Poster: Autonomous LLM-Orchestrated De-obfuscation of Fully Unrolled AES via Zero-State Differential Power Isolation

1st Mani Rupak Gurram
Department of Electrical and Computer Engineering
prairie view a&m university
prairie view, TX, USA
mgurram@pvamu.edu

2nd Akshay Raghavendra Kulkarni
Department of Electrical and Computer Engineering
prairie view a&m university
prairie view, TX, USA
arkulkarni@pvamu.edu

***Abstract*—Fully unrolled cryptographic architectures execute all operational rounds within a single clock cycle, maximizing throughput while generating a massive algorithmic noise floor. This structural combinational noise inherently thwarts standard temporal Side-Channel Analysis (SCA). In this poster abstract, we present an autonomous, AI-driven framework that profiles and defeats the combinational noise floor of a 161,000-gate unrolled AES-128 core. By utilizing a Large Language Model (LLM) operating in a Reasoning and Acting (ReAct) loop, our framework empirically validates the heuristic failure of standard Correlation Power Analysis (CPA), autonomously pivots to a Zero-State Differential Power Isolation methodology, and mathematically cancels multi-round algorithmic noise. The framework successfully isolates localized combinational leakage ($r > 0.318$) and extracts the physical-to-logical routing map introduced by the EDA synthesis compiler.**

1. Introduction

Unlike standard iterative architectures that register the state between rounds, a fully unrolled Advanced Encryption Standard (AES) core [1] instantiates all encryption rounds as a continuous block of combinational logic. While optimal for latency and throughput, the simultaneous switching of 161,000 logic gates creates a massive algorithmic noise floor that inherently masks localized leakage and thwarts standard Correlation Power Analysis (CPA) [2]. Furthermore, Electronic Design Automation (EDA) [3], [4] synthesis routing inadvertently introduces physical-to-logical bit obfuscation during the Place and Route (P&R) phase, driving the critical need for autonomous, AI-orchestrated side-channel extraction.

2. Experimental Setup

The empirical evaluation is conducted using a ChipWhisperer CW305 FPGA target board equipped with an Xilinx Artix-7 device as shown in Figure 1. The physical target is a fully unrolled AES-128 cryptographic core synthesized with approximately 161,000 combinational logic gates. Unlike standard iterative implementations, this architecture contains no intermediate state registers between the ten encryption rounds, enforcing a single-cycle evaluation that maximizes structural combinational noise.

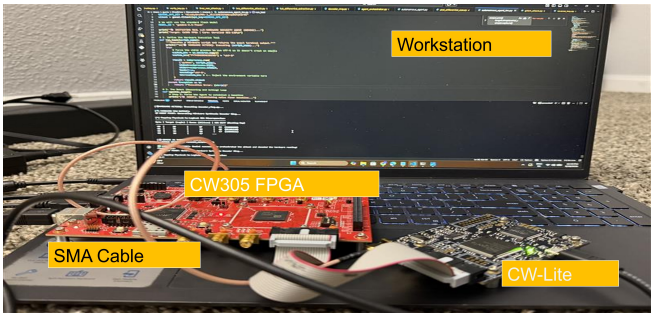

Figure 1. Experimental setup: ChipWhisperer CW305 FPGA target capturing global power via SMA

Power consumption is captured globally via an SMA connector using standard ChipWhisperer capture hardware. The autonomous framework is driven by a local Python orchestrator [6], [7] integrating a Large Language Model (LLM). Operating in a Reasoning and Acting (ReAct) [8] paradigm, the LLM interfaces directly with the hardware API to issue target plaintexts, capture power traces, compute Pearson correlation coefficients (r), and dynamically pivot execution methodologies without human intervention.

3. Autonomous ReAct Pipeline

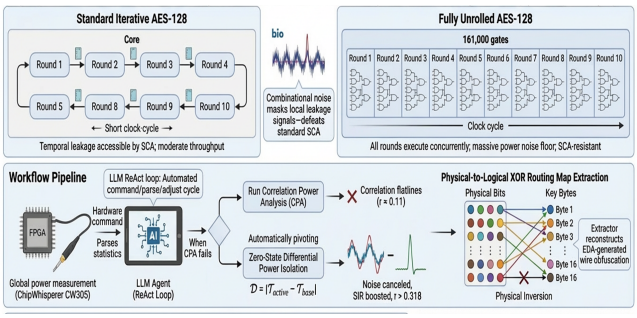

Figure 2. The autonomous LLM-orchestrated side-channel framework defeating unrolled combinational noise

To address the resilience of fully unrolled combinational logic, we propose an autonomous, LLM-orchestrated side-channel framework as shown in Figure 2. The agent initializes global SMA power profiling against the AES core on a target CW305 FPGA. Upon analyzing the real-time statistical output, the agent empirically validates the heuristic failure of standard CPA, diagnosing that the global structural noise successfully masks the localized leakage ($r \approx 0.11$). In response, the orchestrator dynamically pivots to a Zero-State [5]Differential Power Isolation technique [9], [10]. By injecting known-zero states across non-target bytes, the agent mathematically subtracts the multi-round algorithmic baseline from the active trace ($\mathcal{D} = |\mathcal{T}_{active} - \mathcal{T}_{base}|$). This single-channel cancellation dramatically improves the Signal-to-Interference Ratio (SIR), isolating the physical key guess. Finally, the framework maps the extracted silicon footprint against the logical target state, autonomously reverse-engineering a complete 16-byte XOR routing map to de-obfuscate the wire inversions applied by the EDA synthesizer.

4. Preliminary Results

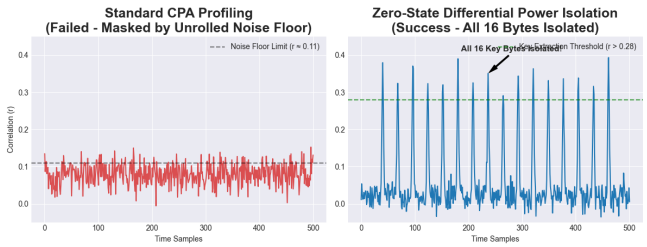

Figure 3. Side-channel correlation traces for the unrolled AES-128 core. (Left) Standard temporal CPA fails to breach the combinational noise floor ($r \approx 0.11$). (Right) The autonomous application of Zero-State Differential Power Isolation successfully cancels multi-round algorithmic noise, isolating all 16 physical key bytes well above the extraction threshold ($r > 0.318$)

The evaluation empirically demonstrated in Figure 3 that standard temporal CPA systematically fails against unrolled hardware, yielding a flat correlation peak ($r \approx 0.11$). However, the autonomous agent successfully bypassed this structural defense. The execution of the Zero-State Differential isolation successfully isolated the target Look-Up Table (LUT) combinational leakage, yielding a definitive correlation spike ($r > 0.318$). Utilizing this localized physical leakage, the LLM framework automatically mapped the physical discrepancies back to the logical key byte targets, producing a verified 16-byte physical-to-logical XOR routing matrix, proving the reversibility of EDA logic synthesis optimization.

5. Conclusion

This work demonstrates that deeply unrolled combinational architectures cannot rely on structural noise as a robust defense against adaptive, differential side-channel methodologies. By integrating LLMs as central reasoning engines, our framework successfully bridges the gap between initial heuristic failure and advanced mathematical isolation. Furthermore, we show that EDA routing optimizations inadvertently act as physical obfuscators that can be systematically reverse-engineered through autonomous agentic profiling.

References

[1] Dworkin, M.J., Barker, E., Nechvatal, J.R., Foti, J., Bassham, L.E., Roback, E. and Dray Jr, J.F., 2001. Advanced encryption standard (AES).

[2] Brier, E., Clavier, C. and Olivier, F., 2004, August. Correlation power analysis with a leakage model. In International workshop on cryptographic hardware and embedded systems (pp. 16-29). Berlin, Heidelberg: Springer Berlin Heidelberg.

[3] Wang, L.T., Chang, Y.W. and Cheng, K.T.T. eds., 2009. Electronic design automation: synthesis, verification, and test. Morgan Kaufmann.

[4] Huang, G., Hu, J., He, Y., Liu, J., Ma, M., Shen, Z., Wu, J., Xu, Y., Zhang, H., Zhong, K. and Ning, X., 2021. Machine learning for electronic design automation: A survey. ACM Transactions on Design Automation of Electronic Systems (TODAES), 26(5), pp.1-46.

[5] A.K.I.S.H.I.T.A., Zero-Value Point Attacks on Elliptic Curve Cryptosystem. ISC 2003.

[6] Fan, S., Cong, X., Fu, Y., Zhang, Z., Zhang, S., Liu, Y., Wu, Y., Lin, Y., Liu, Z. and Sun, M., 2024. Workflowllm: Enhancing workflow orchestration capability of large language models. arXiv preprint arXiv:2411.05451.

[7] Luo, H., Liu, Y., Zhang, R., Wang, J., Sun, G., Niyato, D., Yu, H., Xiong, Z., Wang, X. and Shen, X., 2025. Toward edge general intelligence with multiple-large language model (Multi-LLM): architecture, trust, and orchestration. IEEE Transactions on Cognitive Communications and Networking.

[8] Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K.R. and Cao, Y., 2022, October. React: Synergizing reasoning and acting in language models. In The eleventh international conference on learning representations.

[9] Schramm, K., Leander, G., Felke, P. and Paar, C., 2004, August. A collision-attack on AES: Combining side channel-and differential-attack. In International Workshop on Cryptographic Hardware and Embedded Systems (pp. 163-175). Berlin, Heidelberg: Springer Berlin Heidelberg.

[10] Akkar, M.L. and Goubin, L., 2003, February. A generic protection against high-order differential power analysis. In International Workshop on Fast Software Encryption (pp. 192-205). Berlin, Heidelberg: Springer Berlin Heidelberg., 2003.

Autonomous LLM-Orchestrated De-obfuscation of Fully Unrolled AES via Zero-State Differential Power Isolation

Mani Rupak Gurram, Akshay Raghavendra Kulkarni
Department of Electrical and Computer Engineering, Prairie View A&M University, Prairie View, TX,USA

Abstract

Fully unrolled AES architectures generate massive combinational noise that thwarts standard Side-Channel Analysis. We present an autonomous LLM framework that dynamically defeats this noise floor on a 161,000-gate core. By pivoting to Zero-State Differential Power Isolation, the agent mathematically cancels algorithmic noise, isolates localized leakage ($r > 0.318$), and reverse-engineers the EDA synthesis routing map.

Introduction

Fully unrolled AES architectures compute all encryption rounds within a single clock cycle. While optimal for throughput, the simultaneous switching of 161,000 logic gates creates a massive algorithmic noise floor that inherently masks localized leakage and thwarts standard Correlation Power Analysis (CPA). Furthermore, EDA synthesis routing inadvertently introduces physical-to-logical bit obfuscation, driving the critical need for autonomous, AI-orchestrated side-channel extraction.

Reference: Toru, A.K.I.S.H.I.T.A., Zero-Value Point Attacks on Elliptic Curve Cryptosystem. *ISC 2003.*

Experimental Setup

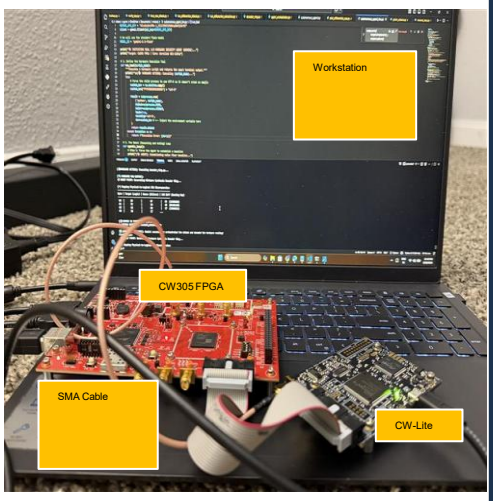

Fig.1.Experimental Setup

- **Target Hardware:** ChipWhisperer CW305 (Artix-7 FPGA) hosting a 161,000-gate fully unrolled AES-128 cryptographic core.
- **Trace Acquisition:** Global power measurements captured synchronously via an SMA cable utilizing a ChipWhisperer-Lite (CW-Lite) capture board.
- **AI Orchestration:** A Python-based Reasoning and Acting (ReAct) loop running on a standard CPU workstation, eliminating the need for expensive local GPU compute.

Proposed Approach

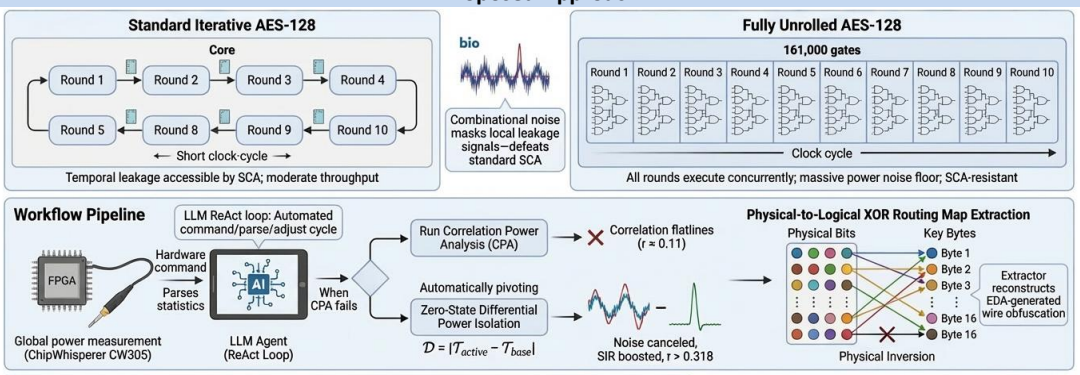

Fig. 2. The autonomous LLM-orchestrated side-channel framework defeating unrolled combinational noise

- **Agentic Profiling:** The LLM orchestrator interfaces with the target CW305 FPGA to capture global SMA power measurements of the 161,000-gate unrolled AES-128 core.
- **LLM Orchestrator Specs:** Powered by the **Gemini API**, requiring minimal local compute (a standard CPU workstation for trace collection and API interfacing). The autonomous diagnosis, methodological pivot, and complete key extraction execute end-to-end in **< 10 minutes**.
- **Failure Diagnosis:** The agent analyzes real-time statistical output, empirically validating that the global structural noise masks localized leakage, causing standard CPA to fail($r \approx 0.11$).
- **Zero-State Differential Pivot:** Diagnosing the heuristic failure, the agent autonomously triggers a methodological pivot, injecting known-zero states across non-target bytes to isolate the active substitution box.
- **Mathematical Noise Cancellation:** The framework mathematically subtracts the multi-round algorithmic baseline from the active trace ($\mathcal{D} = |\mathcal{T}_{active} - \mathcal{T}_{base}|$), dramatically improving the Signal-to-Interference Ratio (SIR).
- **Synthesis Reverse-Engineering:** The agent computationally maps the newly isolated physical silicon footprint against the logical mathematical state, generating a complete 16-byte XOR routing map to de-obfuscate the EDA compiler's wire inversions.

Reference: Schramm, Kai, Gregor Leander, Patrick Felke, and Christof Paar. "A collision-attack on AES: Combining side channel- and differential-attack." In *International Workshop on Cryptographic Hardware and Embedded Systems*, pp. 163-175. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004.

Results

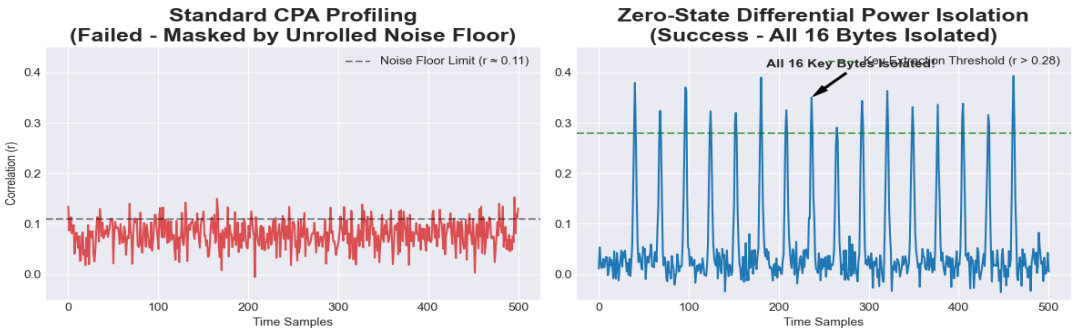

Fig. 3. Standard_CPA VS Zero-State

Side-channel correlation traces for the unrolled AES-128 core. (Left) Standard temporal CPA fails to breach the combinational noise floor ($r \approx 0.11$). (Right) The autonomous application of Zero-State Differential Power Isolation successfully cancels multi-round algorithmic noise, isolating all 16 physical key bytes well above the extraction threshold (r > 0.318)

Conclusion

In conclusion, this work demonstrates that deeply unrolled combinational architectures cannot rely on structural noise as a robust defense against adaptive, differential side-channel methodologies. By integrating Large Language Models (LLMs) as central reasoning engines, our framework successfully bridges the gap between initial heuristic failure and advanced mathematical isolation. Furthermore, we show that EDA routing optimizations inadvertently act as physical obfuscators that can be systematically reverse-engineered through autonomous agentic profiling, ultimately proving that synthesis-level logic decisions leak extractable hardware secrets.