34.2 Side-Channel Attack Counteraction via Machine Learning-Targeted Power Compensation for Post-Silicon HW Security Patching

Qiang Fang*[1], Longyang Lin*[1,2], Yao Zu Wong[1], Hui Zhang[1], Massimo Alioto[1]

[1]National University of Singapore, Singapore, Singapore
[2]Southern University of Science and Technology, Shenzhen, China

*Equally Credited Authors (ECAs)

Counteracting side-channel attacks has become a basic requirement in secure integrated circuits handling physical or sensitive data through cryptography, and preventing information leakage via power and electromagnetic (EM) emissions. Over time, the implementation of protection techniques against power analysis and EM attacks has progressively moved from design-specific (i.e., requiring redesign for their reuse [1], [2], [3]) to design-reusable frameworks [4]-[10], facilitating reuse with no modifications across designs, system security verification, and reducing the area/power overhead through reuse of existing silicon infrastructure across secure design IPs on the same die. Accordingly, embedding protection into regulators has been extensively explored to degrade the attack SNR and increase the minimum traces to key disclosure (MTD) via current equalization [4], a switching regulator with randomized loop control [5], a digital LDO (DLDO) with noise injection [6], a DLDO with randomized thresholds and AES transformations [7], a DLDO based on an edge-chasing quantizer [8], current-domain signature attenuation [9] and an additional time-varying transfer function [10]. Such protections allow design reuse and some degree of power-security flexibility, but have common limitations in that: 1) they indiscriminately compensate the entire large-signal power rather than focusing on small-signal information-sensitive power contributions, preventing power overhead reductions, 2) the level of protection is set at design time, and cannot improve after chip fabrication (no learning), 3) they cannot adapt to mitigate newly discovered side-channel vulnerabilities and attacks. Indeed, power overhead and security upgrade-ability over time are crucial in energy-autonomous systems with long lifespans and in applications where device replacement is expensive or unfeasible (e.g., IoT, implantables).

In this work, a design-adaptive counteraction approach against side-channel attacks is introduced to enable post-silicon upgrade-ability and security fixes over time ("hardware patch"), reuse across ciphers under single- and multi-standard encryption, and targeted compensation of information-sensitive power contributions for low power overhead. The latter are actively compensated by a lightweight machine-learning power model and a power compensator, whose weight updates allow post-silicon improvements and large-scale deployment of security fixes throughout the device lifespan. A 40nm testchip demonstrates adaptability across ciphers and their implementations, while retaining conventional standard-cell-based design for easy adoption, system integration and *in-situ* protection. Hardware patching for new vulnerabilities is demonstrated by introducing and then counteracting a newly proposed attack to PRESENT via weight updates.

The architecture in Fig. 34.2.2 comprises a machine learning unit (MLU) that models the cumulative energy associated with the data-sensitive signal transitions (i.e., correlated to the key) within the crypto-core during an encryption. The MLU input features are directly selected from the cryptographic core via multiplexing (both Hamming weight/distance of intermediate values generated throughout the encryption rounds in both linear/non-linear functions). The MLU output drives a capacitive DAC (capDAC) that draws from the supply an energy contribution offsetting the data-sensitive contributions of the crypto-core, canceling them out. After training (below), the MLU directly compensates the small-signal energy contributions that are truly correlated to the key (as defined by the relevant input features), thereby avoiding indiscriminate higher power compensation of the entire crypto-core power in prior art. More accurate MLU models (e.g., via better training, feature selection, model size) allow finer energy compensation and heightened level of protection (i.e., higher MTD).

Figure 34.2.3 shows the *N*-bit capDAC standard-cell implementation comprising *N* binary-scaled gate clusters, whose transition and energy contribution is enabled only if their corresponding input bit from the MLU is 1 (coarse 9b compensation). The standard-cell capDAC energy naturally tracks the crypto-core across process corners, voltages and temperatures. Automated placement and routing (PNR) in a single design iteration avoids manual optimization and iterative PNR, and the non-linearity due to its irregular placement is compensated by inserting and training redundant LSBs for fine 4b compensation. From Fig. 34.2.3, the resulting DNL (INL) RMS non-linearity is 1.1LSB (3.8LSB). Being in the same power domain as the crypto-core, any information leakage of the above blocks is inherently compensated by including them in the training loop (Fig. 34.2.3). A larger number of training iterations better captures such power contributions, and rapidly improves compensation accuracy, increasing MTD to the state-of-the-art billion-scale trace count (Fig. 34.2.3). In this work, a simple regression model was adopted for the MLU. The MLU does not need to respond instantaneously, as its latency requirement is relaxed (>10 cycles, see below) by DC-DC conversion and supply decap, since they integrate and hence spread the exposure of information-sensitive power contributions over a longer window. In addition to the in-situ power compensation, protection against EM attacks is further enhanced via usage of lower levels of metal for local connectivity/routing (M1…M5), and upper metal for shielding as in [9]. Figure 34.2.3 also introduces a newly discovered attack to the last round of the PRESENT cryptographic algorithm based on bit regrouping, as opposed to conventional attack to its first round. This new attack is used below to illustrate the hardware patching capability described in Fig. 34.2.1.

A power analysis (EM) attack to PRESENT without protection easily discloses the key after MTD=5k traces (7k) from Fig. 34.2.4. The proposed protection configured to counteract the conventional attack to the first round is shown to be effective, raising MTD to more than 1.2B traces. However, when the new attack to the last round is carried out, this protection is expectedly ineffective and the key is easily disclosed in 8k traces. As illustration of the hardware patch concept, the inclusion of features from the last round and model retraining to counteract both the conventional and new attack offer again effective protection, raising the MTD to more than 1.2B for both power analysis and EM attacks. From Fig. 34.2.4, TVLA on an unprotected PRESENT crypto-core under the power analysis (EM) attack reaches the |t|-value target of 4.5 at 700 (800) traces. The proposed protection increases such trace count to 250M (100M), providing an improvement of 357,140× (125,000×) compared to the unprotected version, confirming state-of-the-art robustness against side-channel attacks.

To demonstrate the versatility of the protection in this work and the hardware API concept in Fig. 34.2.1, the attacks were repeated on an AES crypto-core without and with the proposed protection, while retraining the same on-chip model (very same circuit, same power domain) to adapt to the AES algorithm rather than PRESENT. From Fig. 34.2.5, the unprotected AES under power analysis (EM) attack discloses the key after MTD=10k traces (20k), whereas the protected core increases MTD to more than 1.2B. The resulting 120,000× (60,000×) MTD improvement over unprotected AES is in line with the results on PRESENT in Fig. 34.2.4, confirming the consistency of the protection level across cryptographic algorithms. Similarly, TVLA results on the AES core in Fig. 34.2.5 are again consistent with PRESENT.

The above MTD=1.2B observed in the 40nm testchip is in line with the best protection techniques [7], [9], [10], and higher than [1]-[6], [8] by at least three orders of magnitude. The trace count at which TVLA reaches the 4.5 target is higher than prior art, and in particular by up to 3× compared to [7], and 35× higher than [10]. The 8.6% power overhead is lower than prior art, and 5.8× lower than recent demonstrations with highest MTD (equivalent to this work). This is due to the ability of the machine learning power estimator to compensate only the small-signal information-sensitive contributions, rather than the large-signal fluctuations of the overall power. The 69% area overhead over an AES core is 1.4-to-2× higher than techniques with highest MTD [9], [10], and can be reduced in architectures where the protection is shared and reconfigured to support multiple on-chip design IPs (e.g., microprocessor).

As opposed to the (off-chip) machine learning-assisted framework in [3], the proposed approach has an on-chip run-time machine learning model with flexible size, comprehensive feature availability, flexible selection to adapt to different ciphers and different microarchitectural implementations of the same cipher. Hence, the proposed machine learning-based protection offers a higher degree of flexibility and adaptation to different designs and upgradeability to counteract new attacks. This ultimately favors design reuse and also uniquely enables hardware patching against side-channel attacks, maintaining high security standards during the lifespan of a device and even allowing improvements over time.

Acknowledgment:
The authors acknowledge the support from the Singapore National Research Foundation ("SOCure" grant NRF2018NCR-NCR002-0001).

References:

[1] T. Popp et al., "Evaluation of the Masked Logic Style MDPL on a Prototype Chip," *CHES*, Springer, 2007.
[2] D. D. Hwang et al., "AES-Based Security Coprocessor IC in 0.18-μm CMOS with Resistance to Differential Power Analysis Side-Channel Attacks," *IEEE JSSC*, vol. 41, no. 4, pp. 781-792, 2006.
[3] W. Shan et al., "Machine Learning Assisted Side-Channel-Attack Countermeasure and Its Application on a 28-nm AES Circuit," *IEEE JSSC*, vol. 55, no. 3, pp. 794-804, 2020.
[4] C. Tokunaga and D. Blaauw, "Secure AES Engine with a Local Switched-Capacitor Current Equalizer," *ISSCC*, pp. 64-65, 2009.
[5] M. Kar et al., "Improved Power-Side-Channel-Attack Resistance of an AES-128 Core via a Security-Aware Integrated Buck Voltage Regulator," *ISSCC*, pp. 142-144, 2017.
[6] A. Singh et al., "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," *ISSCC*, pp. 404-406, 2019.

2022 IEEE International Solid-State Circuits Conference (ISSCC) | 978-1-6654-2800-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISSCC42614.2022.9731755

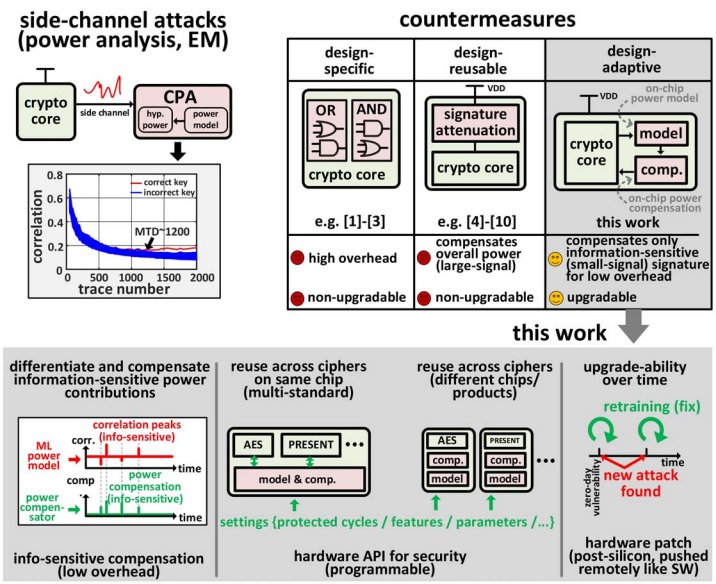


Figure 34.2.1: Different categories of side-channel attack countermeasures.

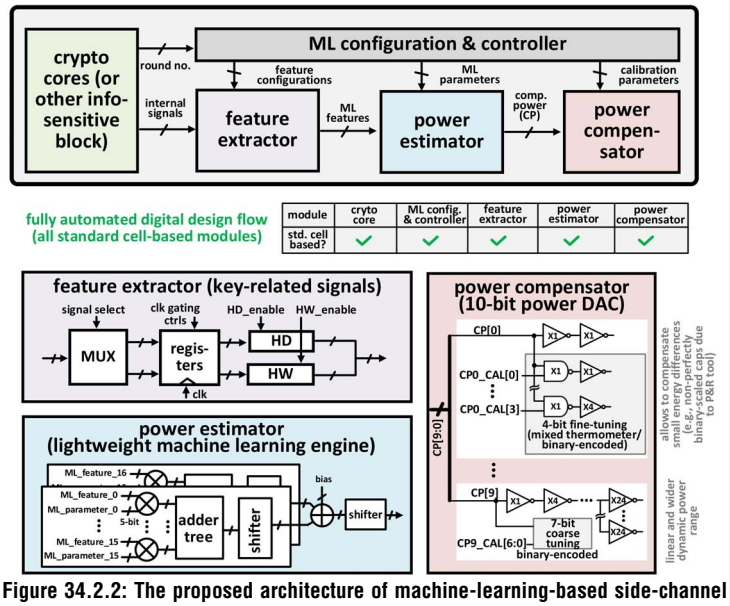


Figure 34.2.2: The proposed architecture of machine-learning-based side-channel attack counteraction.

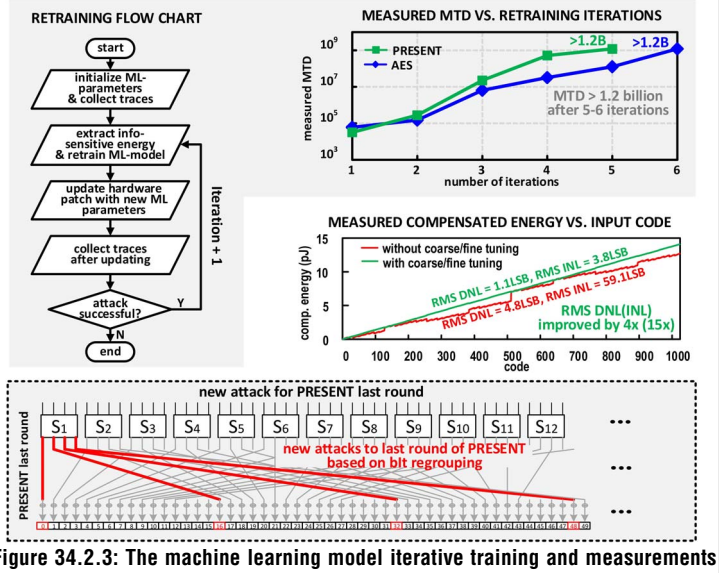
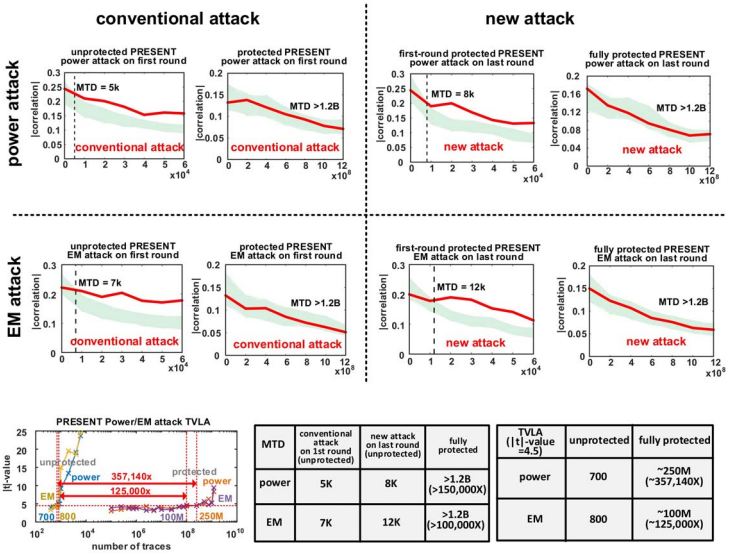


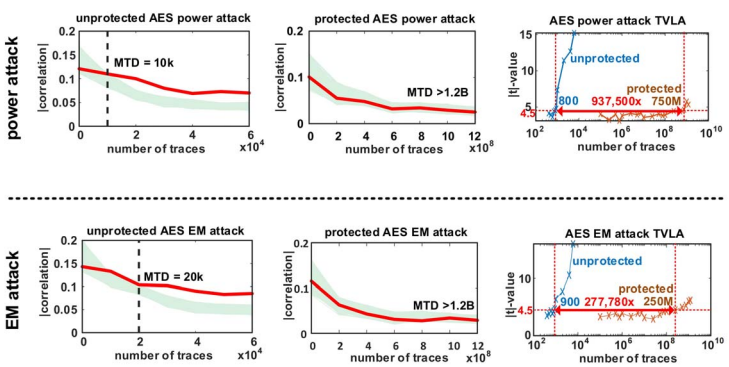
Figure 34.2.3: The machine learning model iterative training and measurements (top); Energy measurements of power DAC (middle); New attack to the PRESENT cryptographic algorithm (bottom).



MTD	conventional attack on 1st round (unprotected)	new attack on last round (unprotected)	fully protected
power	5K	8K	>1.2B (>150,000X)
EM	7K	12K	>1.2B (>100,000X)

TVLA (|t|-value =4.5)	unprotected	fully protected
power	700	~250M (~357,140X)
EM	800	~100M (~125,000X)

Figure 34.2.4: Measured MTD and TVLA for unprotected/protected PRESENT crypto-core for both power and EM attacks.



MTD	unprotected AES	protected AES
power	10K	>1.2B (>120,000x)
EM	20K	>1.2B (>60,000x)

TVLA (|t|-value =4.5)	unprotected AES	protected AES
power	800	~750M (~937,500x)
EM	900	~250M (~277,780x)

Figure 34.2.5: Measured MTD and TVLA for unprotected/protected AES crypto-core for both power and EM attacks.

	This work	ISSCC'21 [10]	ISSCC'20 [9]	ISSCC'20 [8]	VLSI'20 [7]	JSSC'20 [3]	ISSCC'19 [6]	ISSCC'17 [5]	ISSCC'09 [4]	JSSC'06 [2]
counteraction technique	run-time machine learning-based	digital signature attenuation	current domain signature attenuation	edge-chasing quantizer based digital LDO	digital NL-LDO with randomization	ML-based HD redistribution	digital LDO with randomization	integrated buck regulator + LFSR PWM	switched capacitor current equalizer	WDDL dual-rail logic
crypto algorithm	AES, PRESENT	AES	AES	AES	AES	AES	AES	AES	AES	AES
CMOS process	40nm	65nm	65nm	65nm	14nm	28nm	130nm	130nm	130nm	180nm
area overhead	69% [a]	52%	36.7%	27.5%	8%	36%	36.9%	**1%**	7.2%	300%
power overhead	**8.6%**	50%	49.8%	19.4%	10%	32%	32%	5% (exclude IVR)	33%	400%
throughput loss	0%	0%	0%	4.54%	0.7%	0%	10.4%	3.33% (exclude IVR)	50%	25%
attack mode	power, EM	power, EM	power, EM	power	power, EM	power	power, EM	power	power	power
MTD	**AES: >1.2B PRESENT: >1.2B**	**>1.25B(Power) >1.25B(EM)**	**>1B(Power) >1B(EM)**	>7M	>1B	1.5M	8M (power) 6.8M (EM)	>100K	>10M	255K
MTD improvement	AES: >120,000X (power) >60,000X (EM) PRESENT: >150,000X (power) > 100,000X (EM)	**>178,570X (Power) >138,890 (EM)**	>125,000X (Power) >83,333X (EM)	14,000X	>100,000X	446X	3579X(P-AES) 2182X (S-AES)	>20X	>1667X	120X
TVLA	AES: **750M (power) 250M (EM)** PRESENT: 250M (power) 100M (EM)	21.6M (power) 3.6M (EM)	N.A.	N.A.	>250M (power & EM)	not used	not used	N.A.	not used	not used
TVLA improvement	AES: **937,500X (power) 277,780X (EM)** PRESENT: 357,140X (power) 125,000X (EM)	290,000X (power) 70,000X (EM)	N.A.	N.A.	>250,000X (power & EM)	not used	not used	N.A.	not used	not used
standard cell based	YES	YES	NO	NO	NO	YES	NO	NO	NO	YES (custom std-cell)
flexibility & upgradeability	**design-adaptive [b] (run-time on-chip machine learning model)**	design-reusable [c]	design-reusable [c]	design-reusable [c]	design-reusable [c]	design-specific [d]	design-reusable [c]	design-reusable [c]	design-reusable [c]	design-specific [d]

[a] With combined AES and PRESENT cores. Only AES core is 93%
[b] The same protection can be reused for different ciphers or information-sensitive blocks on chip, and can be adapted over time to improve security (flexible features, machine learning model)
[c] The same protection can be reused for different ciphers or their implementations, but cannot be improved over time (no training)
[d] The same protection can be reused for different ciphers or their implementations, but cannot be improved over time (no training)

Figure 34.2.6: Summary of the run-time machine-learning-based counteraction performance and comparison with prior art (best highlighted in bold).

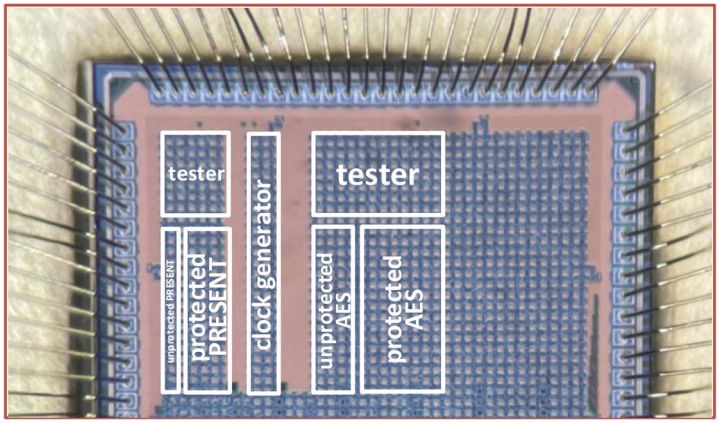


Figure 34.2.7: Die micrograph of the 40nm testchip.

Additional References:
[7] R. Kumar et al., "A SCA-Resistant AES Engine in 14nm CMOS with Time/Frequency-Domain Leakage Suppression using Non-linear Digital LDO Cascaded with Arithmetic Countermeasures," *IEEE Symp. VLSI Circuits*, 2020.
[8] Y. He and K. Yang. "A 65nm Edge-Chasing Quantizer-Based Digital LDO Featuring 4.58 ps-FoM and Side-Channel-Attack Resistance," *ISSCC*, pp. 384-385, 2020.
[9] D. Das et al., "EM and Power SCA-Resilient AES-256 in 65nm CMOS Through >350× Current-Domain Signature Attenuation," *ISSCC*, pp. 424-426, 2020.
[10] A. Ghosh et al., "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," *ISSCC*, pp. 500-501, 2021.

978-1-6654-2800-2/22/$31.00 ©2022 IEEE