Switch Capacitor-Based Time-Varying Transfer Function for FCN and CNN MLSCA-Resistant AES256 in 65-nm CMOS

Archisman Ghosh, *Graduate Student Member, IEEE*, Debayan Das, *Member, IEEE*, Santosh Ghosh, and Shreyas Sen, *Senior Member, IEEE*

***Abstract*—Mathematically secure cryptographic implementations can leak critical information through physical side channels. Machine learning (ML) has facilitated efficient side-channel analysis (SCA), especially on small IoT devices and smart cards. We propose a lightweight, synthesizable technique to enhance ML-based SCA resilience. Our approach introduces a physical time variance technique that specifically targets Deep Neural Network based MLSCA. This brief presents a physical time variance technique that is effective against CNN contrary to the previous state-of-the-art. By eliminating analog units and utilizing a switched capacitor design, it outperforms existing techniques by 5× in terms of traces to train the attacking neural network.**

***Index Terms*—Side channel attack, countermeasures, AES256, time-varying transfer function, generic, low-overhead, synthesizable, MLSCA.**

I. Introduction

COMPUTATIONALLY secure modern encryption engines can be compromised through side-channel attacks, such as power traces [1], EM emanations [2], and timing information [3]. Deep learning techniques have been used to analyze these side channels and recover secret keys with just a few power traces in recent times [4] increasing the threat space significantly for IoT devices which need not operate for a longer time. This brief introduces a low-overhead countermeasure using time-varying power-supply transfer functions implemented with digital-friendly switch capacitor circuits, effectively protecting against different deep neural network-based attacks.

A. Motivation & Related Works

Fig. 1(a) shows the history of switch capacitor (SC) based and time-domain SCA countermeasures. SC-based SCA countermeasure was first proposed by Tokunaga and Blaauw [5] with $> 10M$ minimum traces-to-disclosure (MTD). However, a separate bias is required to clear the residue, which made it analog-like and not easily scalable across technology nodes. DNN-based SCA has two phases, namely a) the training phase, and b) the attack phase. A DNN model is trained using power traces in the training phase as shown in Fig. 1(b). The trained model is used to attack directly in the attack phase. As DNN is already trained it takes a very less number of traces ($M < 10$) reducing attack time significantly, thus increasing threats for smart cards, and IoT devices in a practical scenario. Earlier, time shifting/wrapping-based countermeasures are [6], [7] explored primarily through non-silicon implementations, moved each trace by only a few time samples in a deterministic way (Fig. 1(a)). They have been rendered ineffective against CNN-based SCA. Frequency scaling [7] can also be defeated by analyzing the time-domain power-supply waveform alignment-based attack or frequency domain attacks. Series LDO [8], [9], IVR [10] based countermeasures, proposed recently, show moderate defense against CPA/CEMA (up to 10M MTD) and are not evaluated against MLSCA. Current domain signature attenuation (CDSA) [11], [12], additive masking based solutions [13], and randomized LDO with arithmetic countermeasure [14] have been proven to be protected against deep-learning (DL) SCA attacks. Ghosh et al. proposed a solution of digital signature attenuation which brings the benefit of analog signature attenuation in the digital domain [15], [16], [17]. A memristor-based DPA-resistant AES implementation is explored in [18] using SPICE simulations (not validated in silicon). Recently, ML-based solution is used for side channel resilience [19], which is not evaluated against ML-based SCA strategies.

In this brief, we advance the state-of-the-art of SC countermeasures for SCA making it digital-friendly by removing the residue-reset requirement from [5], combining time-variance into the power supply network through SCs that dynamically changes the transfer function between the local AES supply and observable supply. We demonstrate a switch capacitor-based time-varying transfer function (SC-TVTF) countermeasure resilient to CNN/FCN-based attacks. SC-TVTF controller is fully synthesizable, and switches are synthesizable in automatic place & route(APR) flows which include power gates. Capacitors are placed in layout using DCAP cells by industry-standard APR tools. Synthesizable SC shows slightly worse results against Correlational Power Attack than analog counterpart [16]. However, lower capacitance leads to low area overhead in this technique. Moreover, minimal voltage droop causes no performance overhead. Power comparison with respect to analog counterpart shows our solution is minimum power overhead. Also, analog switch capacitor-based solution is not tested against modern ML-based side-channel attacks. Less time-consuming (theoretically possible in μs)

Manuscript received 2 June 2023; accepted 27 July 2023. Date of publication 4 August 2023; date of current version 8 January 2024. This work was supported in part by NSF under Grant CNS 17-19235, and in part by Intel Corporation. This brief was recommended by Associate Editor R. Venkatesan. *(Corresponding author: Shreyas Sen.)*

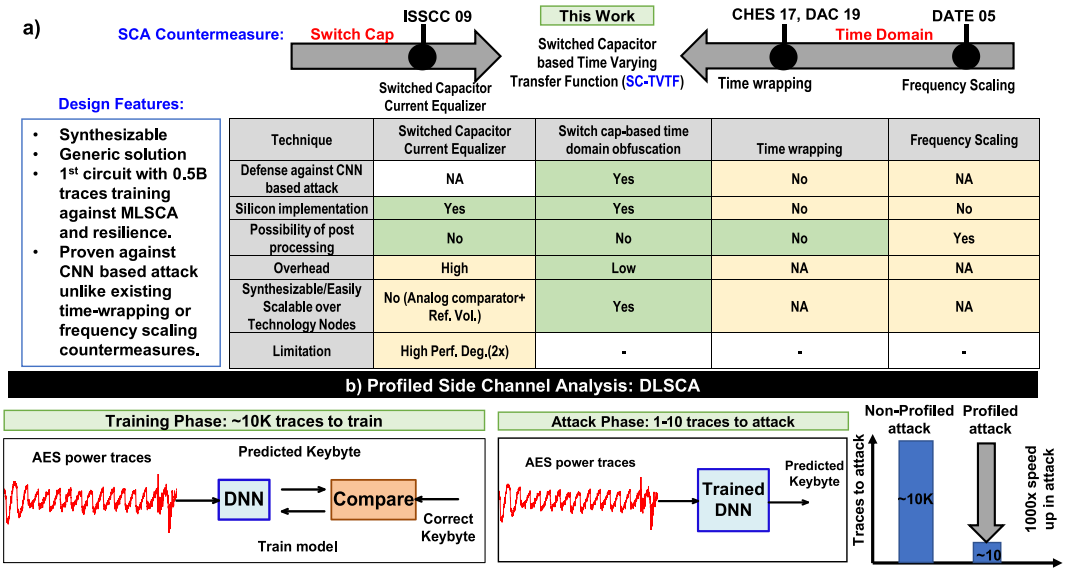
Archisman Ghosh and Shreyas Sen are with the Department of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47097 USA (e-mail: shreyas@purdue.edu).

Debayan Das is with the Electronics Systems Engineering, Indian Institute of Science, Bengaluru 560012, India.

Santosh Ghosh is with the Security and Privacy Research Department, Intel Corporation, Hillsboro, OR 97124 USA.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCSII.2023.3302258.

Digital Object Identifier 10.1109/TCSII.2023.3302258



Technique	Switched Capacitor Current Equalizer	Switch cap-based time domain obfuscation	Time wrapping	Frequency Scaling
Defense against CNN based attack	NA	Yes	No	NA
Silicon implementation	Yes	Yes	No	No
Possibility of post processing	No	No	No	Yes
Overhead	High	Low	NA	NA
Synthesizable/Easily Scalable over Technology Nodes	No (Analog comparator+ Ref. Vol.)	Yes	NA	NA
Limitation	High Perf. Deg.(2x)	-	-	-

Fig. 1. a) State-of-the-art switched cap-based and time-domain countermeasures are shown. This brief extends and merges both concepts to achieve high resiliency against ML-based power SCA. b) Efficacy of Deep Learning-based SCA: traces to perform a successful attack can be reduced by $> 1000\times$ by proper profiling.

M-trace attack [4] is a significant threat against IoT devices. Each time sample captured by the oscilloscope acts as a feature for DNN, scrambling the power traces randomly using the SC-TVTF network makes the feature set different from the training sets as different sampling points are scrambled in time-domain. Hence, it becomes practically harder for DNN to attack in real time providing SCA security.

B. Contributions

This brief has four-fold contributions:

- This 65nm TSMC CMOS IC for the first time demonstrates a time-domain MLSCA resilient technique.
- This brief advances the state-of-the-art of switched capacitor-based solutions by bringing the technique into a completely digital domain and removing the need for external analog bias voltage.
- This brief proposes a generic time-varying transfer function that provides resilience against CNN-based MLSCA, thus advancing time-domain countermeasures.
- Finally, this implementation is a low-overhead, completely synthesizable solution dedicated to IoT devices and smart cards.

Scrambling power traces in the time domain lacks mathematical security but lowers SNR for physical-layer security. Traditional attacks can't breach it, but reverse-engineering/manipulating LFSRs allows for vulnerability. Addressing this open research problem, we propose using verified TRNGs to generate a seed and prevent such theoretical attacks.

II. Circuit Techniques & System Architecture

Fig. 2 shows full SC-TVTF architecture consisting of N (=16) unit switched capacitors (~20pF) and a lightweight SC-TVTF controller. 1 out of N capacitors is randomly chosen to supply the crypto core (parallel AES256) ('Discharge'), while another one is chosen to be charged from the supply ('Charge'), and the rest N-2 (=14) capacitors remain disengaged ('Store'). SC-TVTF controller ensures

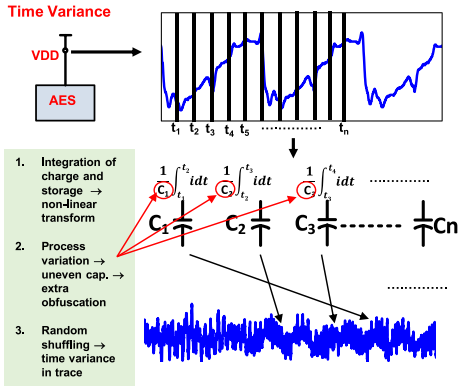


Fig. 2. Working principle of SC-TVTF.

Discharge→Store→Charge→Store→Discharge sequence for each capacitor while ensuring randomness. This architecture allows the transfer of integrated power traces during discharging to random locations on the observable power trace creating a TVTF and removing the need for residue-reset from [5]. The randomness is derived from nested LFSRs, increasing the periodicity of randomization, seeded by external TRNGs. SC-TVTF effectively creates physical time-domain information shuffling, further aided by inherent capacitance mismatch due to intra-chip process variations. Unlike traditional high overhead algorithmic shuffling, SC-TVTF achieves a similar effect by modifying the switch-capacitor power supply network, which has not been explored in prior literature.

The full system architecture is presented in Fig. 3. V_{AES} node is directly connected to V_{DD} node in unprotected mode. However, switch capacitor networks are enabled in protected mode and the power trace is shuffled in the time-domain using an intelligent TVTF controller. The entire capacitor network has a total of 320pF on-chip capacitors. SC-TVTF controller (Fig. 4(a)) consists of 4b and 16b Fibonacci LFSRs for randomization, two 8×4 memories, and two decoders. The working of the SC-TVTF controller is demonstrated using a timing diagram in Fig. 4(b). Randomly generated 4-bit numbers are used to select an address in memory0 and

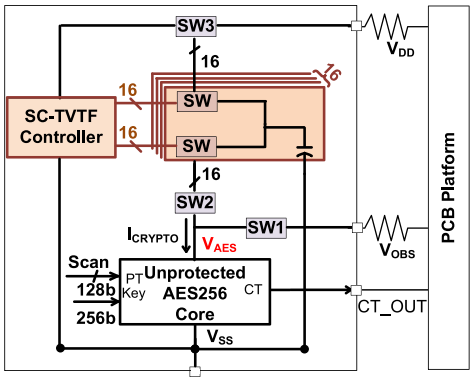


Fig. 3. Full system architecture of SC-TVTF-AES256.

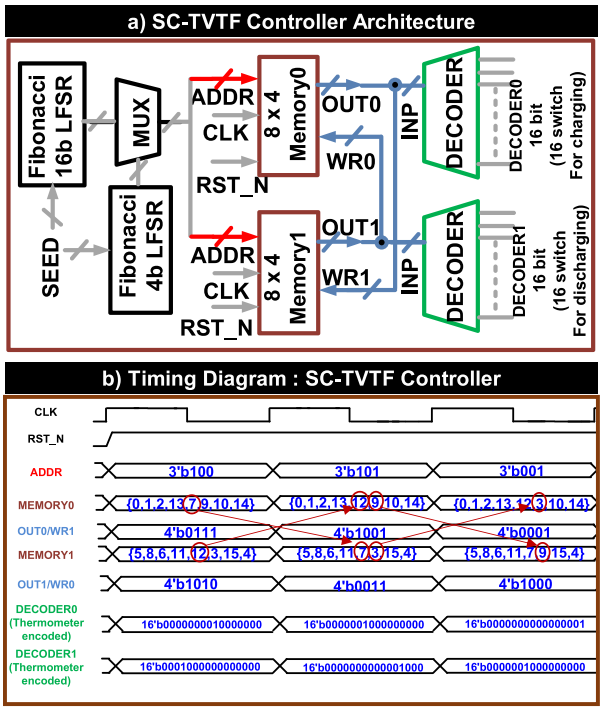
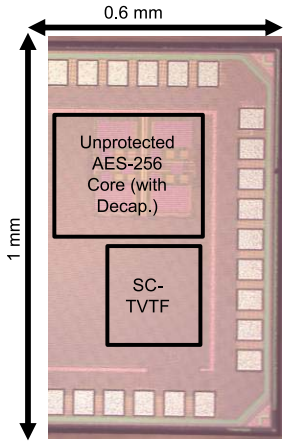


Fig. 4. a) SC-TVTF controller architecture. b) Example timing diagram of SC-TVTF.

memory1. Each address contains a number that acts as a tag for the switched capacitors. Based on the tag, capacitors to be charged/discharged are selected by a decoder. Once selected for charging or discharging, those tags are shifted to alternate memory ensuring no performance degradation unless TVTF operating frequency is sufficiently low. For example, the randomly selected address is 4 as shown in Fig. 4(b) at the beginning, hence 7^{th} and 12^{th} capacitors have been selected for charging and discharging respectively as 7 and 12 are the selected tag at the address 4. Those capacitor tags swapped places between memory0 and memory1 in the next cycle. In the next 2 cycles, address 5 and 0 are randomly selected continuing a similar phenomenon in the SC network.

III. Measurement Results

This 65 nm test chip, being the first of its kind (switch capacitor-based physical time domain obfuscation), shows promising results against MLSCA attacks. Die micrograph and IC specifications are shown in Fig. 5. This IC is taped-out with TSMC 65nm CMOS LP technology and packaged using



Test Chip	Process	65nm CMOS LP
	Package	Wire-bond (Glob-Top Encapsulation)
Unprotected AES-256 Core	Performance	0.8V/10MHz/189uW
	Architecture	128b Datapath (parallel AES)
	Supply Decap	60 pF MOS Cap (0.006 mm²)
	Active Area	0.15 mm²
	Scan Chain	387b
Protected-AES256	Performance	0.8V/10MHz/246uW
	Active Area	0.18 mm²
	Supply Decap	60 pF MOS Cap (0.006 mm²)
	Switched Cap	320pF Switched Cap (0.03 mm²)
	Scan Chain	AES Core: 387b
		Protected core: 143b

Fig. 5. IC micrograph and design specification.

simple wire-bond and globe-top encapsulation. It has a parallel AES256 implementation with a 128b datapath. The active area of the AES256 is $0.15mm^2$. Composite field-based sbox is used earlier as an arithmetic countermeasure against power and EM side channel attack [14] and fault injection attack [20], [21], [22] as well as combined attacks [23], [24]. This type of sbox often leads to low area implementation as well [25]. However, this brief shows the efficacy of time-variance-based physical countermeasure against ML-based power side-channel attack. Hence, this countermeasure can be used with any other arithmetic countermeasure. As our goal is to see relative improvement due to time-domain obfuscation at VDD with respect to unprotected, we use a LUT-based implementation of sbox. Note that, the combined attack is dependent on power traces for correct and faulty ciphertexts of the same plaintext [26]. Due to time domain obfuscation, traces are misaligned hence, any correlational attack is harder to accomplish. However, a thorough analysis will be conducted as part of future work. AES256 is operated at 10MHz for all the SCA experiments and it consumes 189uW power while being operated at 0.8V V_{DD}. The active area of the countermeasure is $0.03mm^2$. Scan chain interface is used for configuring both AES and the countermeasure. Protected mode consumes 248uW power at 10MHz operational frequency and 0.8V V_{DD}. It should be noted that [15] uses similar technique in addition to digital signature attenuation to achieve >1.25B minimum traces to disclosure (MTD). However, TVTF as a standalone countermeasure is first time tested against machine learning based side channel attack here. We conclude that with just time-domain obfuscation, our countermeasure has very high resistance against FCN based or CNN based ML assisted side channel attack. Digital signature attenuation is the main technique explored in [15], [17] where [17] is taped-out in different die. Moreover [17] does not include any time domain obfuscation techniques.

The 1st key byte is targeted for the DNN-based attack. During the training/profiling phase, a fixed plaintext is fed to the AES256 and power traces for all the possible 256 combinations for the 1st key byte is captured using a 5Gsps oscilloscope and fed to the DNN for training. During the testing/attack phase, the trained DNN is then used to recover the 1st correct key byte from the target device (AES256) under attack. The final architectures of the FCN and CNN are shown in Fig. 6.

It should be noted that increasing the number of layers in deep neural network does not necessarily increase or keep the performance constant at the highest level due to overfitting

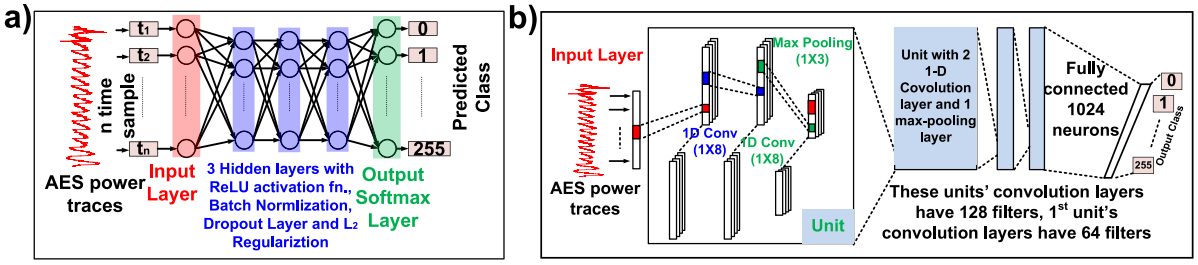


Fig. 6. a) The architecture of FCN used for attack. b) Architecture of CNN used for attack.

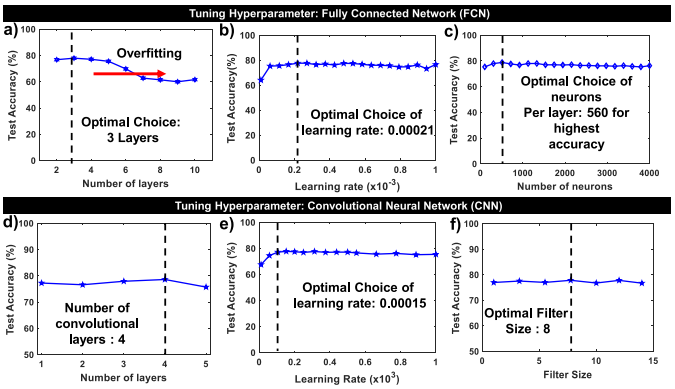


Fig. 7. Tuning hyperparameter for FCN and CNN.

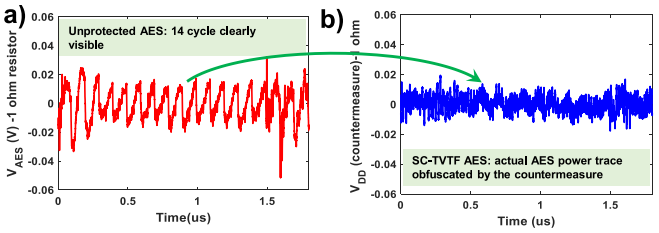


Fig. 8. a) Time domain trace for unprotected AES256 encryption. b) Time domain trace for TVTF-AES256.

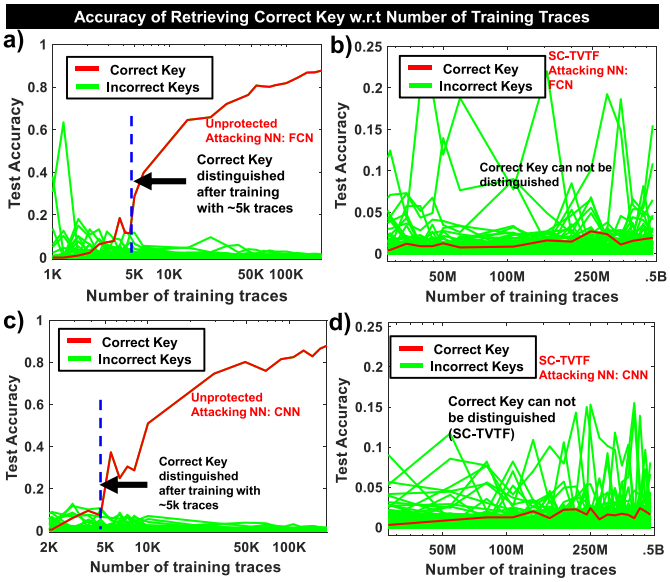


Fig. 9. FCN-based MLSCA results on a) unprotected b) TVTF-AES256. CNN-based MLSCA results on c) unprotected d) TVTF-AES256. Y-axis represents the accuracy of retrieving the correct key and X-axis represents number of traces to train the attacking neural network. Both FCN and CNN-based attacking neural network can retrieve the correct key after being trained with just 5K traces. Both neural networks could not be trained with 500M traces to detect the correct key when the countermeasure is on.

Parameter		This Work	ISSCC'22 [19]	ISSCC'22 [13]	VLSI'20 [14]	ISSCC'20/ CICC'20 [11, 12]	ISSCC'19 [8]	ISSCC '09 [5]
Countermeasure Technique		Switch Cap-based Time Varying Transfer Function	Run-time Machine-Learning	Random Additive masking + Address randomization	NL-DLDO + Arithmetic Solution	Current Domain Signature Attenuation	Randomized Digital LDO	Switched Capacitor Current Equalizer
Attack Technique		ML Attack	CPA	CPA	ML Attack	ML Attack	CPA	DPA
Process		65nm CMOS	40nm CMOS	7nm CMOS	14nm CMOS	65nm CMOS	130nm CMOS	130nm CMOS
Crypto Algorithm		AES-256	AES/ PRESENT	AES-128/256	AES-128	AES-256	AES-128	AES-128
Design Overheads	Area	20% [d]	93%	120%	8%[c]	36.70%	36.9%[b]	33%
	Power	25%	8.6%	120%	10%[c]	49.80%	32%	20%
	Perf.	0%	0%	8%	0.70%	0%	10.40%	50%
MLSCA Analysis	Traces to Train	>0.5B	NA	NA	>100M	>10M	NA	NA

[a]Does not include regulator area/power. [b]Does not include Cap area. [c]Does not include DLDO area/power. Area overhead >150% with DLDO (estimated). [d] Area overhead includes all the extra components without unprotected and calculated with respect to unprotected AES-256.

Fig. 10. Comparison table with respect to state-of-the-art.

which is clearly visible in Fig. 7(a,d). Adding more FCN layers can increase the capacity of the network, allowing it to learn more complex representations. However, too large or too deep model leads to overfitting, where the model memorizes the training data excessively and fails to generalize well to unseen data. Power traces are one-dimensional and simpler datasets with respect to other ML problems such as object detection. Hence, in general, even more than 3 hidden layers lead to overfitting. A lower learning rate leads to a very high training time and epoch. This implies if a low learning rate is used that can lead to improper training and hence attack model does not work the best. To solve this in our final attack model, we start with a high learning rate, then in case of validation accuracy is stuck in a plateau, we reduce the learning rate to reach the optimum. This adaptive approach is well-granted in the machine learning community [27] and provides the optimum results. We use google colab as the ML platform for attacking. A100 gpu (free for use) is used. Data can be processed in 5min 48 seconds with 3 hidden layers. 2 hidden layers only improve 1 minute in terms of time consumed. As the dataset is smaller and 1 dimensional, we do not see much improvement in terms of time/energy consumed. This is the reason behind choosing 3 hidden layers which provide us slightly better performance with respect to 2 hidden layers. Many machine learning problems (e.g., object detection) require to process multidimensional data with a large dataset. Reduction of layer is intuitive in those cases trading off slight inaccuracy due to extremely high training time. Initially, we tried fixing all layers to same parameter and search for the best hyperparameter to achieve the highest accuracy. We observe that 560 neurons provide the highest accuracy as shown in Fig. 7. Now, we run a sweep of different numbers of neurons in different layers and it is observed that

accuracy does not improve much. It is difficult to optimize a perfect set of neurons/layer as we have 3 hidden layers. It is a multi-dimensional optimization problem. We tried to fix the number of neurons of 1st 2 layers and number of neuron is varied in 3rd hidden layer. It is observed that 512, 300, and 860 neurons in consecutive layers lead to similar accuracy. Fig. 8 presents a sample time-domain measurement of unprotected and TVTF design respectively across 1Ω series resistance in the power supply. It is observed that with 200K unprotected traces, both the networks can be trained to achieve 87% accuracy as shown in Fig. 9(a,c). Training with 5K traces is sufficient to recover the correct key during the test/attack phase with high confidence from the AES256, with as low as 5 traces utilizing an M-trace attack using the concept of majority voting [4]. In the protected implementation, the FCN and CNN models could not be trained even after 0.5B traces with similar architecture(Fig. 9(b,d)). Each of 1M traces have been averaged 100 times to get rid of unnecessary measurement noise while training. A similar averaging technique is introduced while attacking as well however, there is no success in such an attack that ensures the efficacy of low-overhead SC-TVTF countermeasure against DNN-based MLSCA. A comparison with previous relevant works has been presented in the table of Fig. 10. This design incurs 20% area overhead and 25% power overhead (considering all the extra components), which is the lowest to date.

IV. Conclusion

Finally, SC-TVTF is an efficient low-overhead technique that provides high MLSCA resilience ($> 100,000\times$ improvement in the number of traces to train compared to unprotected AES and $> 5\times$ compared to state-of-the-art [14]) against advanced DNN-based SCA attacks. NIST has selected Kyber and Dilithium as the new standard for Post Quantum public key cryptography and signature scheme respectively. NIST has also selected ASCON as the new standard for lightweight cryptography (LWC). It is important to note that the proposed countermeasure is a generic one and can be used with any of the crypto cores. For example, Kyber 90's uses AES as pseudo-random number generator. This countermeasure being a generic countermeasure can not only be used in AES of Kyber but also can be used for Kyber in general. This will be done as part of future tape-out and can be evaluated against standard and ML-based attacks as this technique makes the countermeasure completely generic and scalable.

References

[1] P. Kocher et al. "Differential power analysis," in *Proc. CRYPTO*, 1999, pp. 388–397.

[2] P. Kocher, J. Jaffe, B. Jun, and P. Rohatgi "Introduction to differential power analysis," *J. Cryptograph. Eng.*, vol. 1, no. 1, pp. 5–27, Mar. 2011.

[3] P. C. Kocher, "Timing attacks on implementations of Diffie-Hellman, RSA, DSS, and other systems," in *Proc. CRYPTO*, 1996, pp. 104–113.

[4] D. Das, "X-DeepSCA: Cross-device deep learning side channel attack," in *Proc. IEEE/ACM DAC*, 2019, pp. 1–6.

[5] C. Tokunaga and D. T. Blaauw, "Secure AES engine with a local switched-capacitor current equalizer," in *Proc. IEEE ISSCC*, 2009, pp. 64–65.

[6] E. Cagli, C. Dumas, and E. Prouff, "Convolutional neural networks with data augmentation against jitter-based countermeasures," in *Proc. CHES*, 2017, pp. 45–68.

[7] S. Yang, W. H. Wolf, N. Vijaykrishnan, D. N. Serpanos, and Y. Xie, "Power attack resistant cryptosystem design: A dynamic voltage and frequency switching approach," in *Proc. Design Autom. Test Europe*, 2005, pp. 64–69.

[8] A. Singh, M. Kar, S. Mathew, A. Rajan, V. De, and S. Mukhopadhyay, "25.3 A 128b AES engine with higher resistance to power and electromagnetic side-channel attacks enabled by a security-aware integrated all-digital low-dropout regulator," in *Proc. IEEE ISSCC*, pp. 404–406.

[9] S.-H. Cheng, M.-H. Lee, B.-C. Wu, and T.-T. Liu, "A lightweight power side-channel attack protection technique with minimized overheads using on-demand current equalizer," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 69, no. 10, pp. 4008–4012, Oct. 2022.

[10] M. Kar, A. Singh, S. Mathew, A. Rajan, V. De, and S. Mukhopadhyay, "8.1 improved power-side-channel-attack resistance of an AES-128 core via a security-aware integrated buck voltage regulator," in *Proc. IEEE ISSCC*, 2017, pp. 142–143.

[11] D. Das et al. "27.3 EM and power SCA-resilient AES-256 in 65nm CMOS through 350× current-domain signature attenuation," in *Proc. IEEE ISSCC*, 2020, pp. 424–426.

[12] D. Das, J. Danial, A. Golder, S. Ghosh, A. R. Wdhury, and S. Sen, "Deep learning side-channel attack resilient AES-256 using current domain signature attenuation in 65nm CMOS," in *Proc. IEEE Custom. Integr. Circuits Conf. (CICC)*, 2020, pp. 1–4.

[13] R. Kumar et al., "An 8.3-to-18gbps reconfigurable SCA-resistant/dual-core/blind-bulk AES engine in intel 4 CMOS," in *Proc. IEEE ISSCC*, vol. 65, 2022, pp. 1–3.

[14] R. Kumar et al., "A time-/frequency-domain side-channel attack resistant AES-128 and RSA-4K crypto-processor in 14-nm CMOS," *IEEE J. Solid-State Circuits*, vol. 56, no. 4, pp. 1141–1151, Apr. 2021.

[15] A. Ghosh, D. Das, J. Danial, V. De, S. Ghosh, and S. Sen, "36.2 an EM/power SCA-resilient AES-256 with synthesizable signature attenuation using digital-friendly current source and ro-bleed-based integrated local feedback and global switched-mode control," in *Proc. IEEE ISSCC*, vol. 64, 2021, pp. 499–501.

[16] A. Ghosh, D. Das, J. Danial, V. De, S. Ghosh, and S. Sen, "Syn-stellar: An EM/power SCA-resilient AES-256 with synthesis-friendly signature attenuation," *IEEE J. Solid-State Circuits*, vol. 57, no. 1, pp. 167–181, Jan. 2022.

[17] A. Ghosh, D.-H. Seo, D. Das, S. Ghosh, and S. Sen, "A digital cascoded signature attenuation countermeasure with intelligent malicious voltage drop attack detector for EM/power SCA resilient parallel AES-256," in *Proc. IEEE Custom Integr. Circuits Conf. (CICC)*, 2022, pp. 1–2.

[18] M. Masoumi, "Novel hybrid CMOS/memristor implementation of the AES algorithm robust against differential power analysis attack," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 67, no. 7, pp. 1314–1318, Jul. 2020.

[19] Q. Fang, L. Lin, Y. Zu Wong, H. Zhang, and M. Alioto, "Side-channel attack counteraction via machine learning-targeted power compensation for post-silicon HW security patching," in *Proc. IEEE ISSCC*, vol. 65, 2022, pp. 1–3.

[20] M. Mozaffari-Kermani and A. Reyhani-Masoleh, "Concurrent structure-independent fault detection schemes for the advanced encryption standard," *IEEE Trans. Comput.*, vol. 59, no. 5, pp. 608–622, May 2010.

[21] M. Mozaffari-Kermani and A. Reyhani-Masoleh, "A low-power high-performance concurrent fault detection approach for the composite field s-box and inverse s-box," *IEEE Trans. Comput.*, vol. 60, no. 9, pp. 1327–1340, Sep. 2011.

[22] M. Mozaffari-Kermani and A. Reyhani-Masoleh, "A lightweight high-performance fault detection scheme for the advanced encryption standard using composite fields," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 19, no. 1, pp. 85–91, Jan. 2011.

[23] J. Dofe, H. Pahlevanzadeh, and Q. Yu, "A comprehensive FPGA-based assessment on fault-resistant AES against correlation power analysis attack," *J. Electron. Test.*, vol. 32, pp. 611–624, Oct. 2016.

[24] T. Schneider, A. Moradi, and T. Güneysu, "Parti–towards combined hardware countermeasures against side-channel and fault-injection attacks," in *Proc. CRYPTO*, 2016, pp. 302–332.

[25] D. Canright, "A very compact s-box for aes," in *Proc. CHES*, vol. 3659, 2005, pp. 441–455.

[26] Christophe Clavier, Benoît Feix, Georges Gagnerot, and Mylène Roussellet. "Passive and active combined attacks on AES combining fault attacks and side channel analysis," *Proc. Fault Diagn. Toleran. Cryptogr.*, 2007, pp. 10–19.

[27] L. N. Smith, "Cyclical learning rates for training neural networks," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, 2017, pp. 464–472.