ML-DSA-87 : Offensive Top-Down Side-Channel Leakage Assessment Using Pre-Silicon Models

Abstract. Side-channel leakage can break Roots of Trust (post-quantum design components) before silicon. Design-time vulnerability assessment is therefore critical. Architectural choices, toolchain optimizations, and firmware orchestration can create exploitable leakage that is costly to fix post-silicon. We present a hierarchical, offensive pre-silicon methodology that traces causes and effects across functional, firmware, RTL, and gate levels, enabling early detection and clear attribution. In ML-DSA-87, we examine pointwise multiplication (PWM) in signing: the unprotected design is vulnerable to correlation power analysis (CPA), and a two-share masked PWM still exhibits second-order leakage exploitable by higher-order CPA and profiled (template) attacks.

Keywords: Pre-silicon leakage models · Digital Signature · Adams-Bridge accelerator

1 Background

Roots of Trust (RoT). A silicon RoT anchors device identity, secure boot, attestation, and firmware update. Because RoT vouches for the entire platform, compromise at this(physical) layer breaks transitive trust across devices and the supply chain. PQC is critical in this context: quantum-capable adversaries threaten classical signatures and key exchange, putting long-lived identities and attestation paths at risk. However, design-time side channel leakage can erode the credibility of PQC standardization in real RoTs, for example, Caliptra and OpenTitan, by allowing key recovery or bias despite algorithmic soundness [1–3]. Therefore, hardening and verification must occur at design time: pre-silicon, hierarchical side-channel analysis and security sign-off should accompany functional verification to expose the design leakage before tape-out.

Top-down leakage assessment. A hierarchical (top-down) methodology provides two benefits: (i) early discovery during architectural exploration and countermeasure selection; (ii) clear attribution to root causes as effects propagate or disappear through abstraction boundaries. We build on the *Telescope* [4] concept for cross-layer pre-silicon assessment 1 to structure our flow for PQC accelerators.

2 Implementation-Attack Model and Pre-Silicon Equivalence

2.1 How devices leak during crypto

A side channel exposes an information-bearing signal $L(t)$ correlated with sensitive state S while a device executes operations on public data C. Typical sources include power, EM radiation, and timing.

2.2 Power model and pre-/post-silicon alignment

We decompose instantaneous power as

$$P_{\text{tot}}(t) = P_{\text{static}} + P_{\text{Dynamic}} = P_{\text{static}} + (P_{\text{noise}} + P_{\text{op}}(t) + P_{\text{chg}}(t) + P_{\text{data}}(t)),$$

where P_{noise} summarizes electronic/measurement noise and environmental terms, P_{op} captures operation scheduling overheads (clock gating, FSM steps), P_{chg} covers charging/discharging overheads

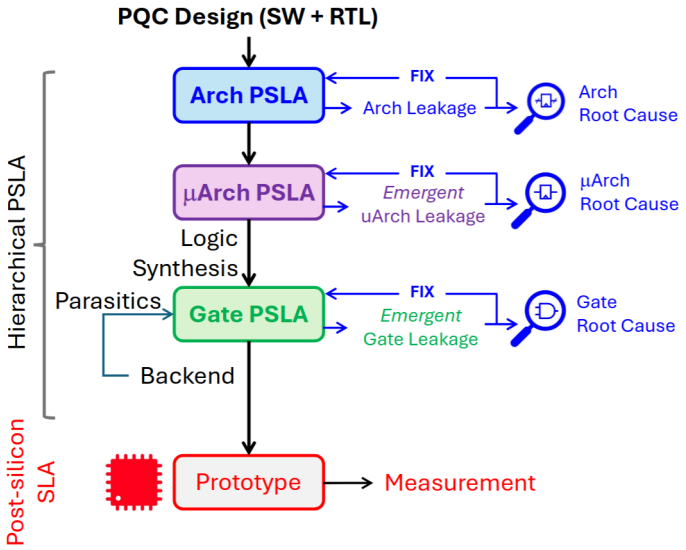

Fig. 1. Overview of the hierarchical pre-silicon leakage assessment flow.

largely invariant for fixed micro-architectural activity, and P_{data} is the data-dependent component carrying secret correlation.

In general, dynamic power is well-approximated by

$$P_{\text{dyn}} = \alpha C_{\text{eff}} V^2 f, \tag{1}$$

where α is the switching activity factor, C_{eff} is effective load capacitance, V the supply voltage, and f the clock frequency. In cycle-accurate simulation, per-register toggle counts (TC) and switching activity in SAIF/VCD act as pre-silicon proxies for α; except for noise, these proxies preserve the same functional dependence on data as post-silicon power/EM traces. Thus, by varying input/state data while holding non-data terms approximately constant, pre-silicon toggle/activity features capture the exploitable variations present post-silicon.

Why pre-silicon? Pre-silicon is where structural fixes (algorithmic masking, share refreshing, register allocation, routing constraints) are still affordable and verifiable. Post-silicon often reveals entangled phenomena (glitches, placement/routing, firmware timing) that are *expensive* to remediate. However, PQC compounds difficulty: large datapaths (NTT/PolyVec), long-running protocols, and heterogeneity (firmware + accelerator) magnify leakage channels and interactions.

2.3 Top-down offensive analysis

Leakage can be *progressive* (amplifies down the stack), *disappearing* (canceled by scheduling), or *emergent* (introduced by synthesis/P&R). A top-down offensive approach tests attacker success at each layer and traces root causes across boundaries (functional→firmware→RTL→gate). We follow the *Telescope*-style layering but adapt it for PQC accelerators at RoT scale [4].

3 Case Study: ML-DSA-87 Signing—Pointwise Multiplication (PWM)

Signing overview. ML-DSA-87 signing computes the expanded matrix $\hat{A}$ from ρ, samples an ephemeral vector $\mathbf{y}$, and forms $\mathbf{w} = A\mathbf{y}$; it then derives $\mathbf{w}_1 = \text{HighBits}(\mathbf{w}, 2\gamma_2)$, hashes $(\mu, \mathbf{w}_1)$ to a challenge seed, samples c, sets $z = \mathbf{y} + c\mathbf{s}_1$, and emits (ς, z, h) after standard rejection checks on $\|z\|_\infty$

Algorithm 1 Signature Generation

1: $\hat{A} \leftarrow \text{ExpandA}(\rho)$ ▷ A has size $k \times \ell \times \mathbb{R}_q$, derived from ρ.
2: $\mu \leftarrow H(\text{tr} \parallel M)$ ▷ 512-bit message hash with $H(\text{PK})$ prefix.
3: $\kappa \leftarrow 0, \quad (z, h) \leftarrow \perp$ ▷ Iteration counter and result tuple.
4: $\rho' \leftarrow \text{random}$ **or** $H(\kappa, \mu)$ ▷ Use hash in deterministic signing.
5: **while** $(z, h) = \perp$ **do** ▷ *Rejection loop*
6: $\quad \mathbf{y} \leftarrow \text{ExpandMask}(\rho', \kappa, \ldots)$ ▷ $\mathbf{y} \in \ell \times \mathbb{R}_q$ sampled from $[-\Upsilon_1, +\Upsilon_1]$.
7: $\quad \mathbf{w} \leftarrow A * \mathbf{y}$ ▷ $\mathbf{w} = \text{NTT}^{-1}\big(\circ\text{NTT}(\mathbf{y})\big)$.
8: $\quad \mathbf{w}_1 \leftarrow \text{HighBits}(\mathbf{w}, 2\gamma_2)$ ▷ Range $\approx (q-1)/(2\gamma_2)$, e.g., $[0, 15]$ or $[0, 43]$.
9: $\quad \varsigma \leftarrow H(\mu, \mathbf{w}_1)$ ▷ ς derived from message and public key.
10: $\quad c \leftarrow \text{SampleInBall}(\varsigma)$ ▷ $c \in \mathbb{R}_q$ with τ nonzero (± 1) coefficients.
11: $\quad \mathbf{z} \leftarrow \mathbf{y} + c \cdot \mathbf{s}_1$ ▷ Prefer storing $\text{NTT}(\mathbf{s}_1)$ as shares.
12: $r_0 \leftarrow \text{LowBits}(\mathbf{w} - c \cdot \mathbf{s}_2, 2\gamma_2)$ ▷ Range is basically $\pm 2\gamma_2$.
13: $\quad$ **if** $\text{MaxAbs}(z) \geq \Upsilon_1 - \beta$ **or** $\text{MaxAbs}(r_0) \geq \gamma_2 - \beta$ **then**
14: $\quad\quad (z, h) \leftarrow \perp$ ▷ reject
15: $\quad$ **else**
16: $\quad\quad h \leftarrow \text{MakeHint}\big(-c \cdot \mathbf{t}_0, \ \mathbf{w} - c \cdot \mathbf{s}_2 - c \cdot \mathbf{t}_0', \ 2\gamma_2\big)$ ▷ $h \in \{0, 1\}^{kN}$.
17: $\quad\quad$ **if** $\text{MaxAbs}(c \cdot \mathbf{t}_0) > \gamma_2$ **or** $\text{CountOnes}(h) > \omega$ **then**
18: $\quad\quad\quad (z, h) \leftarrow \perp$ ▷ reject
19: $\quad\quad \kappa \leftarrow \kappa + 1$ ▷ Advance to create fresh $\mathbf{y}$ next loop.
20: **end while**
21: **return** $\text{Sig} = (\varsigma, z, h)$ ▷ Now non-secret.

and $\text{LowBits}(\mathbf{w} - c\mathbf{s}_2)$. In the NTT formulation, $A\mathbf{y}$ is realized as $\text{NTT}^{-1}\big(\circ\text{NTT}(\mathbf{y})\big)$, with integer reductions applied after pointwise multiplication and at recombination boundaries (Algorithm 1).

Implementation perspective. In ML-DSA-87, the accelerator can be realized in either unprotected or protected form. In Algorithm 1, line 11 (highlighted conceptually in red) is particularly prone to side-channel leakage. The signing operands c and $\mathbf{s}_1$ are polynomials with 250 coefficients each, and the datapath instantiates four NTT modules operating in parallel to accelerate polynomial transforms; each polynomial vector is converted into the NTT domain, followed by pointwise multiplication and modular reduction, before an NTT^{-1} brings results back to the coefficient domain. For tractable and reproducible pre-silicon analysis, we isolate a single pointwise-multiplication (PWM) micro-operation within the transform pipeline: first in an unprotected design to establish side channel leakage baselines, and then in a two-share masked design (Fig. 2), where operands and intermediates are split and recombined with fresh randomness to study second-order effects. The masked PWM microarchitecture maintains strict share separation, fixed-latency reducers, and randomized re-masking between stages, enabling controlled experiments on leakage amplification, disappearance, and emergence across abstraction levels.

4 Leakage Models and Assessment

We use cycle-accurate simulation to extract register-level toggles (TCs). Let $\text{TC}_i[k]$ denote the toggle count of register i within sampling window k (e.g., one micro–operation or sub–operation). All TC features are centered per experiment to remove DC offsets and slow drift.

4.1 RTL Level: Unprotected PWM

Leakage model. We target TC_{acc}, the TC of the register that holds the intermediate $c \cdot s_1$. For CPA we build hypothetical intermediates $H(c, s_1^\star)$ for each key hypothesis $s_1^\star$, map them with a

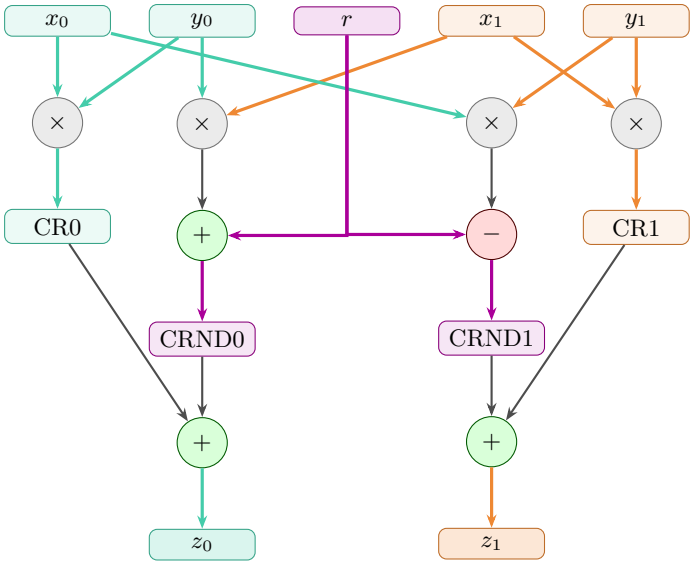

Fig. 2. Two-share masked pointwise multiplication and its datapath (two cycles).

prediction function $\Phi(H)$ (e.g., centered Hamming weight, toggle count of the register), and compute the Pearson correlation

$$\rho(s_1^\star) \;=\; \mathrm{corr}\big(\Phi(H(c, s_1^\star)),\ \mathrm{TC}_{\mathrm{acc}}\big). \tag{2}$$

Expectation. The correct s_1 maximizes ρ as traces increase; the rank of s_1 therefore improves monotonically with the number of traces as shown in the Figure 3.

4.2 RTL Level: Two-Share Masked PWM

Second-order CPA (2oCPA) exploits joint statistics of two leakage samples to defeat first-order masking. We adopt a covariance-based formulation using TC features measured at two (aligned) instants.

Setup and notation. Let C_i be the public input for trace i (mean 0, variance $\mathrm{Var}(C)$), S_1 the fixed secret share, $m_{1,i}, m_{2,i}$ independent zero-mean masks, and r_i fresh randomness with $\mathbb{E}[r_i] = 0$. Shares and partial products are

$$\begin{aligned} x_{0,i} &= m_{1,i}, & x_{1,i} &= S_1 - m_{1,i}, & y_{0,i} &= m_{2,i}, & y_{1,i} &= C_i - m_{2,i}, \\ \mathrm{reg}_{0,i} &= x_{0,i}y_{0,i}, & \mathrm{reg}_{1,i} &= x_{1,i}y_{1,i}, & \mathrm{rnd}_{0,i} &= x_{1,i}y_{0,i} + r_i, & \mathrm{rnd}_{1,i} &= x_{0,i}y_{1,i} - r_i, \end{aligned}$$

and outputs $z_{0,i} = \mathrm{reg}_{0,i} + \mathrm{rnd}_{0,i}$, $z_{1,i} = \mathrm{reg}_{1,i} + \mathrm{rnd}_{1,i}$.

Local linear TC model. For sufficiently wide registers, the TC observable is well approximated by a local linear model [5]:

$$t_{1,i} = \alpha\, s_i + n_{1,i}, \qquad t_{2,i} = \beta\, s_i + n_{2,i}, \qquad s_i \triangleq S_1\, C_i, \tag{3}$$

where gains $\alpha, \beta > 0$ model path asymmetry and $n_{1,i}, n_{2,i}$ are zero-mean noises (glitches, quantization, measurement). Define centered variables $\tilde{t}_{k,i} = t_{k,i} - \overline{t}_k$ and $\tilde{s}_i = s_i - \overline{s}$ with $\overline{(\cdot)}$ denoting the empirical mean.

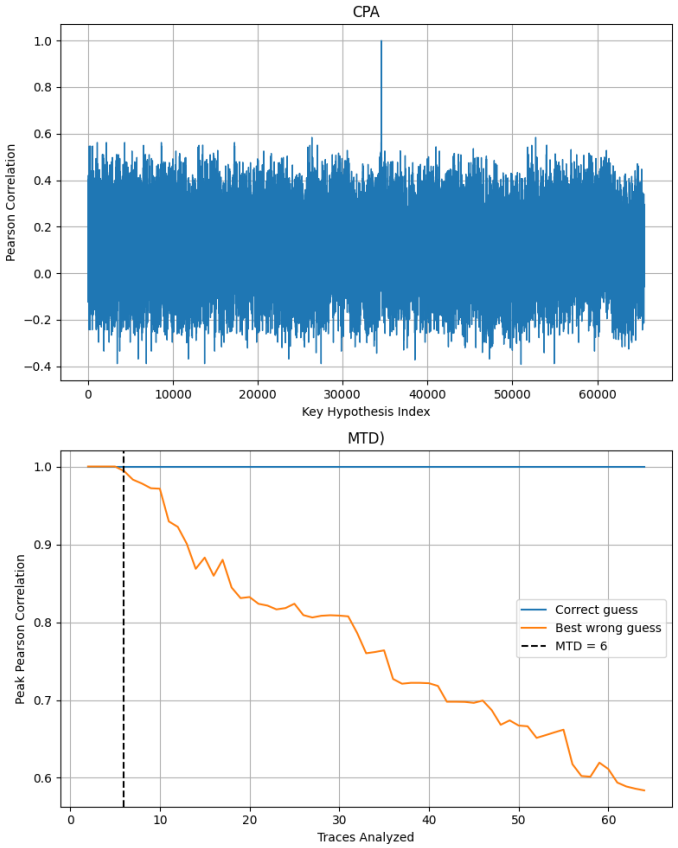

Fig. 3. *Unprotected PWM at RTL.* (a) Correlation vs. number of traces for CPA; the correct key guess is `0x345678`. (b) Key-rank trend shows an MTD of 6 traces.

Second-order features. We build three standard second-order features from $t_{1,i}, t_{2,i}$:

Cross (product).

$$L_{\times,i} \triangleq \tilde{t}_{1,i}\,\tilde{t}_{2,i}, \qquad \mathbb{E}[L_{\times,i} \mid C_i] = \alpha\beta\,\tilde{s}_i^{\,2} + \mathrm{Cov}(n_{1,i}, n_{2,i}). \tag{4}$$

Up to an additive constant, $L_\times$ is proportional to $\tilde{s}^{\,2}$, thereby canceling first-order leakage.

Absolute difference.

$$D_i \triangleq t_{1,i} - t_{2,i} = (\alpha - \beta)s_i + (n_{1,i} - n_{2,i}), \qquad L_{|\Delta|,i} \triangleq |D_i| - \mathbb{E}[|D_i|]. \tag{5}$$

If $n_{1,i}, n_{2,i} \sim \mathcal{N}(0, \sigma^2)$ are i.i.d., then $\mathbb{E}[|D_i| \mid C_i] \approx \sqrt{\frac{2}{\pi}}\sqrt{(\alpha-\beta)^2 s_i^2 + 2\sigma^2}$, monotone in $|s_i|$; after centering, $L_{|\Delta|}$ carries information roughly proportional to s_i^2 and is particularly sensitive to gain mismatch ($\alpha \neq \beta$).

Energy (sum of squares).

$$L_{E,i} \triangleq \tilde{t}_{1,i}^{\,2} + \tilde{t}_{2,i}^{\,2}, \qquad \mathbb{E}[L_{E,i} \mid C_i] = (\alpha^2 + \beta^2)\,\tilde{s}_i^{\,2} + \mathrm{Var}(n_{1,i}) + \mathrm{Var}(n_{2,i}). \tag{6}$$

Apart from a constant offset, L_E scales with $\tilde{s}^{\,2}$ and is robust to sign flips of s_i.

Equations (4)–(6) show that, *after centering*, all three features are affine in $\tilde{s}_i^{\,2}$ (up to model-specific constants). This property enables second-order key recovery.

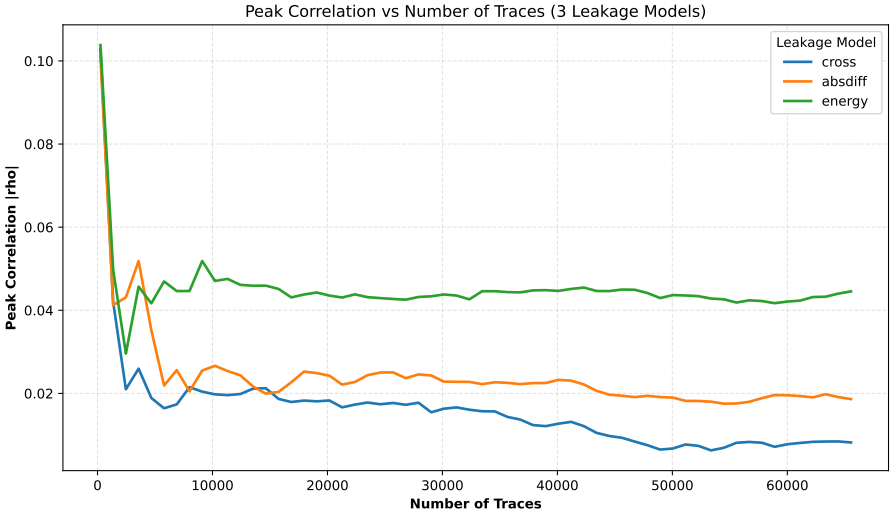

Fig. 4. *Masked PWM at RTL*. Second-order CPA using the *cross*, *absdiff*, and *energy* features: peak |corr| vs. traces.

Predictors and scoring. Let k be a key hypothesis and $p_i(k) = k\,C_i$ with centered $\tilde{p}_i(k)$. To connect gate-level switching to scalar features, we use:

Bit-sliced predictors $h_b(i;k) = ((p_i(k) \gg b)\ \&\ 1) - \Pr[h_b = 1]$ for bit $b \in \{0, \ldots, B-1\}$ and score

$$\hat{\rho}_\bullet(k) \triangleq \max_b \Big|\mathrm{corr}\big(L_{\bullet,i},\, h_b(i;k)\big)\Big|, \qquad \bullet \in \{\times, |\Delta|, E\}. \tag{7}$$

This matches the bit-flip nature of switching and usually yields the sharpest discrimination [6].

Amplitude-like predictors such as $g(i;k) = |\tilde{p}_i(k)|$ or $g(i;k) = (\tilde{p}_i(k))^2$ are even in $\tilde{p}_i(k)$ and correlate with $L_\bullet$ because $\mathbb{E}[L_{\bullet,i} \mid C_i]$ is affine in $\tilde{s}_i^2$:

$$\mathrm{corr}\big(L_{\bullet,i},\, g(i;k)\big) \ \propto\ \mathrm{Cov}\big(\tilde{s}_i^{\,2},\, g(i;k)\big).$$

In practice we rank keys by decreasing $|\hat{\rho}_\bullet(k)|$ from (7).

4.3 2oCPA Statistic

For a chosen feature $L_\bullet$ and predictor $\Psi \in \{h_b, g\}$, the second-order correlation statistic is

$$\rho_\bullet(k;\Psi) \triangleq \frac{\sum_{i=1}^{N} \big(L_{\bullet,i} - \overline{L}_\bullet\big)\,\big(\Psi(i;k) - \overline{\Psi}(k)\big)}{\sqrt{\sum_i \big(L_{\bullet,i} - \overline{L}_\bullet\big)^2}\,\sqrt{\sum_i \big(\Psi(i;k) - \overline{\Psi}(k)\big)^2}}. \tag{8}$$

Under the assumptions above (independent zero-mean noises, non-degenerate C), the expected value of $\rho_\bullet(k;\Psi)$ is maximized at the true key $k = S_1$; empirically, the rank of S_1 improves with the number of traces as highlighted in the Figure 4.

Profiled (Template) Attack—Gaussian Bit-Mode

Profiling assumptions. We assume an attacker with full access to a pre-silicon model can generate profiling traces under random (s_1, c) pairs and collect a feature vector $\mathbf{l} = [\mathrm{TC}_1, \ldots, \mathrm{TC}_6]^\top$ of

Table 1. Distinguishers and minimum traces to disclosure (MTD) for the ML-DSA-87 signing PWM. The unprotected RTL shows immediate first-order leakage: standard CPA recovers the subkey in *six* traces. With two-share masking, first-order leakage is suppressed; however, second-order structure remains: 2oCPA against our centered TC features succeeds at scale ($\approx$ *10k* traces for a 23-bit subkey). A profiled Gaussian bit-mode template further improves efficiency by modeling multi-register structure: the correct key appears within the top-10 candidates after *50* attack traces in our setup. These results highlight (i) why masking must be paired with alignment/gain control and share-isolation policies, and (ii) why pre-silicon, attacker-in-the-loop evaluation is essential for sign-off.

Abstraction	Implementation	Distinguisher	MTD (traces)
RTL	Unprotected PWM	CPA (1st order)	6
RTL	Protected PWM (two-share)	2oCPA	10 k
RTL	Protected PWM (two-share)	Template (Gaussian)	50

register-level toggle counts (six registers covering the two share-coupled datapaths and auxiliaries). For each class u—e.g., the value of a selected bit of the key-dependent intermediate $P = S_1 \cdot C$ ("Gaussian bit-mode")—we fit a Gaussian template

$$\mathbf{l} \,\big|\, \mathcal{C} = u \;\sim\; \mathcal{N}\big(\boldsymbol{\mu}_u,\ \boldsymbol{\Sigma}_u\big), \tag{9}$$

where $\boldsymbol{\mu}_u, \boldsymbol{\Sigma}_u$ are estimated from the profiling set.

Attack phase. Given public c and observed $\mathbf{l}$ for the target device, a key hypothesis $s_1^\star$ induces a class label $g(c[k], s_1^\star)$ for each trace k (e.g., the predicted bit of $P^\star = s_1^\star \cdot C$). We score the hypothesis by the sum of log-likelihoods:

$$\mathcal{L}(s_1^\star) \;=\; \sum_k \log p\big(\mathbf{l}[k] \;\big|\; \mathcal{C} = g(c[k], s_1^\star)\big)\,, \tag{10}$$

and select $\arg\max_{s_1^\star} \mathcal{L}(s_1^\star)$. This profiled distinguisher is complementary to 2oCPA: rather than correlating hand-crafted second-order scalars, it leverages a multivariate model over several registers. The results are highlighted in the Figure 5.

Finally, Table 1 summarizes the vulnerability assessment of the PWM for both unprotected and two-share masked implementations, including the applied distinguishers and the minimum traces to disclosure.

5 Conclusion

Offensive top-down pre-silicon analysis surfaces actionable root causes of second-order leakage in masked PWM for ML-DSA-87. By aligning toggle/activity features with post-silicon observables, designers can fix weaknesses before tape-out and preserve the trust anchored by PQC-enabled Roots of Trust.

Acknowledgments

This work in progress is funded by the Purdue Center for Secure Microelectronics Ecosystem.

References

1. E. Karabulut and K. Upadhyayula, "Side Channel Countermeasures for the Adams Bridge Accelerator," *OCP Global Summit (Security Track)*, Oct. 2024. [Online]. Available: `https://www.youtube.com/watch?v=0H37N6gSIX4`

2. M.-J. O. Saarinen, "Why 'Adams Bridge' Leaks: Attacking a PQC Root-of-Trust," *Hardwear.io USA 2025*, May–Jun. 2025. [Online]. Available: https://www.youtube.com/watch?v=_AERFU3fCns
3. M. Karabulut and R. Azarderakhsh, "Efficient CPA Attack on Hardware Implementation of ML-DSA in Post-Quantum Root of Trust," *IACR ePrint 2025/009*, Jan. 2025. [Online]. Available: https://eprint.iacr.org/2025/009
4. Z. Liu, A. Malnicof, A. Roy, and P. Schaumont, "Telescope: Top-Down Hierarchical Pre-silicon Side-channel Leakage Assessment in System-on-Chip Design," *ACM AsiaCCS*, Aug. 2025. [Online]. Available: https://dl.acm.org/doi/10.1145/3708821.3736216
5. E. Prouff, M. Rivain, and R. Bevan, *Statistical Analysis of Second Order Differential Power Analysis.*
6. C. Mujdei, L. Wouters, A. Karmakar, A. Beckers, J. M. Bermudo Mera, and I. Verbauwhede, "Side-channel Analysis of Lattice-based Post-quantum Cryptography: Exploiting Polynomial Multiplication," imec-COSIC KU Leuven, Belgium.

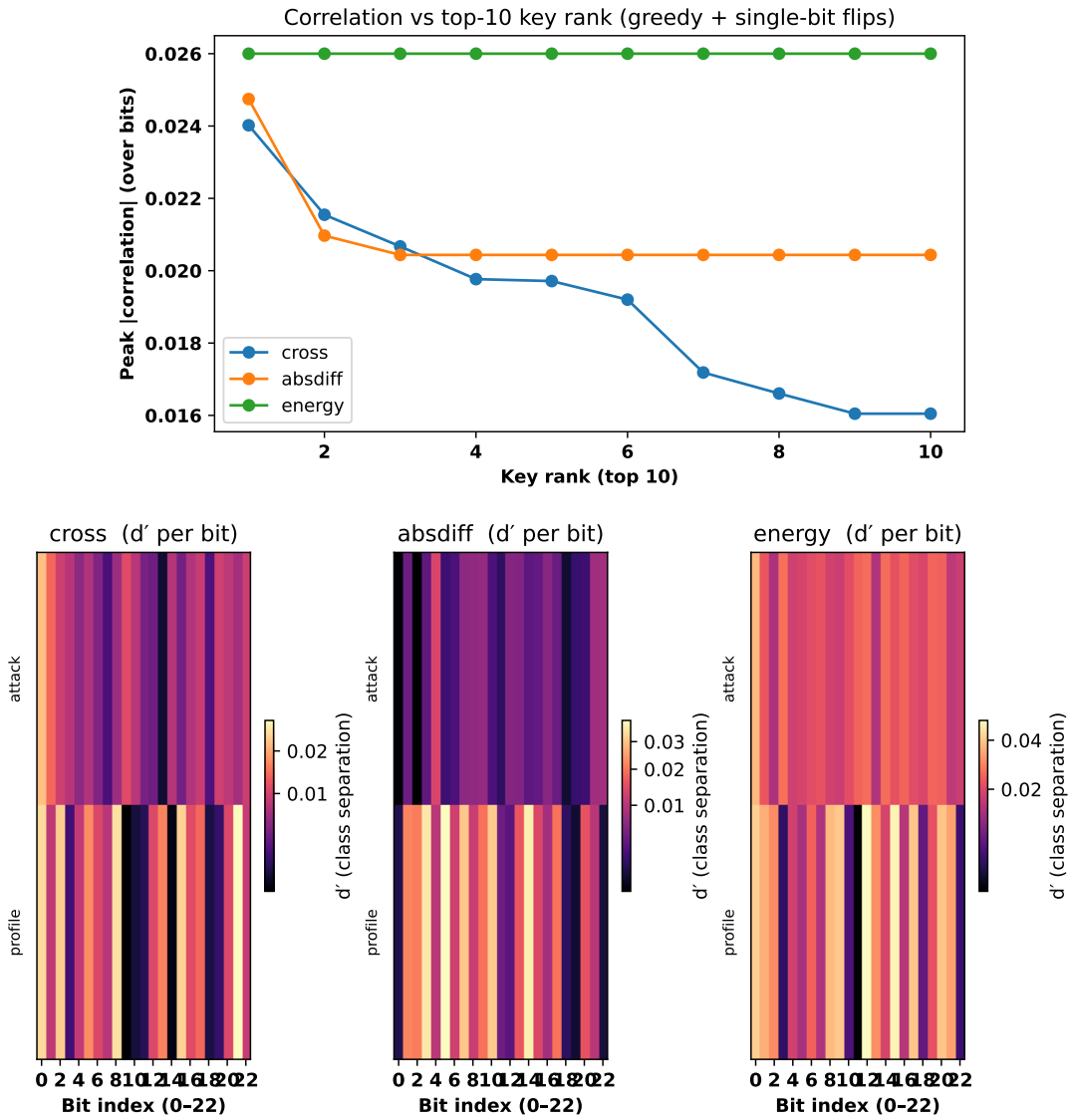

Fig. 5. Profiled (template) attack results for masked PWM. *Top:* Peak absolute correlation versus key rank among the top-10 candidates produced by the greedy key and its single-bit flips (LSB→MSB). The *energy* feature exhibits the strongest separation (nearly flat at the top rank, green), followed by *absdiff* (orange) and *cross* (blue). The monotone decline indicates that the correct key (rank 1) is consistently preferred; nearby single-bit flips degrade predictably. *Bottom:* Bitwise d' separation heatmaps for the three models, shown for *profile* (upper band) and *attack* (lower band). Brighter columns correspond to bits where the Gaussian classes are well separated. The attack bands closely track the profile ones, with mild attenuation due to distribution shift; the *energy* model concentrates separation around mid-significant bits (indices ≈10–14), while *absdiff* benefits from small gain mismatches and *cross* provides stable but lower contrast.