On the Practicality of Relying on Simulations in Different Abstraction Levels for Pre-Silicon Side-Channel Analysis

Javad Bahrami\textsuperscript{1}, Mohammad Ebrahimabadi\textsuperscript{1}, Sofiane Takarabt\textsuperscript{2}, Jean-luc Danger\textsuperscript{3}, Sylvain Guilley\textsuperscript{2,3} and Naghmeh Karimi\textsuperscript{1}

\textsuperscript{1}University of Maryland Baltimore County, Baltimore, US 21250
\textsuperscript{2}Secure-IC S.A.S., Think Ahead Business Line, Paris, France
\textsuperscript{3}LTCI, Télécom Paris, Institut polytechnique de Paris, Paris, France

Keywords: Side-Channel Attacks; Pre-Silicon Evaluation; Toggle Count; SPICE Simulation; Masked Implementations.

Abstract: Cryptographic chips are prone to side-channel analysis attacks aiming at extracting their secrets. Side-channel leakage is particularly hard to remove completely, unless using a bottom-up approach (compositional security). On the contrary, industrial secure-by-design methods are rather relying on a top-down approach: (would-be) protected circuits are synthesized by Electronic Design Automation (EDA) tools. Tracking that no leakage exists at any refinement stage is therefore a challenge. Experience has shown that multiple leakages can resurge out of the blue when a sound RTL design is turned into a technology-mapped netlist. Checking for leaks and identifying them is a challenge. When the netlist is unstructured (e.g., it results from an EDA tool), dynamic checking appears as the most straightforward approach. It is feasible, given only a few thousand execution traces, to decide with a great certainty whether a leakage hides at some time samples within the trace or not. In practice, such easy detection is fostered by the fact that the activity of signals in cryptographic implementations (even more true for masked implementations) is almost maximal (\approx 50\%). The remaining question is about the adequate abstraction level of the simulation. The higher as possible abstractions are preferred, as they potentially capture more situations. However, if the simulation is too abstract, it may model the reality inappropriately. In this paper, we explore whether or not an evenemential simulation (toggle count) is faithful with respect to a low-level simulation (at SPICE level). Our results show that both abstraction levels match qualitatively for unprotected implementations. However, abstract toggle count simulations are no longer connected to real SPICE simulations in masked implementations. The reason is that the effect of the random mask is to mix evenemential simulations (which only reflect "approximately" the SPICE reality) together, in such a way that the useful information is lost. Therefore, masked logic netlist implementations shall be analysed only at SPICE level.

1 INTRODUCTION

Electronic circuits which manipulate sensitive information must implement cryptographic algorithms. Typically, such algorithms are block ciphers, whose security relies on a secret key. Embedded devices, like IoTs, resort to lightweight cryptography, such as PRESENT. The most realistic threat against PRESENT is the side-channel analysis (such as power or electromagnetic analysis). Such attack exploits the activity which leaks information about (sensitive) internal signals. Many works attest of the practical feasibility of such attacks (Heuser et al., 2016).

In order to protect the implementation of cryptographic algorithms against Side-Channel Analysis (SCA), protections are deployed. Random masking (or simply "masking") is a protection against such attacks; it consists in mixing the sensitive information with unpredictable random numbers, also referred to as "masks", in a view to decorrelate the observable side-channel activity from the data handled internally of the device. There are several flavors of masking schemes, which differ by their security and cost. Some schemes are more costly (in terms of gate count and delay) because they address better the peculiarity of hardware implementations. Indeed, software masking schemes mainly focus on protecting variables, whereas hardware masking schemes also protect relevantly the combinational logic which lays between the sequential logic carrying the states.

Precisely, all masked logic can be implemented correctly for registers, which are well identified and controlled resources. This is positive as registers are also the resources which are the most prone to leak-
age. But so, when none of the registers are leaking, the second source of leakage is the combinational logic. This part of the circuit is less predictable, as:

- before tapeout, it is mapped, most probably restructured, and potentially even simplified, and
- upon execution, it features races which might differ on the PVTA (Process, Voltage, Temperature, and Transistor Aging) environmental conditions.

There is therefore the need for tools and methods to evaluate masked netlists. Obviously, preventing is better than curing. Therefore, in this paper, we study pre-silicon techniques. The central question is to determine which pre-silicon technique is reasonable.

2 State of the Art

Predicting whether an implementation is vulnerable to side-channel attacks when still at pre-silicon stage has been addressed by many works. Software simulators can be beneficial in this respect, since software leakage is complex. There are multiple points of interest that can leak, hence checking those mitigations in a program which embeds countermeasures is not trivial. Therefore, multiple simulators have been designed to extract the power leakage. They can be dedicated tools, or can consist in a post-processing of execution traces dumped by a COTS tool.

Program Inferred Power Analysis Simulator (PIN-PAS) is the first exclusive software for analyzing side-channel power written in JAVA which is primarily designed for testing smart-cards (Hartog et al., 2003). This tool is able to simulate the execution of a program (in assembly code) in its virtual environment to subsequently analyze its estimated power. Kirschbaum et al. (Kirschbaum, 2007) designed a simulator based on the Cadence NCSim to generate an accurate leakage tracing. The designed tool extracts power traces based on the gate level netlist, cell information, and gates propagation time to count the transitions. However, it requires a lot of information of hardware implementation of devices which is not always accessible. Debande et al. (Debande et al., 2012) designed a profiled simulator based on the stochastic models and the successive values of device’s registers. This tool is able to extract leakage traces close to the real measurement. This simulator does not need any specific information of the device, however, profiling step should be performed for every new device. SILK was the first fully open-source simulator for side-channel analysis which is unveil in (Veshchikov, 2014). The flexibility of SILK makes users capable to use different model such as Hamming Weight (HW), Hamming Distance (HD), and even user-defined leakage models.

Most of simulations leverage a toggle count leakage model (namely, the Hamming Distance between signals at each trace time sample). Indeed, it has been acknowledged for long in the field of embedded systems security, for instance, in the seminal paper about Correlation Power Analyses (Brier et al., 2004, §2). In this paper, the model focuses on flip-flops. But it has also been extended to combinational logic with fair correlation, e.g., to estimate glitching activity (Liu et al., 2011) (in unprotected circuits tough).

Compared to the aforementioned tools, SPICE simulations are more realistic. They can be considered as the most accurate possible simulations, and are seen as references (Li et al., 2005). Notice that the Synopsys VCS and HSPICE netlists we consider are the same, except that Synopsys VCS reports only an “impulse” for each gate when toggles, whereas HSPICE produces a complete waveform (illustrations will be given in Fig. 3 and Fig. 4).

To fill the gap, in this paper, we revisit the use of toggle count vs SPICE simulations, applied to masked netlists. More precisely, we address the relevance of deploying toggle count in evaluating pre-silicon security level:

- If there is a significant correlation between sensitive variables and the toggle count, then there is obviously a vulnerability, hence the netlist must be patched. Notice that, in the past, the compromise in the netlists has been identified based on toggle count analyses. For example, the flaw in (non-glitch resistant) masking of AND gates has been detected by SPICE in (Mangard et al., 2005, §4) and subsequently qualified based on a toggle count in (Mangard et al., 2006, §2.2.). Recall that “glitches” are races between signals, which are likely to generate transient activity observable through side-channels.

- However, assuming that the toggle count analysis does not reveal signification biases (in terms of correlation with the underlying sensitive variable value), can it be claimed that the netlist is secure? Actually, the absence of correlation can also arise from the fact that the “toggle count may not be a valid abstraction” in some cases. To the best of our knowledge, this question has never been formulated as such and consequently has never been thoroughly studied. In this paper, we carry out a meticulous analysis and conclude that, though SPICE is always suitable a pre-silicon simulation tool, toggle count might not be indicated as

---

1For the sake of illustration, we use HSPICE from Synopsys Inc. as an electrical simulation tool for SPICE netlists.
a definitive argument to claim the trustworthiness of masked netlist.

3 PRESENT S-Box Implementations

In this paper, we target PRESENT cipher and mainly the hardware implementations of 2 unprotected and 5 masking-protected implementations of its S-box module. Table 1 compares the targeted implementations regarding the number of gates, equivalent gates (# gates normalized by the # of equivalent 2-input NAND gates), random bits, and propagation delay.

PRESENT is a lightweight block cipher with 64-bit blocks. It includes 31 rounds each consists of a bitwise XOR operation, a non-linear substitution (S-box) layer and a linear permutation layer. Being non-linear and possessing a contrasted confusion coefficient (Fei et al., 2015), the S-box module is a highly appealing target for adversaries to leak sensitive data.

3.1 Unprotected Implementations

Lookup Table (LUT): This baseline architecture can be implemented using a 4-bit LUT with 16 different addressable locations. It is a direct look-up in the plain table: $\text{S-Box} : [0xc, 0x5, 0x6, 0xb, 0x0, 0xa, 0xd, 0x3, 0xe, 0xf, 0x8, 0x4, 0x7, 0x1, 0x2]$, and is subsequently mapped to gates by the synthesizer.

Optimized Implementation (LUT-OPT): This implementation minimizes the number of non-linear (AND/OR) gates by trying different implementations using Boolean Satisfiability (SAT) solvers (Peralta et al., 2022). It has a longer critical path compared to LUT. Indeed LUT-OPT critical path includes more XOR gates, while LUT has more AND/OR gates; thus LUT-OPT has a longer propagation delay.

3.2 Protected Implementations

Global Lookup Table (GLUT): GLUT masking is a function $\mathcal{F}^2_2 \times \mathcal{F}^2_2 \times \mathcal{F}^2_2 = \mathcal{F}^{12}_2 \rightarrow \mathcal{F}^2_2$ satisfying
\[
Y = \text{GLUT}(A, \text{MI}, \text{MO})
\]
such that:
\[
Y \oplus \text{MO} = \text{S-Box}(A \oplus \text{MI})
\]
where $A$ and $Y$ are the masked input and output, respectively and $\text{MI}$ and $\text{MO}$ are the input and output masks.

Rotating S-box Masking (RSM): In this first-order secure implementation, even 2nd and 3rd order secure based on (Carlet and Guilley, 2013), to minimize the required randomness, 1) the masks set is a subset of the full mask set; 2) the masks used at the S-box output are deduced deterministically from that at the input, by using the next one (in a circular manner) within the mask set.

In our implementation, since there are only 16 masks in the 4-bit PRESENT S-box, we refrain from using a strict subset of the masks. However, we generated output masks based on the available input randomness (Nassar et al., 2012), thus the output mask $\text{MO}$ is generated as below:
\[
\text{MO} = (\text{MI} + 1) \mod 16
\]
where $\text{MI}$ and $\text{MO}$ are integers in $\{0, 1, \ldots, 15\}$, computed via $M \in \mathcal{F}^2_2 \rightarrow \sum_{i=0}^{3} M2^i \in \{0, 1, \ldots, 15\}$ mapping. So, the relationship between RSM and GLUT can be written as:
\[
\text{RSM}(A, \text{MI}) = \text{GLUT}(A, \text{MI}, (\text{MI} + 1) \mod 16)
\]

From an area perspective, RSM has a more compact architecture compared to the GLUT. Note that the genuine version of RSM does not have S-box input on 8-bit, but solely on 4-bit (Guilley et al., 2017a) (wherein randomness comes from S-box shuffling).

ROM-based RSM (RSM-ROM): A stronger implementation of RSM can be realized using a Read-Only Memory (ROM). For this circuit we target logic designs built only from the instantiation of gates in a Boolean library (Giaconia et al., 2007). Here, initially, the datapath is synchronized for any input configuration which makes input-related deviations of leakage small. Then, the structure is designed with a one-hot strategy, i.e., only the required logic is activated, which further contributes to reducing the side-channel footprints.

Gate-Level Masking via Random Sharing (ISW): Proposed by Ishai, Sahai, and Wagner (Ishai et al., 2003), this implementation starts from the LUT-OPT netlist, and gradually replaces the gates with their gadgets, in order to deal with the non-linear gates. In this architecture, the gadget for the AND gate requires 1-bit of randomness, i.e., $R$. Given a random sharing $(A_0, A_1)$ of bit $A$ (where $A = A_0 \oplus A_1$), and a similar sharing for bit $B$, the AND of $A$ and $B$ denoted as $Y$ is computed as below:
\[
\begin{align*}
Y_0 &= ((A_0 \land B_1) \lor R) \lor (A_0 \land B_0) \\
Y_1 &= ((A_0 \land B_1) \lor R) \lor (A_1 \land B_0)
\end{align*}
\]
In the above equations, the order of operations should be followed, thus the implementation must preserve the order of gates in the final netlist. In our implementation, we implemented OR via benefiting from De Morgan’s law $\text{OR}(a, b) = \neg\text{AND}(\neg a, \neg b)$. Since combinatorial gates evaluate their outputs whenever their inputs change, preserving the order can be challenging in ISW due to the race condition. This can lead to a first-order leakage (Roy et al., 2018).
Threshold Implementation (TI): TI is an algorithmic countermeasure against power SCA, which benefits from multi-party computation and secret sharing. TI, unlike ISW, divides input bits into \( d + 1 \) shares, however, its underlying logic does not need to preserve gate ordering. Moreover, thanks to its non-completeness property, in TI any output share only depends on \( d \) shares of each input. Thus, glitches cannot lead to secret information disclosure in TI. In our netlist, terms of order 3 are required, hence 4 shares are needed; we synthesized a TI-compliant fully combinational netlist of PRESENT S-box.

Table 1: Gate-level specification of the targeted S-box implementations.

<table>
<thead>
<tr>
<th># INV</th>
<th>LUT-OPT</th>
<th>GLUT</th>
<th>RSM</th>
<th>RSM-ROM</th>
<th>ISW</th>
<th>TI</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>1</td>
<td>12</td>
<td>20</td>
<td>750</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td># BUF</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td># AND</td>
<td>23</td>
<td>2</td>
<td>580</td>
<td>209</td>
<td>40</td>
<td>1000</td>
</tr>
<tr>
<td># NAND</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td># OR</td>
<td>11</td>
<td>2</td>
<td>254</td>
<td>102</td>
<td>468</td>
<td>0</td>
</tr>
<tr>
<td># NOR</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td># XOR</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td># Gates</td>
<td>41</td>
<td>14</td>
<td>846</td>
<td>331</td>
<td>1750</td>
<td>43</td>
</tr>
<tr>
<td># Eqns. Gates</td>
<td>36</td>
<td>18</td>
<td>846</td>
<td>317</td>
<td>1322</td>
<td>52</td>
</tr>
<tr>
<td>Delay (ps)</td>
<td>160</td>
<td>210</td>
<td>400</td>
<td>330</td>
<td>1710</td>
<td>420</td>
</tr>
<tr>
<td># Random Bits</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

4 Background on Walsh-Hadamard Transform

In practice, glitches may result in leaking sensitive data. However, getting information about when glitches happen, and what information they leak may not be easy during the circuit runtime or even in the SPICE level simulations. To investigate the exploitability of glitches Bahrami et al. use Fourier base transformation (the so-called Walsh-Hadamard) to be able to decompose the device’s power traces into the unmasked input dependent basis, and in turn identify different contribution of leakage from first-order unmasked text \( t \) (Guilley et al., 2017b). The basis is denoted as \( \psi_u \), for all vector \( u \) of \( n \)-bits (\( n = 4 \) for PRESENT). By definition, \( \psi_u \) function is \( t \in F_2^n \mapsto \psi_u(t) = \frac{1}{\sqrt{2^n}}(-1)^{u \cdot t} \), where the operation \( u \cdot t \) denotes the canonical scalar product, i.e., \( u \cdot t = \sum_{i=1}^{n} u_i \wedge t_i \). This basis is orthonormal, in that \( \langle \psi_u | \psi_v \rangle = 0 \) if \( u \neq v \) and 1 otherwise. In sum, \( \psi_u \) shows the interaction between input bits corresponding to the 1 input in \( u \). Here, we ignore the zero-th component, which is the waveform average shape, while extracting all the nonzero components.

Based on (Bahrami et al., 2022), the leakage \( f : t \in F_2^n \mapsto f(t) \in R \) corresponding to (unmasked) S-box input \( t \in F_2^n \), can be calculated as

\[
f(t) = \sum_u (f | \psi_u) \psi_u(t) = \frac{1}{2^{n/2}} \sum a_u(-1)^{u \cdot t}
\]

where \( a_u \) (below) is the Walsh-Hadamard transform of \( f \).

\[
a_u = \frac{1}{2^{n/2}} \sum f(t)(-1)^{u \cdot t}
\]

Using \( (a_u)_{u \in F_2^n} \) coefficients we can characterize the leakage to find out which combination of bits leaks more (if any). Typically, it can be due to a single bit, when \( w_H(u) = 1 \) (where \( w_H \) denotes to Hamming Weight) or otherwise to multiple bits leaking together. The former case is reasonably attributed to an unfortunate “demasking”; while the latter case arises upon complex conditions on the unmasked inputs (connected to “glitch events”) (Bahrami et al., 2022). For example in Fig. 3d, the combination of bit 1 and bit 2 leaks more (shown in solid red).

Interpreting the origin of leakage is out of the scope of this paper. We rather focus on comparing the amount of leakage of various implementations.

5 Results and Discussions

We implemented the add-round-key and S-box operations in the first round of the PRESENT cipher with 80-bit keys in the transistor level for the 7 types of S-box (2 unmasked and 5 masked versions) using a 14-nm Fin-FET – regular voltage threshold corner (rvt) – commercial library, and deployed Synopsys HSPICE for the transistor-level simulations at temperature of 125°C and \( V_{dd} = 0.8 \) V.

The simulated traces contain two parts: the results of key addition and S-box outputs for each initial \( n \)-bit value as well as its following \( n \)-bit value. For the analysis, we considered only the power samples of the second part, i.e., when the cryptographic circuit transitions from initial to final value. We sample the power every 1ps after feeding the related final value till each circuit gets stable. To have a fair comparison among all implementations, the final values are generated randomly, yet such that we have equal number of traces from each class, i.e., equal number of unmasked inputs. The initial values generated randomly such that initial unmasked value is ‘0’ (e.g., \( A_{initial} \oplus MI_{initial} = 0 \) in GLUT). For ISW, TI, and GLUT, totally, we generated 1024 input traces among all possible input combinations, while for RSM and RSM-ROM the total input traces are 256 and for the LUT and LUT-OPT there are 16 traces considering their input sizes. We categorize the power traces into 16 classes based on the value of the unmasked inputs in their final stage, and use the mean trace of each class in our leakage assessments.
To assess the leakage from toggle counts occurring during gate-level simulations, we performed our simulations (using Synopsys VCS) and analysis after back annotating the timing information (SDF file) extracted during the gate-level synthesis.

5.1 Experimental Results

Toggle Count vs. SPICE Power: The first set of results deals with the total power for each class of unmasked plaintext. Recall that we randomly generated input patterns such that in all traces the initial unmasked value is “0000” (referring to class 0) while the final unmasked value can belong to any class (between 0 and 15 in PRESENT). Moreover, to avoid biases the final input patterns were generated such that we have similar number of power traces in each class. Fig. 1 shows the total extracted power values categorized based on the final value of the unmasked input.

Obviously, the unmasked circuitries do not reveal dynamic power consumption for class 0 input as the values represent the dynamic power consumption; thus class 0 for unmasked circuits relates to the case where both initial and final value are 0, so, no dynamic power is consumed. This is in contrast of masking circuits where even we have power consumption for class 0 since owing to the random masks not necessarily initial and final values match.

Based on Fig. 1, the unprotected circuits depict an intra-class match between the toggle count and HSPICE powers, i.e., tentatively class $i$ ($0 \leq i \leq 15$) in both evaluations follow the same trend. However, such trend is not observed in the masked circuits owing to the dependency of power on both input and output masks. The takeaway from these observations is that for unprotected circuits, the toggle counts carry exploitable information and can be used as a source of leakage. However, toggle counts are not exploitable by the attacker in case of protected implementations.

The inter-class correlation between the HSPICE and toggle count results are shown in Table 2. As depicted, this correlation is high for unmasked circuits while low for the masked counterparts. This information shows the practicality of using toggle counts for leaking sensitive data in unmasked circuits, yet not in the masked implementations.

Table 2: Inter-class Pearson Correlation Between Toggle Count and HSPICE powers for each implemented circuit.

<table>
<thead>
<tr>
<th>Circuit</th>
<th>LUT</th>
<th>LUT-OPT</th>
<th>GLUT</th>
<th>RSM</th>
<th>RSM-ROM</th>
<th>ISW</th>
<th>TI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$</td>
<td>0.78</td>
<td>0.78</td>
<td>-0.21</td>
<td>-0.23</td>
<td>-0.21</td>
<td>0.21</td>
<td>0.21</td>
</tr>
</tbody>
</table>

In order to validate that our simulations of masking schemes implementation is correct, we carry out a sanity check on our simulated results. Namely, the next set of results depicted in Fig. 2 illustrates the variance of power values between 16 different classes.
of all 7 implementations in a logarithmic scale for each of HSPICE and toggle count results. Indeed, the results shown in this figure relate to the variances of the bars shown in Fig. 1. The 7 points are almost aligned, which confirms that toggle count is valid estimation of the total HSPICE. This confirms that toggle count is meaningful to summarize global activity but it does not accurately capture data dependencies (as pointed in Table 2) in masked implementations.

**Walsh-Hadamard Leakage Analysis using Toggle Count vs. SPICE Power:** As described earlier, the Walsh-Hadamard transform is used to extract all vulnerabilities as a result of bit interactions (cross-talk) inside implementations. Fig. 3 and Fig. 4 show the post Walsh-Hadamard leakage analysis when using HSPICE and toggle count results, respectively. Interestingly, it is known that TI implementation features glitches; nonetheless Fig. 4g shows that they do not leak information (i.e., all the spectrum have the same magnitude regardless of their polarity which itself does not convey any relevant information). This is expected since those glitches mix only at most \( d = 3 \) shares of each variable randomly split in \( d + 1 = 4 \) shares. As shown, the leakage sources are more distinguishable in HSPICE results depicted in Fig. 3 while the distinguishably is much less when using toggle counts. This confirms that it is not always practical to assess the implementations based on their toggle count data, and more precise simulations such as HSPICE are required.

Another interesting observation here is the leakage of bit 3 (shown in dotted cyan) in most of the implementations when using toggle counts as well as for unprotected implementations when using HSPICE. This is due to the fact that bit 3 in PRESENT S-box has a higher algebraic complexity than other bits.

### 6 Conclusion

Side-channel leakage detection can benefit from fast “toggle count” simulations when netlists are not implementing a masking countermeasure. Otherwise, we show on 5 different masked implementations that toggle count is consistently not related to the actual netlist leakage, as estimated by SPICE simulations. Therefore, leveraging “toggle count” to assess a netlist should be used with caution. Should any bias be found, then obviously the design features a side-channel vulnerability. However, when leakage analysis with “toggle count” features no bias, a refined analysis at SPICE level is required.

### REFERENCES


Figure 3: Walsh-Hadamard coefficients extracted using HSPICE.

Figure 4: Walsh-Hadamard coefficients extracted using Toggle Counts.