to evaluate the side-channel resilience of crypto-systems [11],[12],[22],[23]. The TVLA method also statistically characterizes the probed fields for specially constructed test cases. Here, the DUT is assumed to be a "white box" [9], where evaluators can control the inputs to the chip but not the encryption key. It uses Welch's t-test to compare the means of two sets of observed fields—a reference set (SetA), where inputs are fixed, and a test set (SetB), where the inputs are randomly generated—hypothesizing that information leakage is present if there are significant changes in the means of the two sets. In SetA, one plaintext is repeated N_{SetA} times for a fixed key; in SetB, N_{SetB} randomly generated inputs are encrypted using the same key as SetA. Computing the sample means $\bar{x}^{pc,t/f}_{\text{SetA/SetB}}$ and variances $s^{pc,t/f}_{\text{SetA/SetB}}$ across the $N_{\text{SetA}}/N_{\text{SetB}}$ samples, the Welch t-test is evaluated as:

$$T^{pc,t/f} = \frac{\bar{x}^{pc,t/f}_{\text{SetB}} - \bar{x}^{pc,t/f}_{\text{SetA}}}{\sqrt{s^{pc,t/f}_{\text{SetB}}/N_{\text{SetB}} + s^{pc,t/f}_{\text{SetA}}/N_{\text{SetA}}}} \tag{18}$$

Using the parameters in [11], an example TVLA metric computed for 200 fixed plaintext and 200 random plaintext is shown in Fig. 8. In addition to accurately indicating leakages at ~10 ns/~200 MHz (Fig. 4), the TVLA also shows exaggerated

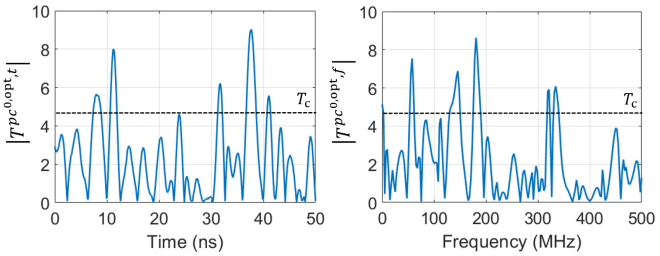


Fig. 8. Time-domain (left) and frequency-domain (right) TVLA metric, evaluated with the probe configuration $pc^{0,\text{opt}}$.

leakage at ~40 ns/~300 MHz. Because test cases are randomized without any restrictions, the TVLA method is leakage-model independent and can be used as a generic approach to analyze side-channel leakage; in contrast, the ANOVA approach described in this work constructs test cases based on the leakage model used in the correlation analysis. The results of TVLA are not necessarily linked, however, to the number of measurements needed to disclose the key [11], [23], [44]. Furthermore, results of low-cost TVLA experiments using fewer encryptions ($N_{\text{SetA}}, N_{\text{SetB}} \approx$ 200-500) may have limited accuracy [11]. Increasing the number of encryptions (N_{SetA}, $N_{\text{SetB}} \approx$ 20000) to improve accuracy [12] is infeasible for fine-grained EM SCA attacks as the acquisition cost would approach the exhaustive search. The computed T-statistic can be compared with a threshold T_c to generate another indicator:

$$Indicator_{\text{TVLA}}^{pc} = \begin{cases} 1 & \text{if } \max\limits_{t/f} T^{pc,t/f} \geq T_c \\ 0 & \text{otherwise} \end{cases} \quad (19)$$

Once probe configurations with 0 TVLA indicator values are eliminated, correlation analysis is performed only with $N_{\text{pc,TVLA}} = \sum_{pc} Indicator_{\text{TVLA}}^{pc}$ configurations. An exhaustive search (TVLA+e) would observe $N_{\text{e}}^{\max}$ encryptions at each configuration. Alternatively, a TVLA-informed search (TVLA+i) similar to that in Section III.E can be used to reduce the measurement costs. In this approach, $N_{\text{scan,TVLA}}^{b}$ scans are performed for each byte b and $mMTD^b \leq N_{\text{e,TVLA}}^{b}$ encryptions are observed with each probe configuration. The acquisition cost of these two protocols are

$$\begin{aligned} Acquis.\ Cost &= (N_{\text{SetA}} + N_{\text{SetB}}) N_{\text{l}} N_{\text{h}} N_{\text{o}} + \\ &\quad N_{\text{e}}^{\max} N_{\text{pc,TVLA}}\ (\text{TVLA} + \text{e}) \\ Acquis.\ Cost &= (N_{\text{SetA}} + N_{\text{SetB}}) N_{\text{l}} N_{\text{h}} N_{\text{o}} + \\ &\quad \textstyle\sum_{b=1}^{16} N_{\text{e,TVLA}}^{b} N_{\text{pc,TVLA}}\ (TVLA + i) \end{aligned} \quad (20)$$

IV. Devices Under Test

This section describes the 9 AES implementations (2 baseline and 7 hardened ones) whose vulnerability to EM SCA attacks is evaluated with the proposed protocol. The countermeasures in these implementations are separated into three categories representing different strategies to secure the chip. The countermeasures tested in this paper are based on existing implementations in [18]-[20], [26],[27].

A. Baseline AES Implementations

The first baseline AES module was implemented on an Artix-7 FPGA with 20 mm × 20 mm chip size tested on the CW305 evaluation board [35]. The evaluation board, which was specifically designed to demonstrate SCA attacks, allowed the clock frequency and supply voltage to be changed. As a baseline scenario, the chip was operated at clock frequency of $f^{\text{clk}} = 20$ MHz and supply voltage of $V^{\text{s}} = 1$ V. This baseline implementation is used as a reference to test 3 repeatability countermeasures (Section IV.B), 1 algorithmic countermeasure (Section IV.C), and 1 physical design strategy (Section IV.D), all implemented on the same FPGA.

The second baseline AES module was an ASIC with 10 mm × 10 mm chip size [36]. The chip was operated at input clock frequency $f^{\text{clk}} = 37.5$ MHz and supply voltage $V^{\text{s}} = 1.1$ V. It is used as a reference for testing 2 physical design strategies implemented on the same chip.

B. AES Implementations with Repeatability Countermeasures

Observed fields depend on the DUT's operating supply voltage and clock frequency. Randomly scaling these parameters can create temporal shifts and modify amplitudes in observed signals, reducing the repeatability of experiments and increasing measurement noise. Three such countermeasures based on EM interference reduction techniques [20] are tested in this paper:

1) Frequency Scaling (FS): Randomizing clock frequency creates delays in the circuit and misaligns measurements over multiple encryptions. While this jitter dithers time-domain signals [14], frequency-domain EM SCA attacks remain effective against this countermeasure. The FS countermeasure was implemented by varying the clock frequency in the range $f^{\text{clk}} = 20\ \text{MHz} \pm 0.25\ \text{MHz}$.

2) Voltage Scaling (VS): Voltage scaling desensitizes peak-to-peak amplitudes of observed fields to the data being encrypted [15]. This countermeasure obfuscates both time- and frequency-domain fields. The VS countermeasure was implemented by varying the input supply in the range $V^{\text{s}} = 1\ \text{V} \pm 0.05\ \text{V}$.

3) Voltage-Frequency Scaling (VFS): This countermeasure combines the VS and FS countermeasures to provide maximum dithering of fields in both time- and frequency-domain [16]. The VFS countermeasure was implemented by simultaneously varying the input supply and clock frequency in the ranges selected in the VS and FS countermeasure (set 2 in [20]).

These countermeasures were implemented on the FPGA, using the programmable clock and voltage supply, such that 5 fixed states of voltage, frequency, or voltage-frequency pairs were chosen within the selected ranges. These countermeasures can be implemented with relatively low overhead [15], [16].

C. AES Implementations with Algorithmic Countermeasures

Countermeasures artificially introducing algorithmic noise typically introduce additional operations/modify data flow in the algorithm. Examples include hiding and masking [19], [26], [27], where exploitable intermediate round outputs are modified to break correlation with observed fields. A majority of countermeasures in this category for AES focus on masking non-linear Sbox operations using novel transformations or changes to existing implementations; e.g., in [19], a byte permutation (BP) network that rearranges bytes randomly was proposed as a precursor to Sbox operations and AES correctness was maintained by using an inverse BP network to re-order bytes at the end of each round. This method showed

limited resilience improvement (~3.2×) for a black-box threat model [19]. Therefore, in addition to the hardened Sbox implementation in [19], a simple Boolean XOR operation for linear operation masking [26], [27] is used to hide the intermediate state register value in this paper. Here, a random "mask" variable $\mathbf{M}_m$ changes $\mathbf{oc}_e^9$ to masked value $\mathbf{oc}_{e,m}^9$ in the penultimate round. The last round begins with "unmasking" and byte-order randomization using the BP network. After Sbox operation, bytes are re-ordered with the inverse BP network, followed by the shift rows operation. The final operations of AES can be summarized as,

$$\begin{aligned} oc_{e,m}^{9,b'} &= oc_e^{9,b'} \oplus M_m^{b'} \\ \tilde{\imath v}_e^{10,b} &= ShiftRows\left(BP^{-1}\left[Sbox\left(BP\left[oc_{e,m}^{9,b} \oplus M_m^b\right]\right)\right]\right) \\ oc_{e,m}^{10,b} &= \tilde{\imath v}_e^{10,b} \oplus k^{10,b} \end{aligned} \quad (21)$$

This countermeasure was implemented on the FPGA using the nominal clock frequency and input supply. While it can be an effective countermeasure, masking incurs significant area and delay overheads [19]; moreover, it can be vulnerable to higher-order attacks [26], [27] outside the scope of this paper, where the mask is attacked first, followed by the key.

D. AES Implementations with Physical Design Strategies

These countermeasures minimize data-dependent variations in observed fields by implementing dedicated signal attenuation hardware [13], modifying the chip's physical design [18], or shielding the module [17]. These may not be effective at all frequencies of interest, can increase packaging costs, or increase the area overhead. In this paper, 3 such countermeasures are implemented: In the first one, a 25-μm thick aluminum foil is placed over the FPGA to attenuate fields and degrade EM SCA attacks. The other two countermeasures are implemented on the ASIC and involve changes to the AES module's power grid. The first design implements a "twisted pair" grid structure [18]; the second one uses wider and thicker power rails to shield signals from lower metal layers [18].

V. Setup and Baseline Results

This section presents the measurement setup and results for the baseline FPGA and ASIC implementations of AES-128. The proposed method is compared to alternatives in terms of acquisition costs. All spatial maps of fields and computed statistics in this section were obtained with the *x*-oriented probe.

A. Measurement Setup

The setup used a 1-mm H-field probe from Langer [10],[33] fixed at $h_0 = 0.5$ mm to scan both chips at $N_l = 51 \times 51$ locations in $N_o = 2$ orientations. This initial search space [6] can be expanded as per the probe's resolution and the technology node used for implementation. For each encryption, measurements were recorded for the last clock cycle of AES, at a sampling rate of 10 GS/s. To boost the amplitude of measured fields, the probe was connected to a 30 dB amplifier [10], [33]. A Keysight DSOS054A oscilloscope was used to capture signals. The oscilloscope could store up to 10000 waveforms in its memory and had sufficient processing capability to perform analysis on these waveforms, removing potential bottlenecks resulting from data transfer [33]. The probe was positioned using a Riscure EM Probe station. More details on the measurement equipment are given in [33]. All equipment and chip inputs were controlled using an automated script. The ASIC used an additional Arduino interface, which acts as an intermediary during the transfer of plaintext and keys from the main computer. To demonstrate the spatial resolution, maps of time- and frequency-domain fields at information leaking time/frequency samples are plotted in Fig. 9 for the two DUTs, averaged over 30 repeated measurements. These composite images are obtained one pixel/measurement at a time by re-positioning the probe and repeating the encryption.

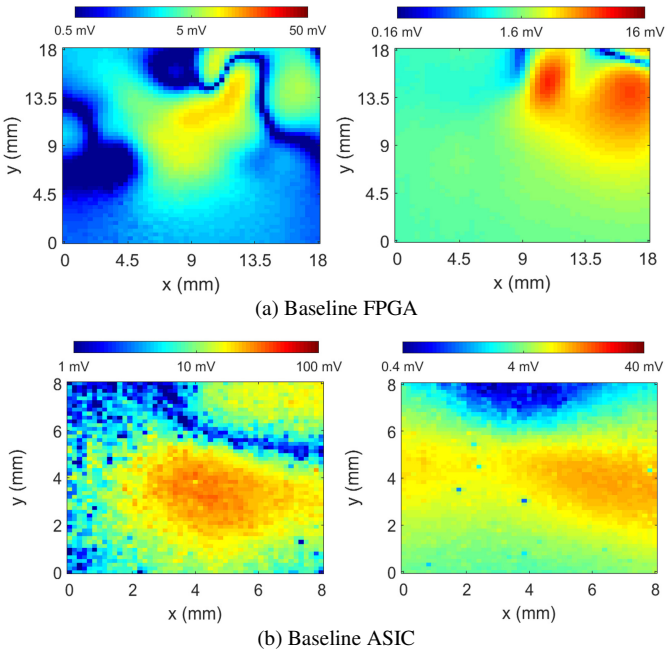
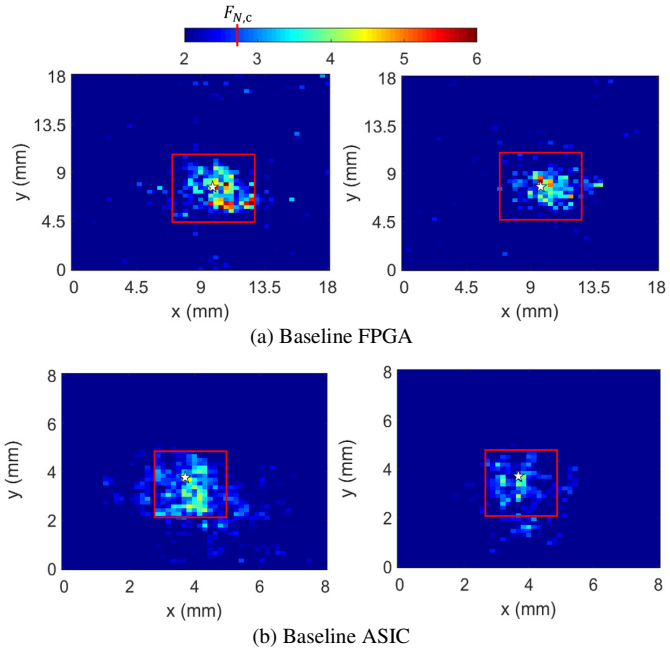


Fig. 9. Spatial map of (a) time-domain signals at ~8 ns (left) and frequency-domain signals at ~160 MHz (right) for the FPGA module detailed in [10], and (b) time-domain signals at ~6 ns (left) and frequency-domain signals at ~100 MHz (right) for the ASIC module detailed in [18].

B. Proposed Protocol Results

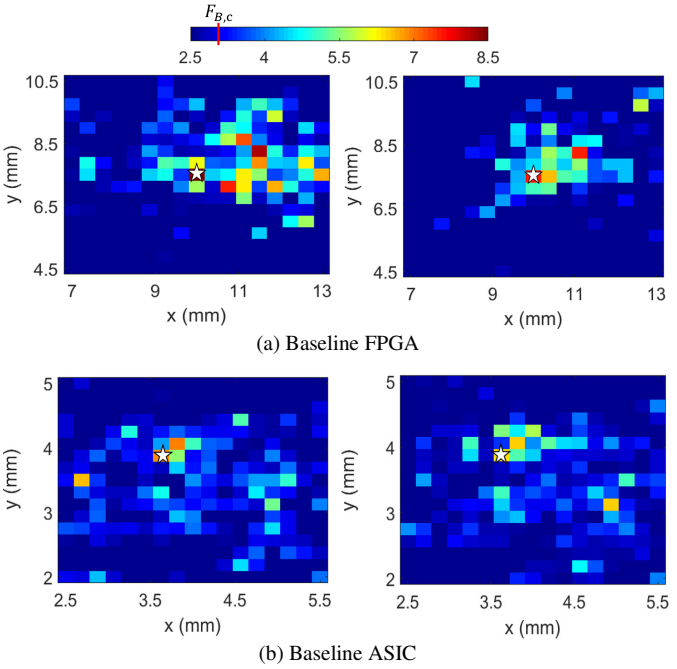
The Stage I $F_N^{b,pc,t/f}$ metric was computed by repeating the $N_{e,I} = 17$ encryptions detailed in Section III.C $N_{r,I} = 30$ times. Spatial maps of the maximum $F_N^{0,pc,t/f}$ are plotted in Fig. 10; a large portion of the configurations with high F-values were located inside the areas marked with red boxes. Fig. 10 shows that frequency-domain analysis discarded more configurations in Stage I. The Stage II $F_B^{b,pc,t/f}$ metrics were computed by repeating the $N_{e,II}^b = 15$ encryptions detailed in Section III.D $N_{r,II} = 10$ times and averaging the signals. Spatial maps of the maximum $F_B^{0,pc,t/f}$ are shown in Fig. 11 only for the areas marked with red boxes in Fig. 10 for simplicity (high F-ratio configurations outside the red boxes were also evaluated in Stage II). Then, configurations whose maximum $F_B^{0,pc,t}$ were larger than $F_{B,c}$ were tested in Stage III to find the optimal probe configurations, using at most $N_{scan,III}^b = 2/3\ (6/8)$ scans in time/frequency domain for the baseline FPGA (ASIC). Each scan incremented the estimate $mMTD_s^{b,est}$ by 500. The acquisition costs of the protocol are listed in Table I. The table shows that the time-domain evaluation required ~1.2× (~1.1×)



(a) Baseline FPGA

(b) Baseline ASIC

Fig. 10. Spatial map of $\max_t F_N^{0,pc,t}$(left) and $\max_f F_N^{0,pc,f}$ (right) for the baseline (a) FPGA [10] and (b) ASIC [18]. Optimal configurations are shown with stars.



(a) Baseline FPGA

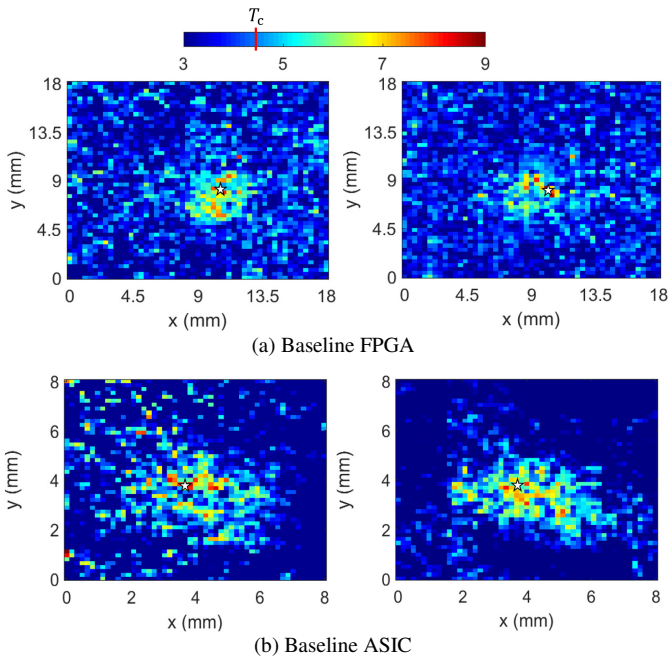
(b) Baseline ASIC

Fig. 11. Spatial map of $\max_t F_B^{0,pc,t}$(left) and $\max_f F_B^{0,pc,f}$ (right) for the baseline (a) FPGA [10] and (b) ASIC [18] in the locatoins inside the red boxes in Fig. 10. Optimal configurations are shown with stars.

more measurements than the frequency-domain one for the FPGA (ASIC).

C. Cost Comparison to Alternative Methods

Let's first compare the proposed method for evaluating EM SCA vulnerability to emulating correlation-analysis attacks. Using an exhaustive scan, where $N_e^{\max} = 20000$ encryptions are observed with every probe configuration in the search space, correlation analysis would require ~10^8 measurements. The acquisition cost can be lowered with adaptive scan



(a) Baseline FPGA

(b) Baseline ASIC

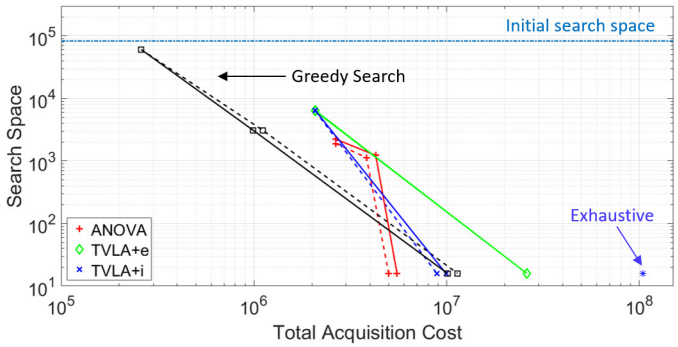
Fig. 12. Spatial map of $\max_t T^{pc,t}$(left) and $\max_f T^{pc,f}$ (right) for the baseline (a) FPGA [10] and (b) ASIC [18]. Optimal configurations are shown with stars.

TABLE I
PROPOSED ANOVA METHOD'S COSTS

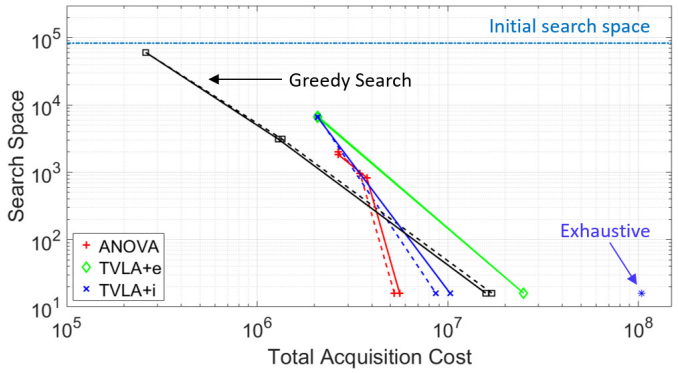
Acquisition Cost	Baseline FPGA		Baseline ASIC	
	Time Domain	Frequency Domain	Time Domain	Frequency Domain
Stage I (×10^6)	2.65	2.65	2.65	2.65
Stage II (×10^6)	1.62	1.18	1.11	0.81
Stage III (×10^6)	1.24	1.16	1.83	1.77
Total (×10^6)	5.52	4.49	5.59	5.24

protocols. Here, the greedy search protocol [10] was implemented with a pre-characterization stage: Every probe configuration was used to observe fields for $N_e^{\text{pre}} = 50$ random encryptions and configurations where the standard deviation was < 0.1 mV were removed from the search space. In Phase I, $N_{\text{scan,I}} = 2$ (3) scans were performed for the FPGA (ASIC) and optimal configurations were identified by using $N_{\text{e,2,I}} = 5000$ ($N_{\text{e,2,I}} = 5000$ and $N_{\text{e,3,I}} = 8000$) encryptions; in phase II, $N_{\text{scan,II}} = 2$ (2) scans were performed for each byte. The final costs of implementing the protocol on the baseline FPGA (ASIC) were found to be ~1.0/1.1×10^7 (~1.6/1.7×10^7) measurements in time/frequency domain. Therefore, the proposed protocol was observed to be ~17-22× cheaper than the exhaustive approach and ~2-3× cheaper than the adaptive acquisition approach for the baseline cases.

Next, let's compare the proposed ANOVA-based method to TVLA-based alternatives. Here, TVLA was implemented using $N_{\text{SetA}} = N_{\text{SetB}} = 200$ encryptions for both baseline implementations. Spatial maps of the maximum $T^{pc,t/f}$ are shown in Fig. 12. Numerous "false positives" are observed throughout the search space, especially for the FPGA. Using the TVLA+e protocol on the FPGA (ASIC) required ~2.6/2.8×10^7 (~2.5/2.4×10^7) measurements in time/frequency domain. The TVLA+i protocol required $N^b_{\text{scan,TVLA}} = 2/3$ (6/8) scans and ~9.9/10.2×10^6 (~8.8/8.6×10^6) measurements in time/frequency



(a) Baseline FPGA



(b) Baseline ASIC

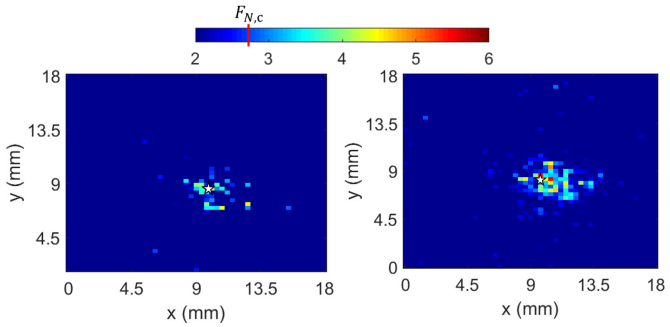
Fig. 13. Reduction of search space for the optimal probe configuration in time (solid) and frequency domain (dashed) for the baseline (a) FPGA [10] and (b) ASIC [18]. Unlike the exhaustive- and greedy-search protocols, which emulate correlation analysis by actual attackers with restricted access, the TVLA and ANOVA protocols accelerate the process by computing statistical metrics.

domain. Therefore, the proposed protocol was observed to be ~4-5× cheaper than the TVLA+e and ~1.5-2× cheaper than the TVLA+i method for the baseline cases.

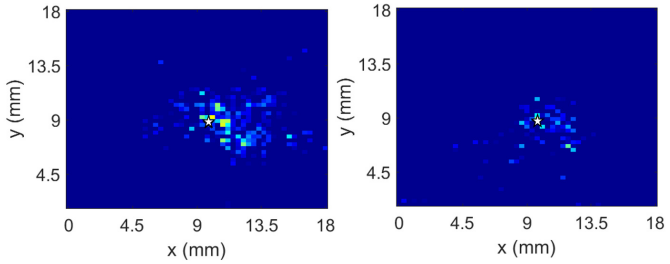
All protocols identified similar information-leaking configurations and minimum MTDs, although each protocol required different acquisition cost to reach the final result. All protocols began with the same maximum search space ($N_o \times N_l = 2 \times 51 \times 51$ configurations), at 0 acquisition cost, and ended with 16 optimal configurations (one for each byte) after accruing the acquisition cost of the measurements. The costs of the protocols are plotted in Fig. 13, along with the reduction of the search space at each stage/phase. The search space size at the end of each stage is the sum of remaining possible probe configurations identified for each byte. Fig. 13 shows that the methods' performances were rather insensitive to whether time- or frequency-domain signals were used and that the proposed protocol outperformed the alternatives for the baseline implementations. Whether the same observations apply to hardened implementations is presented next.

VI. Results For Countermeasures

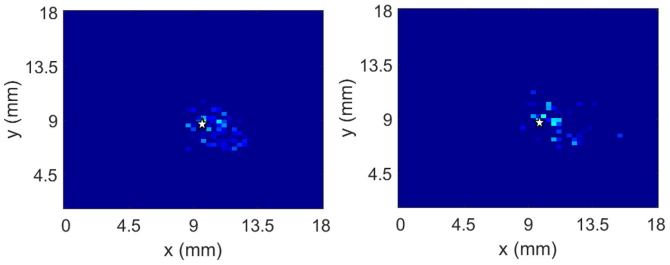
This section details the results of evaluations of AES implementations hardened by the countermeasures described in Section IV and the measurement setup detailed in Section V. For each class of countermeasures, spatial maps of $F_N^{b,pc,t/f}$ and/or $F_B^{b,pc,t/f}$ are shown in Sections VI.A-C. Section VI.D presents the costs of evaluating the counter-measures along with the improvement in resilience. For countermeasures with



(a) Frequency Scaling (FS) countermeasure



(b) Voltage Scaling (VS) countermeasure



(c) Voltage-Frequency Scaling (VFS) countermeasure

Fig. 14. Spatial map of $\max_t F_N^{0,pc,t}$ (left) and $\max_f F_N^{0,pc,f}$ (right) for the FPGA implementing three countermeasures that increase the measurement noise. Optimal configurations are shown with stars.

$mMTD^b > N_e^{max} = 20000$, the cost of the greedy-search protocol is replaced by the cost of the exhaustive scan.

A. Countermeasures Increasing Measurement Noise

The countermeasures FS, VS, and VFS detailed in Section IV.B increase the measurement noise in signals. Because they increase variance within repeated measurements, these counter-measures should degrade $F_N^{b,pc,t/f}$. Spatial maps of the maximum $F_N^{0,pc,t/f}$ are plotted in Fig. 14 for the 3 hardened implementations. The results can be compared to those for the baseline FPGA in Fig. 10; the optimal probe configurations were found to be the same in all cases.

The FS countermeasure could improve the resilience of the module against time-domain EM SCA attacks but had negligible impact on frequency-domain ones. Although shifts in time domain should not impact the magnitude of signals in freq uency domain, delaying/hastening the signal still caused some minor variations in the frequency-domain EM SCA attack; this is because measurements were time-gated to the nominal clock period [20]. The VS countermeasure could i mprove the resilience of the module against both time- and frequency-domain EM SCA attacks, although the impact was more apparent in the frequency-domain approach. Voltage scaling affects the fields disproportionately in time domain, in

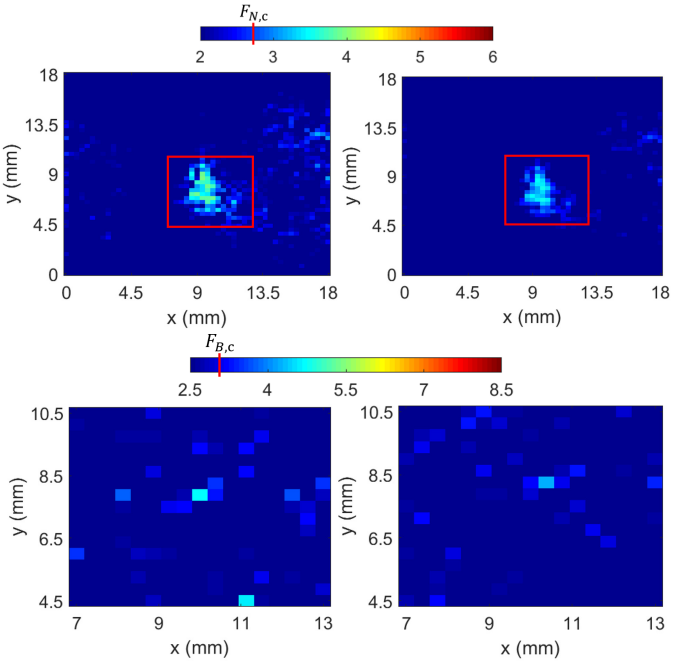


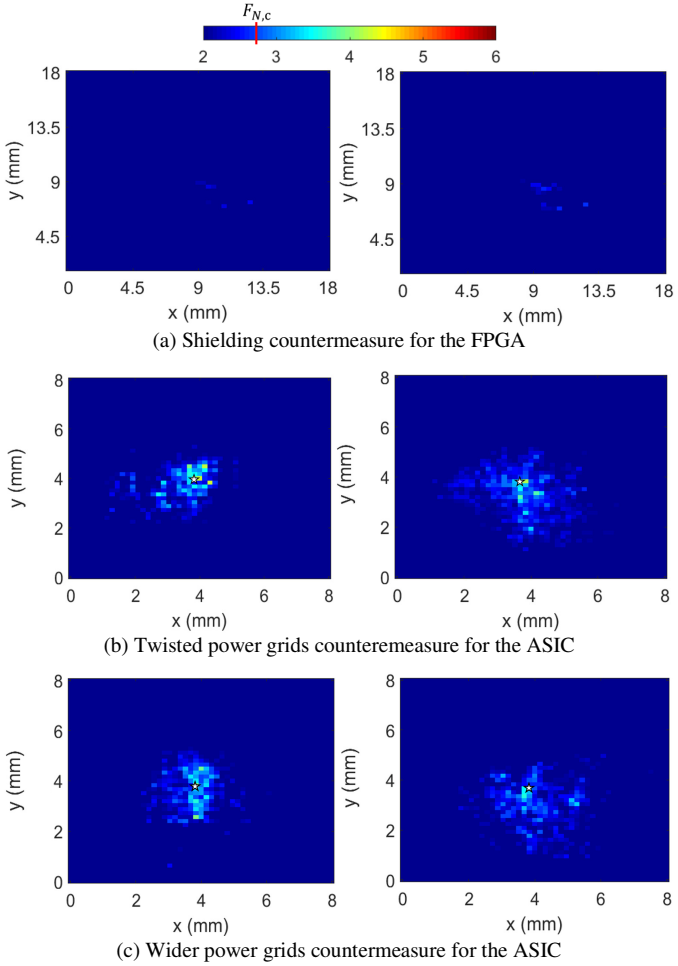
Fig. 15. Spatial map of $\max_t F_N^{0,pc,t}$ (top-left), $\max_f F_N^{0,pc,f}$ (top-right), $\max_t F_B^{0,pc,t}$ (bottom-left), and $\max_f F_B^{0,pc,f}$ (bottom-right) for the FPGA implementing the masking countermeasure that increases algorithmic noise.

particular, more variance was observed around signal peaks, while at other time intervals signals were more repeatable [20]. The VFS countermeasure could improve the resilience of the module against both time- and frequency-domain EM SCA attacks. Because this countermeasure combines the previous two countermeasures, the first two stages of the proposed ANOVA method could identify only a few promising configurations with either time- or frequency-domain signals. The proposed method required ~7.2/5.5×10^6, ~6/5.7×10^6, and ~6.9/7×10^6 measurements to identify the optimal probe configurations for the FPGA hardened with the FS, VS, and VFS countermeasure in time/frequency domain.

B. *Countermeasure Increasing Algorithmic Noise*

The masking countermeasure detailed in Section IV.C increases the algorithmic noise. Because it performs additional uncorrelated computations, this countermeasure should primarily degrade $F_B^{b,pc,t/f}$. Spatial maps of the maximum $F_N^{0,pc,t/f}$ and $F_B^{0,pc,t/f}$ are plotted in Fig. 15. Comparing the results to that for the baseline FPGA in Figs. 10-11 shows that more configurations were eliminated compared to the baseline at the end of Stage I in addition to Stage II, because randomly masking the state register increases signal variance for repeated encryptions as well as increasing algorithmic noise from uncorrelated computations. More importantly, it was observed at the end of Stage III that none of the probe configurations could disclose *any* key byte after $N_e^{\max}$ encryptions.

Unlike the adaptive scan protocols, which would potentially need the same number of measurements as an exhaustive scan (~10^8) to reach this conclusion, the proposed ANOVA method required only ~8.3/7.6×10^6 measurements in time/frequency domain.



(a) Shielding countermeasure for the FPGA

(b) Twisted power grids counteremeasure for the ASIC

(c) Wider power grids countermeasure for the ASIC

Fig. 16. Spatial map of $\max_t F_N^{0,pc,t}$ (left) and $\max_f F_N^{0,pc,f}$ (right) for the three countermeasures attenuating target signals. Optimal configurations, if present, are shown with stars.

C. *Countermeasures Attenuating Target Signals*

The physical design strategies detailed in Section IV.D attenuate the target signals. Because they also reduce the variance of the target signals, these countermeasures should degrade both $F_N^{b,pc,t/f}$ and $F_B^{b,pc,t/f}$. Spatial maps of the maximum $F_N^{0,pc,t/f}$ are plotted in Fig. 16, and can be compared with baseline results in Fig. 10.

The shielded FPGA revealed no configurations of interest at the end of Stage I, failing to disclose the AES key; this is to be expected as the shield is 3-4 skin depths thick at the information leaking frequencies. While the physical design strategies in [18] revealed few configurations of interest, these configurations were successful in recovering the key, providing limited improvement in resilience. The dense wider power-grid structure revealed marginally fewer configurations compared to the twisted power-grid countermeasure.

The proposed method required ~2.7×10^6 measurements using both time- and frequency-domain analysis to evaluate the shielding countermeasure. The evaluation of the twisted power-grid structure, the time-/frequency-domain analysis required acquisition cost of ~6.6/7.3×10^6 measurements. The evaluation of the dense wider power grid structure in time/frequency domain required acquisition cost of ~8/8.1×10^6 measurements.

TABLE II
EFFECTIVENESS OF COUNTERMEASURES AND THE COST OF EVALUATION

DUT	Improvement over Baseline for TD/FD Attack	Most Effective Attack	Acquisition Cost of Alternatives vs. ANOVA for Most Effective Attack		
			Adaptive Scan	TVLA +e	TVLA +i
FPGA Baseline	1×/ 1×	TD	1.8×	4.8×	1.8×
ASIC Baseline	1×/ 1×	TD	2.7×	4.7×	1.6×
FPGA with FS	6.3×/ 1.1×	FD	2.3×	4.8×	1.8×
FPGA with VS	4.7×/ 8.6×	TD	3.0×	6.9×	3.6×
FPGA with VFS	13.6×/ 12.1×	TD	4.3×	6.0×	4.6×
FPGA with Masking*	>30×/ >24×	-	13.1×	5.3×	5.3×
FPGA with Shielding*	>30×/ >24×	-	37.0×	3.6×	3.6×
ASIC with Twisted Power Grid	1.5×/ 1.4×	TD	1.9×	7.7×	4.3×
ASIC with Wider Power Grid	2.4×/ 2.1×	TD	1.7×	7.1×	4.2×

* AES key not disclosed

D. Marginal and Acquisition Cost Comparison

Next, the effectiviness of the countermeasures are evaluated and the costs of the proposed ANOVA method are compared to those of the alternatives when countermeasures are present.

For the baseline FPGA (ASIC), using the optimal configurations identified in Section V, the marginal cost of disclosing keys was only ~$1.1/1.4\times10^4$ (~$3.5/4.4\times10^4$) measurements in time/frequency domain, i.e., disclosing the AES key required ~3× more measurements for the ASIC.

When the FPGA was hardened with the FS, VS, and VFS countermeasures, using the optimal configurations identified in Section VI-A, the attackers could disclose the AES key with ~$6.9/1.5\times10^4$, ~$0.5/1.2\times10^5$, and ~$1.5/1.7\times10^5$ measurements in time/frequency domain, respectively. Comparing these marginal costs to those of the baseline FPGA shows that these countermeasures improve the module's resilience to EM SCA attack significantly. These are easy to implement countermeasures that require relatively small design overhead.

When the FPGA was hardened with masking or shielding, because no key bytes could be disclosed by observing $N_{\mathrm{e}}^{\max}$ encryptions, the marginal cost of disclosing the key was $> 16N_{\mathrm{e}}^{\max}$, i.e., $>3.2\times10^5$ measurements; thus, these countermeasures improve the module's resilience to EM SCA attack by >30/24× in time/frequency domain. Masking considerably improves the security of the chip at the cost of larger area and delay overheads [19]. While a very simplistic shield was used here, practical use of shielding can incur large packaging costs [17]. Additionally, incorrect shielding can block higher-frequency contributions to measurement noise and potentially *reduce* the module's resilience.

When the ASIC was hardened with twisted and dense wider power grid, using the optimal configurations identified in Section VI-C, the attackers could disclose the AES key with ~$5.3/5.9\times10^4$ and ~$8.5/9.2\times10^4$ measurements in time/ frequency domain, respectively. For these physical design strategies, while no logic blocks were added, implying little to no power overhead, layout changes increase the module's area.

The resilience of the 9 AES implementations against fine-grained EM SCA attacks and the costs of this evaluation are shown in Table II. In Table II, the resilience improvement is calculated as the ratio of an implementation's marginal cost over that of the baseline module. The improvement for security evaluation is quantified by dividing the acquisition costs of the alternative methods by that of the proposed method. In each case, both time- and frequency-domain EM SCA attacks were performed but the acquisition costs are compared only for the attack that had the lower marginal cost.

Table II shows that among all countermeasures, masking and shielding countermeasures were most effective in improving the chip's security. In all 9 cases, the ANOVA method required the fewest measurements to evaluate the EM SCA security of the AES implementation. Applying the proposed method was ~1.7-37× cheaper than the adaptive scan protocol, ~3.6-7.7× cheaper than the TVLA followed by exhaustive correlation analysis, and ~1.6-5.3× cheaper than the TVLA-informed correlation analysis. The protocol was particularly efficient when evaluating the most secure implementations.

VII. CONCLUSION

An ANOVA-based measurement method was presented to evaluate fine-grained EM SCA vulnerability of cryptographic modules. The method was used to evaluate 2 baseline and 7 hardened implementations of the AES algorithm against fine-grained EM SCA attacks. The method is implemented in multiple stages; in the first two stages, it eliminates probe configurations posing the lowest risks by estimating the contribution of measurement and algorithmic noise in observed fields, in the last stage it applies correlation-analysis informed by the risk estimates identified in the previous stages to actually reveal the AES key. The method assumes a gold-box threat model and uses specifically chosen inputs and encryption keys in order to evaluate measurement and algorithmic noise with few measurements. The (gold-box) ANOVA method required upto ~37×, ~7.7×, and ~5.3× fewer measurements than the (black-box) greedy-search correlation analysis, the (white-box) TVLA followed by exhaustive correlation analysis, and the (white-box) TVLA-informed correlation analysis, respectively. The proposed method is particularly efficient for evaluating the most secure chips, such as the shielded-FPGA implementation, where it discards ineffective measurement configurations at a relatively low acquisition cost. Thus, it enables rapid empirical evaluation of how effective a countermeasure is for hardening a cryptographic module against fine-grained EM SCA attacks.

The proposed method can be used with alternative methods [40]-[41] in Stage III, if the set of probe configurations can be sufficiently condensed in Stages I and II. The proposed method can also be extended to evaluating other computing systems by suitably modifying definitions of target and background processes; e.g., a related ANOVA method was used in [4] to evaluate the security of a general-purpose embedded system.

ACKNOWLEDGMENT

The authors thank Dr. J. Kulkarni and M. Wang for sharing the hardened ASIC implementations of AES.

REFERENCES

[1] M. Vuagnoux and S. Pasini, "An improved technique to discover compromising electromagnetic emanations," in *Proc. IEEE Int. Symp. Electromagn. Compat.*, pp. 121-126, July 2010.

[2] A. Zajic and M. Prvulovic, "Experimental demonstration of electromagnetic information leakage from modern processor-memory systems," *IEEE Trans. Electromagn. Compat.*, vol. 56, no. 4, pp. 885-893, Aug. 2014.

[3] F. Werner *et al.*, "A method for efficient localization of magnetic field sources excited by execution of instructions in a processor," *IEEE Trans. Electromagn. Compat.,* vol. 60, no. 3, pp. 613-622, June 2018.

[4] V. V. Iyer and A. E. Yılmaz, "Using the ANOVA F-statistic to isolate information-revealing near-field measurement configurations for Embedded Systems," in *Proc. IEEE Int. Symp. Electromagn. Compat.,* Aug. 2021.

[5] Y.-I. Hayashi *et al.*, "Efficient evaluation of EM radiation associated with information leakage from cryptographic devices," *IEEE Trans. Electromagn. Compat.*, vol. 55, no. 3, pp. 555–563, Jun. 2013.

[6] V. V. Iyer and A. E. Yilmaz, "Using the ANOVA F-statistic to rapidly identify near-field vulnerabilities of cryptographic modules," in *Proc. IEEE Int. Microw. Symp.,* June 2021.

[7] L. Sauvage, S. Guilley, and Y. Mathieu, "Electromagnetic radiations of fpgas: High spatial resolution cartography and attack on a cryptographic module," *ACM Trans. Reconfigurable Technol. Syst.*, 2009.

[8] M. Alam *et al*. "One&Done: a single-decryption EM-based attack on OpenSSL's constant-time blinded RSA," in *Proc. USENIX*, pp.585-602, Aug 2018.

[9] V. Iyer, M. Wang, J. Kulkarni, and A. Yilmaz, "A systematic evaluation of EM and power side-channel analysis attacks on AES implementations," in *Proc. IEEE ISI,* Nov. 2021.

[10] V. V. Iyer and A. E. Yilmaz, "An adaptive acquisition approach to localize electromagnetic information leakage from cryptographic modules," in *Proc. IEEE Texas Wireless Symp.,* Mar. 2019.

[11] J. Danial, D. Das, S. Ghosh, A. Raychowdhury, and S. Sen, "SCNIFFER: low-cost, automated, efficient electromagnetic side-channel sniffing," *IEEE Access*, vol. 8, pp. 173414-173427, Sep. 2020.

[12] M. Wang *et al.*, "Galvanically isolated, power and electromagnetic side-channel attack resilient secure AES core with integrated charge pump based power management," in *Proc. IEEE CICC*, Apr. 2021.

[13] D. Das, M. Nath, B. Chatterjee, S. Ghosh, and S. Sen, "STELLAR: a generic EM side-channel attack protection through ground-up root-cause analysis," in *Proc. IEEE HOST*, May 2019.

[14] J.-S. Coron and I. Kizhvatov, *An Efficient Method for Random Delay Generation in Embedded Software*, CHES, Berlin: Springer, 2009, vol. 5747.

[15] C. Sui, J. Wu, Y. Shi, Y. Kim, and M. Choi, "Random dynamic voltage scaling design to enhance security of NCL S-box," in *Proc. IEEE MWSCAS*, Aug. 2011.

[16] A. Singh *et al*., "Improved power/EM side-channel attack resistance of 128-bit AES engines with random fast voltage dithering," *IEEE J. Solid-State Circ*., vol. 54, pp. 569–583, Feb. 2019.

[17] M. Yamaguchi *et al.*, "Development of an on-chip micro shielded-loop probe to evaluate performance of magnetic film to protect a cryptographic LSI from electromagnetic analysis," in *Proc. IEEE Int. Symp. Electromagn. Compat.*, pp. 103–108, Jul. 2010.

[18] M. Wang *et al.*, "Physical design strategies for mitigating fine-grained electromagnetic side channel attacks," in *Proc. IEEE CICC*, Apr. 2021.

[19] G. Li, V. Iyer, and M. Orshansky, "Securing AES against localized EM attacks through spatial randomization of dataflow," in *Proc. IEEE HOST,* May 2019.

[20] V. Iyer, A. Thimmaiah and A. Yilmaz, "Testing the resilience of cryptographic modules against fine-grained time- and frequency-domain EM side-channel analysis attacks," in *Proc. IEEE ICEAA,* Aug 2021.

[21] T. Miyuki and Y. Hayashi, "AES cipher keys suitable for efficient side-channel vulnerability evaluation," Cryptology ePrint Archive, Rep. 2014/770, 2014.

[22] G. Becker, ''Test vector leakage assessment (TVLA) methodology in practice,'' in *Proc. Int. Cryptograph. Module Conf.*, Sep 2013.

[23] C. Whitnall and E. Oswald, "A cautionary note regarding the usage of leakage detection tests in security evaluation", Cryptology ePrint Archive, Rep. 2019/703, 2019.

[24] F. Unterstein *et al.*, "Dissecting leakage resilient prfs with multivariate localized em attacks," in *Proc. COSADE*, Jul. 2017.

[25] NIST FIPS Pub. "197: Advanced encryption standard (aes)". *Federal information processing standards publication*, 197(441):0311, 2001.

[26] M. Nassar, Y. Souissi, S. Guilley, J.-L.Danger, "RSM: a small and fast countermeasure for AES, secure against 1st and 2nd-order zero-offset SCAs," in *Proc. DATE*, Mar. 2012, pp.1173-1178.

[27] M.-L. Akkar and C. Giraud, An Implementation of DES and AES Secure against Some Attacks, CHES, ser. LNCS, LNCS, Ed.,. Springer, May 2001, pp. 309–318.

[28] I. Buhan, L. Batina, Y. Yarom, and P. Schaumont, "SoK: design tools for side-channel-aware implementations." Jun. 2021. Available: *ArXiv* abs/2104.08593.

[29] K. A. Remley *et al.*, "Millimeter-wave modulated-signal and error-cector-magnitude measurement with uncertainty," *IEEE Trans. Microw. Theory Tech.*, vol. 63, no. 5, pp. 1710-1720, May 2015.

[30] K. Freiberger, H. Enzinger, and C. Vogel, "A noise power ratio measurement method for accurate estimation of the error vector magnitude," *IEEE Trans. Microw. Theory Tech.*, vol. 65, no. 5, pp. 1632-1645, May 2017.

[31] B. F. Jamroz *et al.*, "Accurate monte carlo uncertainty analysis for multiple measurements of microwave systems," in *Proc. IEEE MTT-S Int. Microw. Symp.*, Jun. 2016.

[32] C. Fager and K. Andersson, "Improvement of oscilloscope-based RF measurements by statistical averaging techniques," in *Proc. IEEE MTT-S Int. Microw. Symp.,*Jun. 2006.

[33] V. V. Iyer, "An adaptive measurement protocol for fine-grained electromagnetic side-channel analysis of cryptographic modules," M.S. thesis, Univ. of Texas, Austin, Aug. 2019.

[34] J. Heyszl *et al*., *Strengths and Limitations of High-resolution Electromagnetic Field Measurements for Side-channel Analysis*, Lecture Notes in Computer Science, Berlin, Germany: Springer, 2012, vol. 7771.

[35] ChipWhisperer, Github Repository [online], available: https://github.com/newaetech/chipwhisperer

[36] Aoki laboratory, Tohoku University, Japan [online], available: http://www.aoki.ecei.tohoku.ac.jp/crypto/web/cores.html

[37] D. Fujimoto et al., "On-chip power noise measurements of cryptographic VLSI circuits and interpretation for side-channel analysis," in *Proc. IEEE Int. Symp. on Electromagn. Compat.*, pp. 405–410, Sep. 2013.

[38] E. Peeters, F.X. Standaert, J.J. and Quisquater, "Power and electromagnetic analysis: Improved model, consequences and comparisons". *Integr. VLSI J.*, vol. 40, pp.52-60, Jan. 2007.

[39] A. Kumar, C. Scarborough, A. E. Yilmaz, and M. Orshansky, "Efficient simulation of EM side-channel attack resilience," *in Proc. ICCAD*, pp. 123 – 130, Nov. 2017.

[40] R. Gilmore, N. Hanley, and M. O'Neill, "Neural network based attack on a masked implementation of AES," *in Proc. ICCAD*, pp. 106 – 111, Nov. 2015.

[41] T. Kubota *et al.*, "Deep learning side-channel attack against hardware implementations of AES," *Microprocessors and Microsystems*, vol. 87, Nov. 2021.

[42] S. Mangard, E.Oswald, and T. Popp, *Power analysis attacks: Revealing the secrets of smart cards.* Springer Science & Business Media, Vol. 31. 2008.

[43] Information Technology Standard, "Security techniques – testing methods for the mitigation of non-invasive attack classes against cryptographic modules," *International Organization for Standardization*, Geneva, CH, 2016.

[44] C. Whitnall and E. Oswald, "A critical analysis of ISO 17825 ('Testing Methods for the mitigation of non-invasive attack classes against cryptographic modules')," in *Proc. Int. Conf. Theory Appl. Cryptol. Inf. Secur.*, Dec. 2019, pp. 256–284.