Deep Learning Side-Channel Attack Resilient AES-256 using Current Domain Signature Attenuation in 65nm CMOS

Debayan Das[1], Josef Danial[1], Anupam Golder[2], Santosh Ghosh[3], Arijit Raychowdhury[2], Shreyas Sen[1]
[1]Purdue University, IN, USA; [2]Georgia Institute of Technology, GA, USA; [3]Intel Labs, OR, USA

***Abstract*—This article, for the first time, demonstrates an efficient circuit-level countermeasure to prevent deep-learning based side-channel analysis (DLSCA) attacks on encryption devices. Machine learning (ML) SCA, particularly DLSCA attacks have been shown to be extremely effective as it can potentially reveal the secret key of the cryptographic device with as low as a single trace, by offloading the heavy-lifting on the profiling phase where the model learns the correlated leakage patterns of the key. This work presents a current-domain signature attenuation (CDSA) hardware embedding an AES256 engine fabricated in 65nm CMOS technology to suppress the current signature by >350× before it reaches the power supply pin accessible to an attacker. Measurement results show that a 256-class deep neural network (DNN) model for DLSCA attack can be fully trained (>99.9% test accuracy) using only <5K power traces from the unprotected AES256, while the DNN model for the protected CDSA-AES256 cannot be trained even with 10M traces.**

Keywords—Current Domain Signature Attenuation; Machine Learning Power Side-Channel Attack; Security; Deep Neural Network; Countermeasure; Cryptography.

I. Introduction

Cryptographic algorithms are integral to today's internet-connected devices to provide security and integrity of data. Although these algorithms cannot be broken using brute-force cryptanalytic attacks, they are implemented on a physical platform which leak critical information in the form of power consumption, electromagnetic radiation, timing, and so on. This work focuses on the power SCA attacks, specifically profiling attacks, and demonstrates a physical countermeasure to prevent deep-learning based power SCA.

Non-profiled attacks include differential and correlational power analysis (DPA/CPA) which directly attack a target device utilizing statistical correlation, while profiling SCA attacks comprise of building an offline template (model) using an identical device and the attack is performed on a similar device with much fewer traces [1], [2].

A. Motivation

DLSCA utilizes a DNN model for each key byte (of AES256) by training it on traces collected by varying the key byte [1]. As shown in Fig. 1(a), power traces for profiling (training) the 256-class DNN are captured from the test chip running AES256 (protected/unprotected mode) with a fixed plaintext (PT) and varying the 1st key byte and labeling each trace with the corresponding key byte value. During the DLSCA attack phase, unseen traces (for the same PT used in training) are fed to the trained DNN to predict the correct key byte. Fig. 1(b, c) shows an overview of the DLSCA, where most of the heavy-lifting is done during the training phase as it takes few thousands of traces to train the DNN, while an attack becomes feasible in a single trace, increasing the threat surface significantly [2].

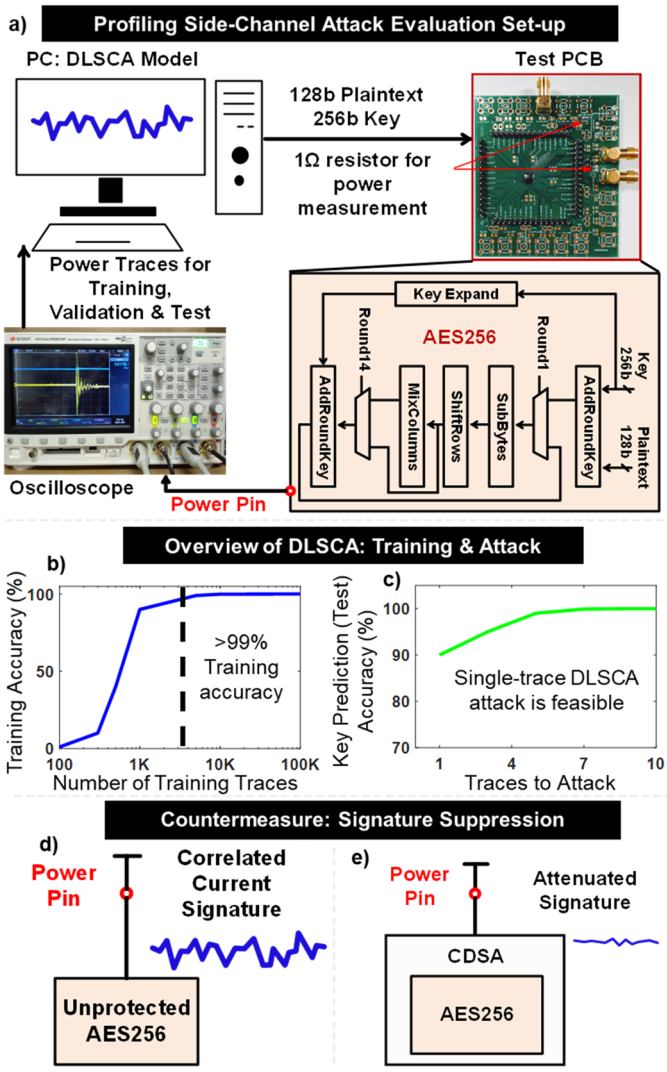


Figure 1: (a) DL SCA attack set-up on the AES256 with the 65nm test chip; (b, c) Overview of DLSCA attack showing the feasibility of a single-trace attack [2] on an unprotected crypto implementation. (d, e) Concept of the CDSA hardware.

Fig. 1(d, e) shows an overview of the proposed CDSA circuit, which suppresses the correlated current signature significantly, motivated by the fact that the minimum traces to disclosure (MTD) is inversely proportional to the square of the

This work was supported in part by NSF under Grant CNS 17-19235, CNS 16-24731 (CAEML), SRC (Grant 2720.001) and Intel Corporation.

978-1-7281-6031-3/20/$31.00 ©2020 IEEE

signal to noise ratio (SNR): $MTD \propto \frac{1}{SNR^2}$ [3]. By reducing the SNR drastically, CDSA ensures that the ML model does not learn the leakage pattern within a reasonable number of traces.

B. Contribution

The key contributions of this work are:

- This work, for the first time, utilizes CDSA hardware involving a high output impedance current source (CS) embedding a crypto engine to demonstrate high DLSCA resilience, by providing >350× signature attenuation in 65nm CMOS.
- DLSCA attack is demonstrated on an unprotected AES256 engine using only <5K measured power traces to train the 256-class DNN.
- Measured results from the CDSA-AES show that the DNN could not be trained even with 10M traces, thwarting DLSCA attack. Moreover, it is a generic low area/power overhead SCA countermeasure and can be extended to any crypto algorithm without any performance degradation.

II. Background & Related Work

Existing logical and architectural countermeasures involving time-domain or clock-jitter based obfuscations have been shown to be defeated using convolutional neural network (CNN) which learns the side-channel leakage even in presence of trace misalignments [4]. Also recently, masking-based countermeasures have been shown to be ineffective against DLSCA attacks [5], [6].

Circuit-level on-chip power SCA countermeasures include charge recovery logic [7], switched capacitor current equalizer [8], [9], integrated voltage regulator (IVR) [10], and all-digital low-dropout (LDO) regulator [11], which suffer from performance degradation, high power/area overheads because of large embedded passives, as well as EM leakage from large metal-insulator-metal (MIM) capacitor top plates. Simulations of shunt LDO based regulators have been shown to be effective for power SCA resistance [3]. None of these have been evaluated against DLSCA attacks yet.

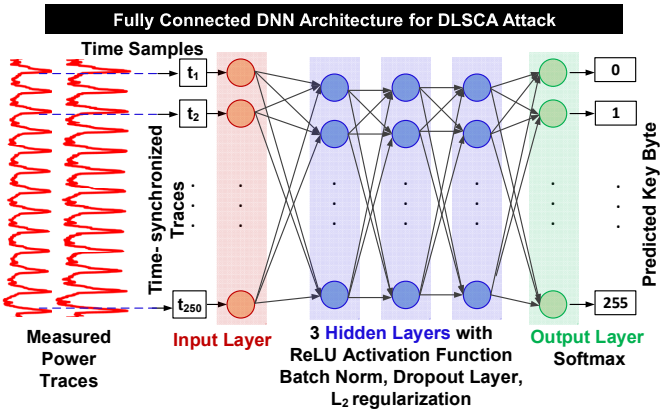


Figure 2: 256-class DNN Architecture for the DLSCA attack on the unprotected AES256 and CDSA-AES256. The input layer consists of 250 neurons (number of time samples in each power trace), followed by the 3 hidden layers with 1K neurons each and finally the 256-neuron output layer to predict the correct key byte. This work uses a fully-connected DNN as the captured traces are time-aligned (synchronized with a trigger pulse for the end of encryption).

Recently, CDSA has been shown to be extremely resilient shown against traditional non-profiled CPA/CEMA attacks [12]. This work, for the first time, evaluates the efficacy of the CDSA hardware against DLSCA attacks on AES-256.

III. DLSCA Attack on the Unprotected AES256 Core

The 65nm test chip contains both unprotected and protected (CDSA) implementations of AES256 (refer Fig. 7(a)). For profiling, we capture power traces from the unprotected core and build the DNN model. Once the training is completed, the DNN model can then be used to attack (classify unseen traces).

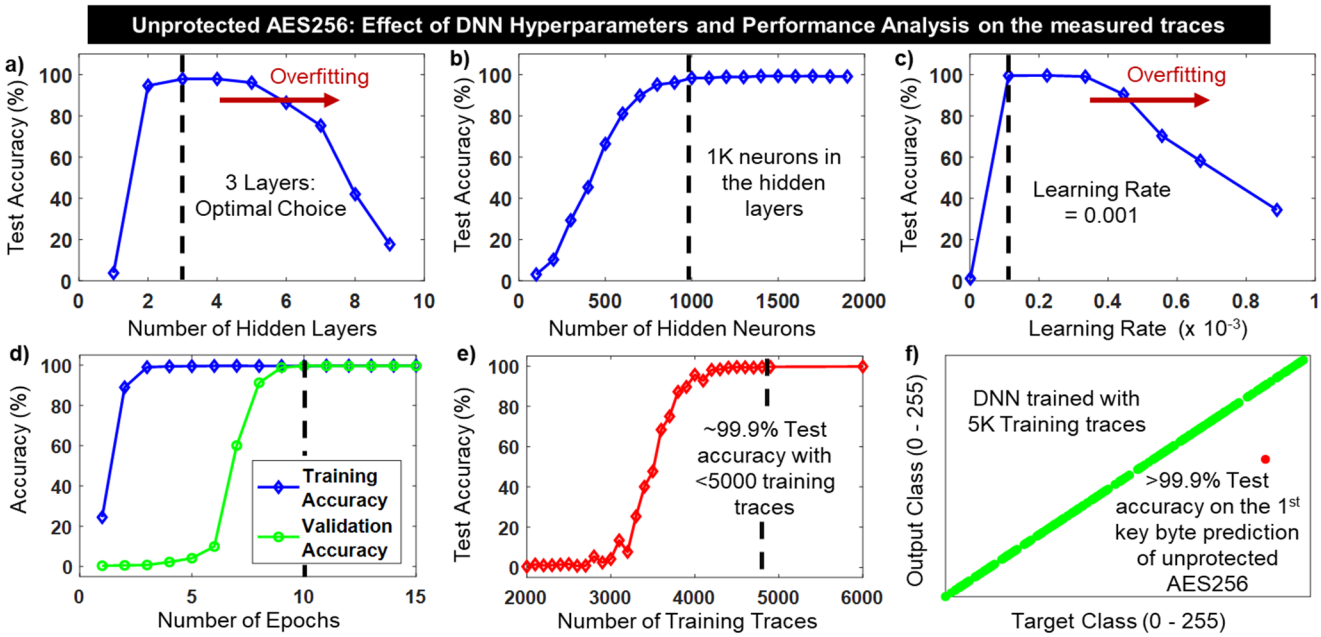


Figure 3: DLSCA attack on the unprotected AES256: (a-c) Effect of the hyperparameters (number of hidden layers, hidden neurons in each layer, learning rate) on the test accuracy of the fully-connected DNN for 5K training traces. (d) Training/Validation accuracy reaches 99.9% within 10 epochs with 5K training traces. (e) Test accuracy of the DNN reaches ~99.9% with <5K training traces with 10 epochs. (f) Confusion plot of the test traces showing >99.9% test accuracy of the DLSCA.

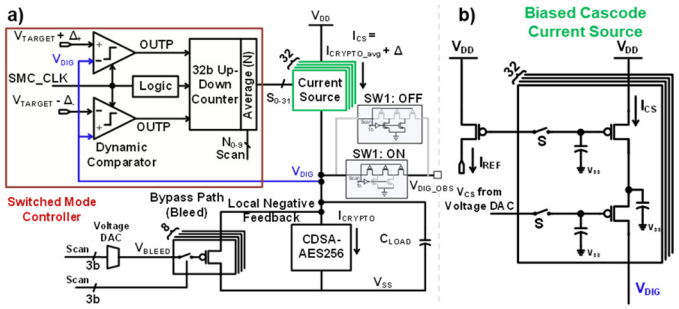


Figure 4: (a) System architecture showing the circuit details of the CDSA, (b) Biased cascode current source (CS) with high output impedance helps achieve significant crypto current signature suppression. Note that highly isolating switches (SW1) are kept for V_{DIG} observability.

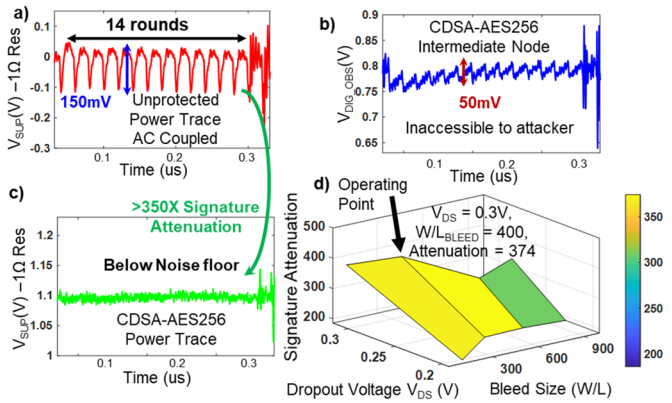


Figure 5: (a-c) Time-Domain Measurement Waveforms showing >350x signature attenuation for the CDSA-AES256 power trace. (d) Design space exploration shows the dependence of attenuation on dropout voltage (V_{DS}) and size of the PMOS bleed.

A. DNN Architecture

Fig. 2 shows the DNN architecture for the DLSCA attack. The input layer consists of 250 neurons (number of time samples in each measured power trace), followed by three hidden layers, each with Rectified Linear Unit (ReLU) non-linear activation, batch normalization, a dropout layer (20%), and L_2 regularization to prevent overfitting and finally the output layer with 256 neurons, which predicts the correct key byte in a single trace utilizing the softmax function.

B. Choice of Hyper-parameters

Fig. 3(a-c) shows the effect of the hyperparameters on the DNN test set accuracy. Three hidden layers with 1K neurons each and a learning rate of 0.001 is the most optimal choice for the unprotected AES256 traces.

C. Performance Analysis

Fig. 3(d, e) shows that the training and validation accuracy of the DNN reaches >99.9% within 10 epochs, and the test accuracy on the unseen traces reaches ~99.9% with only <5K training traces. The test confusion plot (Fig. 3(f)) reveals that only 1 key byte value (marked in red) out of the 256 was misclassified by the DNN, demonstrating a successful DLSCA attack on the 1st key byte of the unprotected AES256.

IV. Current Domain Signature Attenuation Hardware

The main idea of the countermeasure is to embed the crypto core within the CDSA, such that the correlated current signature is significantly suppressed, and the supply current becomes almost constant (independent of the crypto current).

A. Design of the CDSA

The CDSA circuit (Fig. 4) utilizes digitally-tunable cascode current source (CS) with high output impedance to power the AES. The goal of the CDSA circuit is to provide an average load (AES) current plus a small delta current that leaks through the bypass PMOS bleed path to ground, providing local negative feedback leading to the ability to support any $I_{AES_{avg}}$ in between two quantized current levels of the CS (i.e. aids in analog regulation without a high-power shunt-loop). The CS consists of 32 PMOS slices, 16 of which are turned on nominally. The unit current (~94µA) of the CS is chosen to be higher than the key-dependent variation in $I_{AES_{avg}}$ (~72µA), so that the key-dependent information in average DC current is not transferred to supply current (DC regulation) and is leaked by the bleed PMOS, making the design highly secure from an information-theoretic viewpoint. A slow digital switched-mode control (SMC) LDO tracks and regulates the voltage across the AES (V_{DIG} between $V_{TARGET}+\Delta_+$ and $V_{TARGET}-\Delta_-$) by turning on or off the required number of PMOS CS slices. It should be noted that the SMC LDO is a low-BW loop and has a dead band of 50mV, such that it remains disengaged during steady-state operation of the CDSA-AES circuit. Two dynamic comparators compare V_{DIG} with $V_{TARGET}+\Delta_+$ and $V_{TARGET}-\Delta_-$ respectively, and a 32-bit up-down counter with averaging (to control the loop frequency) controls the appropriate number of CS slices to be turned on.

Unlike traditional series LDOs, the supply current in CDSA does not track the AES current. Instead, we choose to tolerate the ~30-50mV voltage droop across the AES engine (V_{DIG} is guard-banded to ensure no performance degradation at the cost of some power overhead), and the high impedance (r_{ds} >10KΩ) CS on top ensures that the current fluctuation at the supply is attenuated by AT = $\omega_{AES}C_L r_{ds}$, i.e. >350× ($i_{CS} = \frac{v_{DIG}}{r_{ds}}$, $v_{DIG} = \frac{i_{AES}}{\omega C_L}$, $AT = \frac{i_{AES}}{i_{CS}} = \omega_{AES}C_L r_{ds}$). The use of cascode CS biased in subthreshold saturation increases r_{ds} by ~10× compared to one-stack CS, allowing 10× reduction in C_L (only 150pF, iso-attenuation) across the crypto engine.

B. Time-Domain Measurements & Design Space Exploration

The shunt path PMOS bias (near-threshold operation) as well as number of PMOS legs ON are scan controllable to analyze the effect of the extra bleed current on signature attenuation. Time-domain measurements of the unprotected AES vs. CDSA-AES show a signature attenuation of >350× for the power traces (Fig. 5). Design space exploration of the CDSA-AES reveals the optimal operating point at dropout voltage of 0.3V across the CS stage and a bleed size of 400. The unprotected AES is powered with 0.8V input and consumes ~1mA average current at 50MHz (refer Fig. 7(b)).

V. DLSCA Attack on the Protected CDSA-AES256 Core

The captured traces from the CDSA-AES256 are now fed to the 256-class DNN for profiling. Fig. 6(a) shows that the DNN does not train on the protected traces (even with 10M traces and 100 epochs) as the signature remains deeply buried under the system noise (without any additional noise injection). Fig. 6(b)

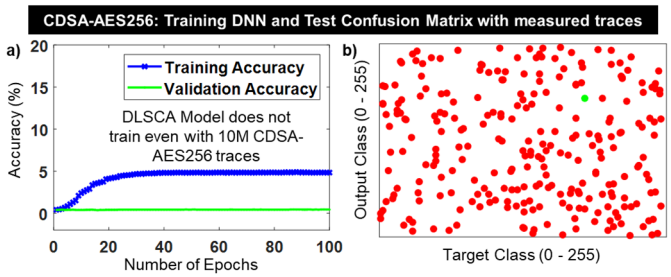


Figure 6: DLSCA attack on the CDSA-AES: (a) Training/validation accuracy does not improve even with 10M traces. (b) Test confusion matrix shows a random trend (~0.3% test accuracy) with numerous misclassifications.

shows the confusion matrix for the unseen test traces from the CDSA-AES256. As we can expect, the DNN does not classify the key bytes correctly (red dots represent misclassifications) and the accuracy is close to random ($\frac{1}{256}$ ~0.3%).

A. Comparison with the State-of-the-Art Countermeasures

Fig. 7(c) shows a comparison with the state-of-the-art existing circuit-level countermeasures. While none of the previous countermeasures have been evaluated against DLSCA attacks, CDSA is the first circuit-level technique demonstrating DLSCA resilience.

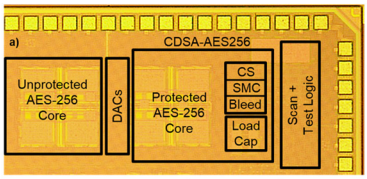
Compared to the unprotected AES256 implementation, the DLSCA immunity is significantly improved by >2000× (>10M compared to 5K traces for training), at the expense of 49.8% power and 36.7% area overheads. It should be noted that the countermeasure is generic and can be used with any other crypto engine, or a combination of multiple crypto engines without any performance overheads.

VI. Remarks & Conclusion

The system developed in 65nm CMOS embeds the crypto core (AES256) within a CDSA hardware such that the critical signature is highly attenuated, to thwart DLSCA attacks. The DNN model which was trained within 5K traces for the unprotected AES256, could not be trained even with 10M traces for the CDSA-AES (Table 1). The >350× signature attenuation of the CDSA promises an improvement of >350^2×, which implies protection up to >600M traces for the DNN training. However, being time-limited due to our trace capture framework for DLSCA, we could demonstrate DLSCA resilience up to 10M traces.

Note that a fully connected DNN is chosen for the DLSCA attack as the traces are perfectly aligned in time (using the on-chip trigger pulses for end of encryption), and hence CNN is not necessary. Also, for the CDSA, signature attenuation is fundamental to the correlated leakage and hence CNNs would not provide any extra benefit over fully connected DNNs for low SNR scenarios. Although the assumption of a fixed plaintext for profiling the DNN may not be most practical for a real attack, it provides a methodology for fast leakage assessment in the machine learning domain and allows to evaluate the efficacy of a countermeasure.

Finally, CDSA is a low-overhead technique to provide high resiliency against DLSCA attacks (>2000×) without any performance degradation and can be extended to any other crypto algorithm.



b)		
Unprotected AES-256 Core	Performance	0.8V/50MHz/0.8mW
	Architecture	256b parallel AES
	Active Area	0.15 mm²
CDSA-AES256	Input	1.1V
	Active Area	0.205 mm²
	Current Efficiency	91.74%@1mA load (90uA through bleed)
	Load Cap	150pF MOS Cap (0.015 mm²)

c) Parameter			This Work	VLSI'15 [7]	ISSCC '09 [8]	ISSCC'17 [10]	ISSCC'19 [11]
Countermeasure Technique			Current Domain Signature Attenuation	Charge Recovery Logic	Switched Capacitor Current Equalizer	Integrated Buck Regulator	Digital LDO with randomization
Attack Technique			ML Attack	Statistical Correlational/Differential Attacks			
Process			65nm CMOS	65nm CMOS	130nm CMOS	130nm CMOS	130nm CMOS
Crypto Algorithm			AES-256	AES-128	AES-128	AES-128	AES-128
Standalone AES Power/Frequency			0.8mW@ 50MHz, 0.8V	138mW@1.3 2GHz	33mW@ 100MHz	10.5mW@ 40MHz	10.9mW@ 80MHz, 0.84V
Design Overheads	Area		36.7%[c]	25%	33%	1%[a]	36.9%[b]
	Power		49.8%[c]	30%	20%	5%*[a]	32%
	Perf.		0%	0%	50%	3.33%	10.4%
SCA Analysis	Time/Freq Domain		Time, Freq	Time	Time	Time, Freq	Time, Freq
	MTD/ Traces to Train	CPA	[12]	940K	>10M	>100K	8M
		DLSCA	>10M	-	-	-	-

[a]Does not include regulator area/power, [b]Does not include MIM Cap area, [c]Power overhead includes the dropout voltage across CS, the excess bleed current drawn during the steady-state operation, as well as the V_{DD} guard-banding to ensure no performance degradation. Area overhead includes the load cap area, CS, SMC loop, and the bleed path.

Figure 7: (a, b) Chip Micrograph and design summary of the system. (c) Comparison with state-of-the-art countermeasures.

Efficacy of CDSA-AES256: Summary

	Traces to Train	Traces to Attack
Unprotected AES256	<5K	~1 (Single-trace attack)
CDSA-AES256	>10M	- (Model could not be trained)

Table 1: Summary of CDSA-AES256 countermeasure against DLSCA attacks.

References

[1] G. Hospodar, B. Gierlichs, E. De Mulder, I. Verbauwhede, and J. Vandewalle, "Machine learning in side-channel analysis: a first study," *J. Cryptogr. Eng.*, vol. 1, no. 4, p. 293, Oct. 2011.

[2] D. Das et al., "X-DeepSCA: Cross-Device Deep Learning Side Channel Attack," in *Proceedings of the 56th Annual Design Automation Conference 2019*, New York, NY, USA, 2019, pp. 134:1–134:6.

[3] D. Das et al., "ASNI: Attenuated Signature Noise Injection for Low-Overhead Power Side-Channel Attack Immunity," *IEEE TCAS-1* 2018.

[4] E. Cagli, C. Dumas, and E. Prouff, "Convolutional Neural Networks with Data Augmentation against Jitter-Based Countermeasures -- Profiling Attacks without Pre-Processing," CHES 2017.

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

[6] H. Maghrebi, T. Portigliatti, and E. Prouff, "Breaking Cryptographic Implementations Using Deep Learning Techniques," 921, 2016.

[7] S. Lu, Z. Zhang, and M. Papaefthymiou, "1.32GHz high-throughput charge-recovery AES core with resistance to DPA attacks," in *2015 Symposium on VLSI Circuits (VLSI Circuits)*, 2015, pp. C246–C247.

[8] C. Tokunaga and D. Blaauw, "Secure AES engine with a local switched-capacitor current equalizer," in ISSCC 2009, pp. 64-65,65a.

[9] C. Tokunaga and D. Blaauw, "Securing Encryption Systems with a Switched Capacitor Current Equalizer," *JSSC*, pp. 23–31, Jan. 2010.

[10] M. Kar et al., "8.1 Improved power-side-channel-attack resistance of an AES-128 core via a security-aware integrated buck voltage regulator," in *ISSCC* 2017, pp. 142–143.

[11] A. Singh et al., "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 *ISSCC* 2019, pp. 404–406.

[12] D. Das et al., "27.3 EM and Power SCA-Resilient AES-256 in 65nm CMOS Through >350× Current-Domain Signature Attenuation", accepted in ISSCC 2020.