Invited Paper: Side Channel Vulnerability Analysis of Flexible Neuromorphic Circuits

Priyanjana Pal*, Brojogopal Sapui*, and Mehdi B. Tahoori
Karlsruhe Institute of Technology, Germany
{priyanjana.pal, brojogopal.sapui, mehdi.tahoori}@kit.edu

***Abstract*—The rapid advancement of flexible electronics (FE) has driven significant innovation across diverse sectors, including healthcare, wearables, smart packaging, and IoT devices, owing to their adaptability, lightweight form factor, and cost-effectiveness compared to traditional silicon-based electronics. A key computing paradigm in this domain is bespoke classifiers, where model parameters are hardcoded in neuromorphic hardware to meet strict area, power, and cost constraints. By tailoring bespoke hardware to specific tasks, these circuits achieve significant accuracy under tight resource budgets but also introduce distinct security vulnerabilities. The intrinsic flexibility of substrates, unconventional manufacturing processes, and limited protective packaging make such systems particularly vulnerable to security threats, with side-channel attacks (SCAs) being a critical concern. In this work, we systematically investigate SCA vulnerabilities in bespoke TFT-based multilayer perceptron (MLP) classifiers, considering both analog (flexible analog multilayer perceptron (*f-AMLP*)) and digital (flexible digital multilayer perceptron (*f-DMLP*)) realizations. For digital classifiers, we apply correlation power analysis (CPA), leveraging well-established leakage models from silicon-based systems. For analog classifiers, where leakage is continuous, nonlinear, and strongly influenced by device-level variability, we develop a tailored convolutional neural network (CNN)-based regression attack capable of extracting inputs from noisy power traces. Experimental results across benchmark datasets show that *f-DMLP*s can be compromised with 70–85% cumulative attack success rate (ASR) after ≈ 4k–5k traces using CPA, while *f-AMLP*s, though slower to attack initially, reach up to 90–95% ASR after ≈ 8k–9k traces with CNN-based approach.**

I. Introduction

Despite significant progress in silicon-based electronics, particularly regarding power efficiency and transistor miniaturization, their rigid nature, complex manufacturing process, and high production costs restrict their suitability for low-cost, sustainable, and flexible consumer-edge applications [1, 2]. Rapidly growing sectors such as smart packaging, flexible displays, drug delivery systems, RFID tags, smart bandages, and IoT devices require thin, lightweight, cost-efficient, and bespoke fabrication techniques that conventional silicon technologies cannot economically fulfill [2]–[4].

In such scenarios, Flexible electronics (FE) stand out as a viable alternative. FE commonly use thin-film transistor (TFT) technologies, such as amorphous indium-gallium-zinc oxide (a-IGZO). These materials offer key benefits including mechanical flexibility, transparency, and low production costs, making them ideal for various edge scenarios [1, 2]. However, unlike rigid silicon-based devices, flexible TFT systems lack robust protective packaging and operate under fluctuating environmental conditions, which inherently increase their vulnerability to external influences, including mechanical stresses and temperature shifts, increasing security vulnerabilities.

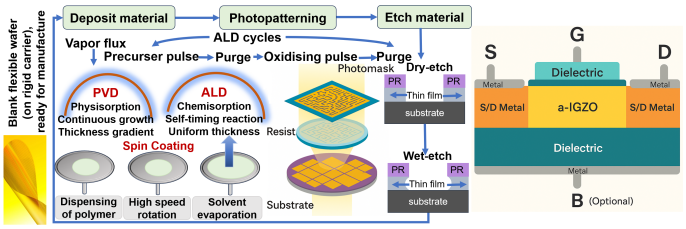


Fig. 1: Fabrication flow of a-IGZO TFT, including material deposition (PVD, ALD, spin coating), photo-patterning, and etching. Insets show flexible substrate, and device structure [12].

A promising computing paradigm in this FE domain is flexible bespoke classifiers, where model parameters are directly hardcoded in neuromorphic[1] hardware to meet strict area, power and cost constraints of FE technology. These architectures are specifically tailored to target tasks and are realized in both analog and digital forms. Analog multilayer perceptron (MLP) often based on unipolar (n-type-only) designs leverage continuous-valued computation to achieve compact, energy-efficient designs [4]–[8]. Digital counterparts, while generally more resilient to environmental noise, still exhibit discrete leakage patterns. Prior work has shown the effectiveness of such bespoke architectures for sensor data processing, real-time classification, and edge AI tasks, particularly where low-cost, low-power and mechanical flexibility are essential [9]–[11].

A critical yet often overlooked security concern in TFT-based FE systems is their susceptibility to side-channel attacks (SCAs), which exploit unintended information leakage such as power consumption or electromagnetic radiation during intermediate computations. Due to their minimal protective packaging, mechanical flexibility, and sensitivity to environmental variations, TFT-based *f-AMLP* classifiers are inherently more exposed to power-based SCAs compared to conventional silicon systems. In analog designs, the continuous-valued signals further amplifies leakage potential, creating substantial security and privacy risks for sensitive domains such as healthcare, wearables etc [2, 3].

Security evaluation of digital silicon-based MLPs is a mature field, offering well-established methodologies such as correlation power analysis (CPA), to be directly adapted to *f-DMLP* classifiers. However, extending these methods to *f-AMLP* is sig-

*Authors contributed equally to this work.

[1]In the cited works, the term "neuromorphic" is used broadly to describe hardware implementations inspired by neural network architectures for parallel computation and target classification tasks, specifically analog and digital multilayer perceptron (MLP)–based classifiers.

© 2025 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

nificantly more challenging. The intrinsic nonlinear transistor characteristics, device-specific threshold voltage shifts, parasitic effects, mechanical stress-induced variations, and strong sensitivity to environmental conditions such as temperature and humidity produce highly complex and noisy leakage profiles. These characteristics limit the effectiveness of traditional statistical models and necessitate the use of advanced, data-driven approaches to capture and exploit subtle correlations in leakage. By applying standard digital SCA techniques to *f-DMLP* and contrasting them with tailored NN-based attacks for *f-AMLP* counterparts, our work provides a comparison of attack success, complexity, and trace efficiency, offering new insights into the distinct security challenges of FE. In short, the contributions of this work are as follows:

- Side-channel vulnerability analysis of TFT-based flexible bespoke classifiers, in both analog (*f-AMLP*) and digital (*f-DMLP*) forms, using realistically simulated power traces incorporating device-level nonidealities and noise sources.
- CNN-based regression attack framework tailored to the continuous, nonlinear, and noise-distorted leakage of *f-AMLP* circuits, and benchmark its effectiveness against a standard CPA attack on the digital counterpart.
- Evaluation in benchmark datasets shows that the *f-DMLP* reaches a high attack success rate (ASR) within 4k to 5k traces, while the *f-AMLP*, despite slower initial growth, achieves a higher ASR (>90%) after ≈8.8k traces, showing the different leakage behavior and attack complexity.

The rest of this paper is structured as follows: Sec. II introduces FE, discusses related works, and presents key preliminaries. Sec. III outlines the proposed power side-channel analysis techniques for flexible TFT-based *f-AMLP* and *f-DMLP* classifiers, including methods to recover critical classifier parameters. In Sec. IV, we evaluate our proposed approach using four benchmark datasets, highlighting input data-dependent vulnerabilities. Finally, Sec. V summarizes our work.

II. Preliminaries

A. Flexible Electronics (FE)

FE offer unique advantages such as mechanical flexibility, lightweight, and compatibility with various flexible substrates including plastics, metals, and polymers. Unlike conventional rigid silicon-based electronics, FE can conform to dynamic or curved surfaces, making them suitable for wearable health monitors, medical devices, smart packaging, and IoT [2, 3].

Structurally, FE utilize thin-film transistors (TFTs), based on a-IGZO, produced via low-temperature processing techniques that enable cost-effective and scalable manufacturing. In a typical process flow, the fabrication process step in Fig. 1 begins by cleaning a rigid glass substrate, spin-coating a flexible substrate onto it, and curing at an appropriate bake temperature. An a-IGZO semiconductor layer is deposited next (e.g., via atomic layer deposition (ALD), or physical vapor deposition (PVD)) to form the active region. Due to material and process constraints, a-IGZO TFT is unipolar (only n-type devices), which requires the use of resistor–transistor logic (RTL) for the circuit realization and imposes restrictions on its design complexity. The integration density is also limited to a few thousand devices, with feature sizes in the µm range, leading to prominence on compact, application-specific bespoke designs.

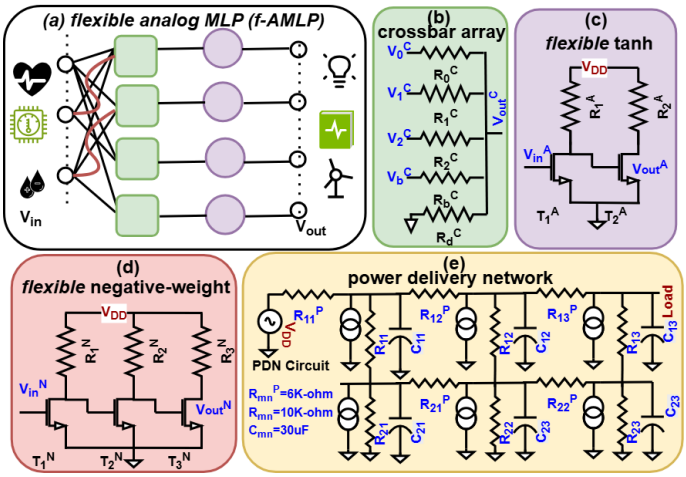


Fig. 2: (a) Schematic of a bespoke *f-AMLP* [16] that receives sensor signals and yields outputs to subsequent devices. Circuit primitives: (b) 3-input, 1-output resistor crossbar.(c) Inverter-based negative weight circuit. (d) Tanh-like activation. (e) Power-delivery network (PDN) for stable power supply.

B. Flexible Multilayer Percepton Classifiers (f-MLPs)

Both analog and digital implementations presented in these works [4, 11, 13]–[15] are bespoke flexible neuromorphic classifers, where model parameters are hardcoded directly into hardware during fabrication to fit tight area–power budgets, thereby reducing computation overhead.

1) Flexible Analog Multilayer Percepton Classifiers (f-AMLPs): Fig. 2(a) illustrates the full *f-AMLP* architecture, Fig. 2 (b-d) show the building blocks of *f-AMLP*, including nonlinear circuits. Fig. 2 (e) shows a design for a power delivery network providing stable power to *f-AMLP* primitives.

a) Resistor Crossbars: The resistor crossbar array for flexible neuron, as shown in Fig. 2(a), generates an output voltage by combining multiple inputs. The output can be calculated using basic Ohm's Law and Kirchhoff's Laws as:

$$V_z = \frac{g_1}{G}V_0 + \frac{g_2}{G}V_1 + \frac{g_3}{G}V_2 + \frac{g_b}{G}V_b, \tag{1}$$

where g_i represents individual conductances and G is the total conductance. By selecting specific resistor values, the circuit calculates a weighted sum of input voltages.

b) Flexible Negative Weight Circuit: Resistor crossbars alone can only produce positive weights. To handle negative weights, inverter-based circuits are used, as shown in Fig. 2(c). These circuits invert input voltages to represent negative weights, following a modified negative tanh function:

$$\mathrm{neg}(V_z) = -\left(\eta_1^N + \eta_2^N \cdot \tanh\left(\left(V_z - \eta_3^N\right)\cdot \eta_4^N\right)\right). \tag{2}$$

where $\boldsymbol{\eta}^N = [\eta_1^N, \eta_2^N, \eta_3^N, \eta_4^N]$ are auxiliary parameters that modify the original tanh function, which is ultimately determined by the physical quantities $\boldsymbol{q}^N = [R_1^N, R_2^N, R_3^N, W_1^N, L_1^N, W_2^N, L_2^N, W_3^N, L_3^N]$ in the circuit. Here, W_i^N and L_i^N are geometric features (width and length) of the

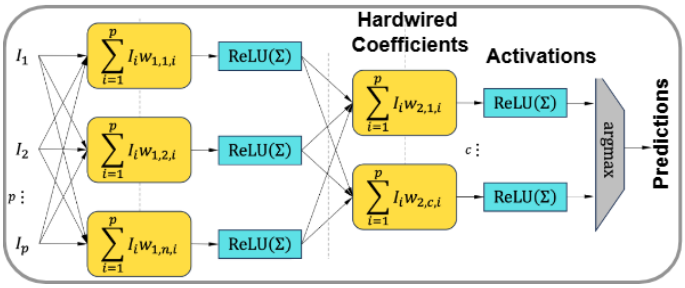


Fig. 3: Bespoke architecture of *f-DMLP* [17].

transistor T_i^{N}. Here, the superscript $(\cdot)^{\mathrm{N}}$ denotes the variables in the Negative weight circuits. The shape of this function can be customized by changing the circuit parameters, allowing fine-tuning for specific tasks.

c) Flexible Activation Circuits: After processing by the resistor crossbars, signals pass through flexible activation circuits that mimic the activation functions commonly used in neural networks (NNs). The activation circuit, shown in Fig. 2(c), uses a parameterized tanh function:

$$V_{\mathrm{a}} = \mathrm{ptanh}(V) = \eta_1^{\mathrm{A}} + \eta_2^{\mathrm{A}} \cdot \tanh\left(\left(V - \eta_3^{\mathrm{A}}\right) \cdot \eta_4^{\mathrm{A}}\right), \quad (3)$$

with the auxiliary parameters $\boldsymbol{\eta}^{\mathrm{A}} = [\eta_1^{\mathrm{A}}, \eta_2^{\mathrm{A}}, \eta_3^{\mathrm{A}}, \eta_4^{\mathrm{A}}]$ determined by $\boldsymbol{q}^{\mathrm{A}} = [R_1^{\mathrm{A}}, R_2^{\mathrm{A}}, W_1^{\mathrm{A}}, L_1^{\mathrm{A}}, W_2^{\mathrm{A}}, L_2^{\mathrm{A}}]$. Similarly, $\boldsymbol{q}^{\mathrm{A}}$ can also be trained to fit specific target tasks. Also, the superscript $(\cdot)^{\mathrm{A}}$ denotes the variables in the tanh circuits.

2) Flexible Digital Multilayer Percepton Classifiers (f-DMLPs): Low-cost embedded machine learning (ML) systems often target specific tasks, which leads to a requirement for specialized ML circuit designs. Previous work [18] has shown that customizing low-cost large-area electronic circuits for specific ML models and datasets can greatly improve efficiency. Despite these improvements, the circuits created were still not suitable for practical large-area electronics due to high area and power usage [19]. Approximate computing (AxC) techniques as shown in Fig. 3 help reduce hardware requirements but usually lead to reduced accuracy. Building upon this, another approach in [11] combined bespoke design with approximation techniques for essential components like multipliers, accumulators, and activation functions. Weights are quantized to powers of two, allowing multiplication to be replaced with simple bit-shifts or interconnect rewiring, thereby eliminating dedicated multipliers in the multiply–accumulate (MAC) units [11]. Accumulation is performed using low-complexity approximate adders [10], which further reduce area and power while maintaining acceptable accuracy for the target tasks. This approach resulted in substantial savings in area and power, making complex ML classifiers feasible for large-area electronics.

C. Side-Channel Analysis

Side-channel analysis (SCA) involves extracting sensitive information from cryptographic or computational hardware implementations through unintended physical leakage channels, including power consumption, electromagnetic emissions, timing variations, or acoustic signals. These leakage patterns unintentionally reveal internal states or operations, allowing attackers to recover secret parameters, inputs, or processed data. Traditionally, statistical methods such as correlation power analysis (CPA) and template attacks have been widely used, relying heavily on explicit leakage models and assumptions of linear or easily-modeled leakage relationships.

Recently, ML and DNN-based SCA have emerged as powerful tools capable of capturing subtle, and complex leakage behaviors. ML-based approaches can autonomously identify and exploit unique patterns within power or electromagnetic traces, even under noisy and unpredictable measurement conditions. Various regression models and classification frameworks have shown superior performance in parameter recovery, significantly outperforming traditional statistical techniques. The combination of FE's environmental sensitivities and the enhanced capabilities of neural-network-driven SCA underscores the critical need to assess vulnerabilities.

D. Related Works

Previous studies have demonstrated significant vulnerabilities of ML classifiers to SCA. For instance, [20] recovered NN weights using power analysis; [21] evaluated masking techniques for SCA resistance; [22] proposed OpenSCA to assess ML leakage; [23] extended this to CIM and NVM-based architectures [24], revealing memory-centric attacks.

However, side-channel vulnerabilities in large-area electronics like FE-based ML classifiers remain highly underexplored. Unique features such as lack of rigid packaging, flexible substrates, and strain sensitivity may increase leakage. PDN design plays a vital role in leakage suppression [25], yet stable PDNs are difficult to implement in FE. Parallel studies have addressed reliability and variability in large-area electronics [4, 16], but dedicated security related study in this domain are still lacking.

III. Proposed Side Channel Attack Methodology

A. Threat Model

We consider an adversary performing a non-invasive power-based side-channel attack on flexible bespoke ML classifiers. We assume the attacker can physically access the targeted FE device during normal operation and passively measure power consumption traces using standard measurement equipment (e.g., oscilloscope). The attacker does not have direct, invasive access to the device's internal circuitry or memory. Furthermore, we assume the attacker possesses knowledge about the general architecture of the classifier and the ability to collect a sufficient number of power measurements under varying operational conditions.

The primary goal of the attacker is to recover sensitive intellectual property (IP) embedded within the device, specifically the analog input voltage and inference outcomes. Successful extraction of these parameters and intermediate results from the measured power traces would compromise both the confidentiality of the processed data and the proprietary nature of the classifier design. The practical relevance of this threat model aligns with realistic scenarios involving flexible ML classifiers used in sensitive applications, such as medical diagnostics, wearable technologies, and secure IoT devices.

B. CPA-based Attack on f-DMLP Classifiers

1) Characterization of Side-Channel Leakage: Digital implementations of *f-DMLP* classifiers on FE platforms inherently simplify computational tasks through techniques such as multiplier approximations, gate pruning, and weight quantization. These simplifications lead to distinct and predictable power consumption patterns that strongly correlate with logical state transitions in digital circuits. Consequently, the leakage characteristics in *f-DMLP* classifiers manifest as deterministic switching activities, inherently less complex and more directly observable compared to the continuous and nonlinear leakage in analog implementations.

2) CPA Attack and Leakage Extraction: For CPA attacks, we adopt the Hamming distance (HD) leakage model, which accurately captures the power consumption variations resulting from transitions between digital logic states. In our CPA-based attack methodology, we record power measurements from the FE-based *f-DMLP* classifiers during their operation. These power traces inherently encode the switching activities that correspond directly to internal computational states.

Mathematically, the CPA procedure involves computing Pearson correlation coefficients between the collected power traces $P(t)$ and the predicted Hamming distance values HD_i for each hypothesized state transition.

$$\rho_i = \frac{\text{Cov}(P(t), HD_i)}{\sigma_{P(t)} \cdot \sigma_{HD_i}}, \tag{4}$$

where ρ_i represents the correlation coefficient for the i-th hypothesis, Cov denotes the covariance, and σ indicates the standard deviation. High correlation values clearly indicate successful identification of internal digital states or parameter values. This direct correlation approach efficiently extracts leakage information without the computational overhead associated with machine learning techniques. The detailed correlation results validating this methodology are presented in Sec. IV.

C. ML-based Attack on f-AMLP Classifiers

1) Profiling of Side-Channel Leakage: Flexible analog classifiers implemented with TFT technology inherently exhibit intricate leakage characteristics arising from complex nonlinear transistor behaviors, including device-specific threshold voltage shifts, nonlinear current-voltage (I-V) characteristics, and parasitic capacitances. Additionally, the leakage signals from these devices are significantly influenced by mechanical deformation and environmental factors such as temperature and humidity, adding to their unpredictability and complexity.

2) Unique Exploitation of Leakage via CNN-based Regression: To effectively exploit such subtle and highly nonlinear leakage signals, traditional statistical methods such as CPA or ML models such as logistic regression or support vector machines (SVM) are insufficient due to their inherently linear or shallow nonlinear decision boundaries. Conversely, our regression model leverages convolutional neural networks' intrinsic hierarchical structure and deep feature extraction capability to capture complex correlations within the leakage signals effectively.

Moreover, the CNN regression model uniquely identifies critical leakage signatures that correspond to specific input analog voltages processed by the TFT-based classifier. By employing convolutional layers, our model extracts spatial-temporal leakage patterns at multiple resolutions. This capability can be mathematically described as follows:

$$\mathbf{Y}^{(l)} = f\left(\mathbf{W}^{(l)} * \mathbf{Y}^{(l-1)} + \mathbf{b}^{(l)}\right), \tag{5}$$

where $\mathbf{Y}^{(l)}$ represents the extracted feature maps at layer l, $\mathbf{W}^{(l)}$ denotes convolutional kernels optimized to isolate leakage-specific features, $\mathbf{b}^{(l)}$ represents biases, and $f(\cdot)$ denotes nonlinear activation functions. Through successive layers, the CNN transforms noisy power traces into a refined set of features directly correlated to analog inputs, which a simpler ML model cannot achieve due to their limited modeling capabilities.

3) Training, Modeling, and Validation for CNN-based Attack: Our CNN architecture for FE leakage signals uses several convolutional layers with increasing feature extraction capabilities, nonlinear ReLU activation functions, and dropout layers with a selected dropout rate of 0.3 to mitigate overfitting. During training, we employ a supervised regression approach using a comprehensive dataset generated from SPICE simulations, which accurately represent realistic leakage patterns including controlled Gaussian and PDN-induced noise.

The training process involves iterative back-propagation using the Adam optimizer, optimizing network parameters (kernels, biases, and dropout probabilities) to minimize the mean squared error (MSE) between the predicted and actual input voltages. A validation subset of our generated traces is consistently monitored to prevent overfitting and ensure robust generalization of the trained model. After training, the model's effectiveness is quantitatively validated through extensive testing, showing high regression accuracy and minimal residual error distribution. This rigorous training, modeling, and validation process ensures the CNN's precision and reliability in recovering inputs from noisy and distorted *f-AMLP* leakage, confirming our successful CNN-based side-channel attack.

IV. Evaluation

A. Experiment Setup

To assess the effectiveness of the proposed method, the experimental analysis of *f-AMLP* was performed using SPICE simulations in Cadence Virtuoso[2] using the Pragmatic *FlexIC Gen3* PDK [26] combined with Python [27]-based processing and modeling scripts. The *f-DMLP* were gate-level netlists mapped to flexible logic cells; dynamic power was obtained from vector-based (VCD/FSDB) switching activity. Hardware costs are tabulated in Tab. I.

PDN modeling: We included a PDN as a distributed RC ladder (Fig. 2 (e)), capturing rail resistances and on-board decoupling. We used $R^{\text{P}}_{mn} = 6\,\text{k}\Omega$, $R_{mn} = 10\,\text{k}\Omega$, and $C_{mn} = 30\,\mu\text{F}$ for the supply path and decoupling network, and sweep these within process tolerances during sensitivity analysis. The PDN was placed in series with *VDD* and the

[2]https://www.cadence.com/en_US/home.html

classifier load; instantaneous supply current $i_{DD}(t)$ and power traces were recorded at the PDN output node.

TABLE I: Hardware Costs of Bespoke *f-MLPs*

Dataset	Topology	Analog (*f-AMLP*)			Digital (*f-DMLP*))		
		Area (mm^2)	Power (μW)	Acc. (%)	Area (mm^2)	Power (μW)	Acc. (%)
Iris	4-8-3	0.04	3.39	96.51	4.831	72.01	95.43
Cardio	21-12-3	0.09	8.81	86.81	1.60	240.40	85.11
Pendigits	16-10-10	0.07	7.17	53.44	3.21	410.01	89.60
Seeds	7-4-3	0.04	3.14	88.10	6.20	110.70	86.21

1) Attack Setup of f-DMLP: We performed CPA on the *f-DMLP* classifier. Initially, the design was constrained to incorporate multiplier approximations, gate pruning, and weight quantization [10], inspired by standard techniques employed in silicon-based digital designs to simplify the computational overhead. Power traces were captured under controlled experimental conditions, ensuring consistency in environmental parameters and measurement setups. Using the HD leakage model, we computed predicted switching activities corresponding to state transitions within the digital circuits. Pearson correlation coefficients were then calculated between these predicted switching behaviors and the experimentally captured power traces. The CPA parameters, such as the trace length, the number of traces, and statistical significance criteria, adhered closely to standard practices from digital VLSI security evaluations.

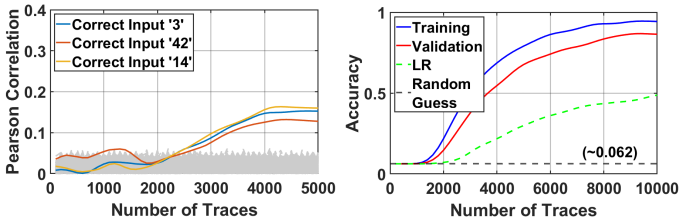
2) Attack Setup of f-AMLP:

a) Dataset Generation: We extracted data dependent power measurement consisting of 10,000 samples to realistically emulate side-channel leakage from *f-AMLP* classifiers. Each input sample represented analog input voltages uniformly distributed between 0 and 1. Corresponding power traces were modeled using nonlinear leakage functions characteristic of *f-AMLP* circuit primitives, i.e. nonlinear circuits (see Fig. 2), including polynomial relationships. To reflect realistic measurements and environmental variability, we injected controlled Gaussian noise and introduced random outliers into the dataset, comprising $\approx$ 5% of total samples.

b) Training Setup of f-AMLP: The dataset was divided into train (60%), test (20%) and validation (20%) sets to facilitate unbiased evaluation of the side-channel attack. For modeling, we developed a CNN-based regression attack comprising two convolutional layers (kernel size 5) followed by Rectified Linear Unit (ReLU) activations and dropout layers (0.3 dropout rate) to enhance model robustness and generalizability. The CNN was trained using the Adam optimizer (learning rate 0.001) with a mini-batch size of 64 over 50 epochs, monitoring training and validation loss to ensure stable convergence and evaluate model effectiveness.

B. CPA based SCA results on f-DMLP classifiers

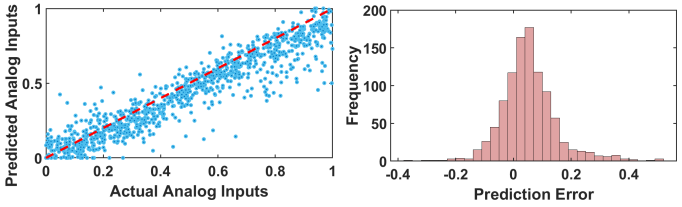
Correlation plots of the CPA-based attack for *f-DMLP* classifiers (Fig. 4a) illustrate how clearly the model exploits power-leakage signals. Pearson correlation values for correct input predictions initially exhibit minimal improvement due to noise and fewer traces. However, after $\approx$ 1,500 traces, they distinctly separate from incorrect inputs, rapidly increasing and gradually saturating near moderate correlation values (around 0.2-0.25).



(a) Correlation progression on digital *f-DMLP* classifiers, illustrating gradual separation of correct inputs' (3 among 21) correlations from incorrect inputs (shown in grey), highlighting the attack.

(b) CNN-based regression accuracy (training vs. validation) on analog *f-AMLP* classifiers, demonstrating effective generalization and robustness despite measurement noise for Iris dataset.

Fig. 4: Comparison of side-channel attack effectiveness on analog and digital FE classifiers (Iris dataset).



(a) Correlation showing successful recovery of input voltages from side-channel power traces, highlighting the vulnerability.

(b) Distribution of residual prediction errors showing concentrated accuracy around zero, validating model precision and reliability.

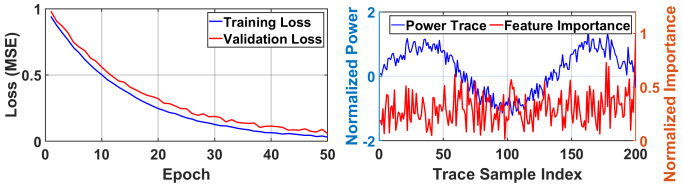
Fig. 5: Robustness of CNN-based side-channel regression attack on analog *f-AMLP* implementation (Iris dataset).

The clear separation after sufficient trace collection demonstrates CPA's effectiveness in identifying switching activities associated with internal logic states. These results validate that digital implementations exhibit predictable and effectively exploitable leakage characteristics via CPA, thus highlighting their practical vulnerability.

C. CNN-based SCA on f-AMLP classifiers

1) Attack Robustness : In contrast to the discrete switching behavior of digital designs, *f-AMLP* implementations exhibit complex, nonlinear leakage patterns shaped by TFT nonidealities, parasitic effects, and environmental variability. Fig. 4b shows training and validation accuracy trends for our CNN-based regression model attacking an *f-AMLP* classifier. The model gradually learns to separate correct analog input predictions from incorrect ones despite substantial noise, achieving robust generalization beyond 8,000 traces. This slow initial accuracy rise, followed by rapid escalation, highlights the need for advanced profiling techniques to extract weak but information-rich features in FE-specific measurement noise.

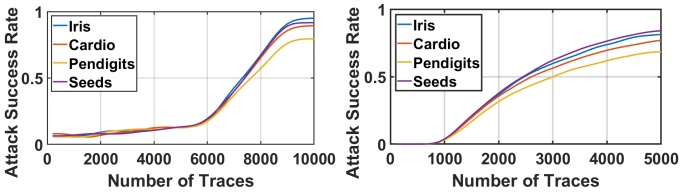
Further validation is provided in Fig. 5, where the scatter plot (Fig. 5a) presents the predicted versus actual analog input voltages obtained from our CNN-based regression model for the *f-AMLP* classifier. The dense clustering of points along the diagonal line indicates that the predicted values closely match the ground truth throughout the input range, with minimal bias or variance. Even at the extremes of the analog input domain, where non-idealities and PDN-induced voltage drops are most pronounced, predictions remain closely aligned with the ideal



(a) Learning curves showing CNN training and validation loss across epochs during regression-based modeling. Rapid initial convergence demonstrates effective learning of analog leakage patterns.

(b) Saliency map illustrating feature importance of CNN regression model on input power traces. High-importance regions reveal specific segments of the power trace most influential for accurate analog input prediction.

Fig. 6: Attack progress on *f-AMLP* classifiers (Iris dataset).



(a) ASR progression for CNN-based attacks on analog *f-AMLP* classifiers, highlighting slow initial growth followed by rapid escalation after a few hundred traces, indicative of complex but exploitable analog leakage.

(b) CPA-based cumulative ASR on *f-DMLP* classifiers, demonstrating distinct dataset-specific trends and faster initial convergence to moderately high values, clearly showing simpler leakage characteristics compared to analog scenarios.

Fig. 7: Attack success rate comparison across diverse datasets.

diagonal, reflecting the model's ability to learn highly nonlinear leakage-to-value mappings.

The residual error histogram (Fig. 5b) provides complementary insight into prediction accuracy and error distribution. The narrow, Gaussian-like distribution sharply centered at zero demonstrates the absence of significant bias in the regression output, while the low spread confirms strong prediction consistency across all tested inputs. The lack of heavy tails in the distribution suggests that the CNN avoids large outlier errors, even under noisy measurement conditions.

2) Attack Progress and Feature Attribution: To better understand how the CNN learns to exploit analog leakage, we examine intermediate training behavior and feature attribution. The loss curves in Fig. 6a show rapid reduction of both training and validation MSE within the first $\approx$ 20 epochs, indicating fast convergence despite high measurement noise. Validation loss closely follows training loss, confirming minimal overfitting and strong generalization.

The saliency map in Fig. 6b provides complementary insight by highlighting power-trace segments most influential to input prediction. High-importance intervals coincide with analog signal transitions in the crossbar summation and activation stages, where nonlinearities and circuit-specific switching generate distinctive leakage signatures. This targeted exploitation of specific time windows underscores the CNN's capability to focus on the most informative segments of the trace, enabling successful attacks in scenarios where traditional statistical methods fail.

D. Attack Success Rate Comparison: Analog vs. Digital

The cumulative attack success rate (ASR) trends for the analog f-AMLP and digital f-DMLP implementations across four benchmark datasets (Iris, Cardio, Pendigits, Seeds) are shown in Fig. 7a and Fig. 7b, respectively. The progression curves highlight a fundamental trade-off between attack speed: the number of traces required to reach a given success probability, and the final achievable accuracy.

For the analog case, CNN-based regression must learn and generalize from complex, noise-rich leakage patterns arising from TFT device nonlinearities, PDN noise, and measurement distortions. Consequently, the ASR grows slowly during the initial $\approx$2,000–3,000 traces, reflecting the difficulty of extracting reliable features under high noise conditions. After this slow start, the model enters a rapid escalation phase between $\approx$3,000 and $\approx$4,700 traces, where learned features begin to strongly correlate with the target inputs. The ASR then continues to improve, eventually saturating at $\geq$90.

In contrast, the digital f-DMLP, targeted with CPA, exhibits a faster initial ASR growth, with most datasets achieving over 45% success within $\approx$1,500–2,000 traces. This rapid convergence is attributed to the discrete nature of switching activity in digital logic, where leakage correlates strongly and consistently with Hamming distance predictions. Saturation occurs earlier at $\approx$4,000–5,000 traces, but at lower final ASR values (70–85%) compared to the analog case. This reflects the ease of early exploitation due to simpler leakage models, yet a ceiling in achievable accuracy since digital leakage is more structured and contains fewer exploitable nonlinear components. This trade-off in attack speed versus final ASR has direct implications for threat modeling in FE, as the optimal defense strategy must consider both early-stage and long-term attack resilience.

V. Conclusion

Flexible electronics offer promising opportunities for low-cost, application-specific neuromorphic computing. However, their inherent susceptibility to variation, environmental stress, and the absence of rigid packaging also introduce side-channel information leakage. In this work, we introduce a robust CNN-based regression attack model used to exploit subtle, nonlinear leakage patterns in flexible analog MLP classifiers and achieve accurate recovery of input voltages, and also compared with state-of-the art digital flexible classifiers. Our results demonstrate that while digital designs can be compromised with relatively few traces, their exploitable leakage limits the attacker's ultimate success. In contrast, analog implementations resist early exploitation but yield stronger information over time, making them more vulnerable to high-trace profiling attacks.

VI. Acknowledgment

This work has been supported by the the European Research Council (ERC) (Grant No. 101052764) .

References

[1] S. Kim, "Inkjet-Printed Electronics on Paper for RF Identification (RFID) and Sensing," *Electronics*, vol. 9, no. 10, p. 1636, 2020.

[2] A. U. Alam *et al.*, "Fruit Quality Monitoring with Smart Packaging," *Sensors*, vol. 21, no. 4, p. 1509, 2021.

[3] Q. Sun *et al.*, "Smart Band-Aid: Multifunctional and Wearable Electronic Device for Self-Powered Motion Monitoring and Human-Machine Interaction," *Nano Energy*, vol. 92, p. 106840, 2022.

[4] P. Pal *et al.*, "Neural Architecture Search for Highly Bespoke Robust Printed Neuromorphic Circuits," in *IEEE/ACM ICCAD '24*, 2024.

[5] H. Zhao, P. Pal, M. Hefenbrock, Y. Wang, M. Beigl, and M. B. Tahoori, "Neural evolutionary architecture search for compact printed analog neuromorphic circuits," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, pp. 1–1, 2024.

[6] P. Pal *et al.*, "Analog printed spiking neuromorphic circuit," in *IEEE DATE*, 2024, p. 6 S.

[7] H. Zhao, M. Hefenbrock, M. Beigl, and M. B. Tahoori, "Highly-dependable printed neuromorphic circuits based on additive manufacturing," *Flexible and Printed Electronics*, vol. 8, no. 2, p. 025018, jun 2023. [Online]. Available: https://dx.doi.org/10.1088/2058-8585/acd8cd

[8] P. Pal, A. Studt, T. Gheshlaghi, M. Hefenbrock, M. Beigl, and M. B. Tahoori, "Spikesynth: Energy-efficient adaptive analog printed spiking neural networks," in *44th ACM/IEEE International Conference on Computer Aided Design (ICCAD 2025)*. Institute of Electrical and Electronics Engineers (IEEE), 2025.

[9] E. Ozer, J. Kufel, J. Biggs, G. Brown, J. Myers, A. Rana, A. Sou, and C. Ramsdale, "Bespoke machine learning processor development framework on flexible substrates," in *2019 IEEE International Conference on Flexible and Printable Sensors and Systems (FLEPS)*, 2019, pp. 1–3.

[10] F. Afentaki, G. Saglam, A. Kokkinis, K. Siozios, G. Zervakis, and M. B. Tahoori, "Bespoke approximation of multiplication-accumulation and activation targeting printed multilayer perceptrons," in *2023 IEEE/ACM ICCAD*. IEEE, Oct. 2023. [Online]. Available: http://dx.doi.org/10.1109/ICCAD57390.2023.10323613

[11] F. Afentaki *et al.*, "Bespoke approximation of multiplication-accumulation and activation targeting printed multilayer perceptrons," in *2023 IEEE/ACM (ICCAD)*, 2023, pp. 1–9.

[12] E. Services, "S9-e3_pragmatic flexics – part 3: Introducing pragmatic flexic platform gen 3," https://www.youtube.com/watch?v=rAQlsL8fR00, Apr. 2025, published April 14, 2025; accessed April 19, 2025.

[13] M. B. Tahoori, E. Ozer, G. Zervakis, K. Balaskas, and P. Pal, "Computing with printed and flexible electronics," in *2025 IEEE European Test Symposium (ETS)*, 2025, pp. 1–9.

[14] P. Pal, F. Afentaki, H. Zhao, G. Saglam, M. Hefenbrock, G. Zervakis, M. Beigl, and M. B. Tahoori, "Fault sensitivity analysis of printed bespoke multilayer perceptron classifiers," in *2024 IEEE European Test Symposium (ETS)*. IEEE, 2024, pp. 1–6.

[15] P. Pal, T. Gheshlaghi, H. Zhao, M. Hefenbrock, M. Beigl, and M. Tahoori, "Print-safe: Printed ultra-low-cost electronic x-design with scalable adaptive fault endurance," *ACM Trans. Embed. Comput. Syst.*, Aug. 2025, just Accepted. [Online]. Available: https://doi.org/10.1145/3758096

[16] H. Zhao *et al.*, "Highly-Bespoke Robust Printed Neuromorphic Circuits," in *Design, Automation and Test in Europe (DATE)*. IEEE, 2023.

[17] G. Armeniakos *et al.*, "Model-to-circuit cross-approximation for printed machine learning classifiers," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol. 42, no. 11, pp. 3532–3544, 2023.

[18] M. H. Mubarik *et al.*, "Printed machine learning classifiers," in *2020 53rd Annual IEEE/ACM MICRO*, 2020, pp. 73–87.

[19] G. Armeniakos *et al.*, "Cross-layer approximation for printed machine learning circuits," in *2022 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2022, pp. 190–195.

[20] L. Batina, S. Bhasin, D. Jap, and S. Picek, "CSI NN: Reverse engineering of neural network architectures through electromagnetic side channel," in *28th USENIX Security Symposium (USENIX Security 19)*. Santa Clara, CA: USENIX Association, Aug. 2019, pp. 515–532. [Online]. Available: https://www.usenix.org/conference/usenixsecurity19/presentation/batina

[21] A. Dubey, R. Cammarota, and A. Aysu, "Maskednet: The first hardware inference engine aiming power side-channel protection," in *2020 IEEE International Symposium on Hardware Oriented Security and Trust (HOST)*, 2020, pp. 197–208.

[22] W. Wei, L. Liu, M. Loper, K.-H. Chow, M. E. Gursoy, S. Truex, and Y. Wu, "A framework for evaluating gradient leakage attacks in federated learning," 2020. [Online]. Available: https://arxiv.org/abs/2004.10397

[23] B. Sapui, J. Krautter, M. B. Mayahinia, and Others, "Power side-channel attacks and countermeasures on computation-in-memory architectures and technologies," in *IEEE European Test Symposium (ETS)*, 2023.

[24] B. Sapui and M. B. Tahoori, "Power side-channel analysis and mitigation for neural network accelerators based on memristive crossbars," in *2024 29th Asia and South Pacific Design Automation Conference (ASP-DAC)*, 2024, pp. 612–617.

[25] R. Selvam and A. Tyagi, "Power distribution network capacitive decoupling for side-channel resistance," in *2021 IEEE International Symposium on Smart Electronic Systems (iSES)*, 2021, pp. 183–188.

[26] Pragmatic, "Flexic Platform Gen3," https://www.pragmaticsemi.com/foundry/flexic-platform-gen-3, 2025.

[27] A. Paszke *et al.*, "Pytorch: An Imperative Style, High-performance Deep Learning Library," in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds. Curran Associates, Inc., 2019, pp. 8024–8035.