This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

The Low Area Probing Detector as a Countermeasure Against Invasive Attacks

Michael Weiner, Salvador Manich, Rosa Rodríguez-Montañés, and Georg Sigl

***Abstract*— Microprobing allows intercepting data from on-chip wires as well as injecting faults into data or control lines. This makes it a commonly used attack technique against security-related semiconductors, such as smart card controllers. We present the low area probing detector (LAPD) as an efficient approach to detect microprobing. It compares delay differences between symmetric lines such as bus lines to detect timing asymmetries introduced by the capacitive load of a probe. Compared with state-of-the-art microprobing countermeasures from industry, such as shields or bus encryption, the area overhead is minimal and no delays are introduced; in contrast to probing detection schemes from academia, such as the probe attempt detector, no analog circuitry is needed. We show the Monte Carlo simulation results of mismatch variations as well as process, voltage, and temperature corners on a 65-nm technology and present a simple reliability optimization. Eventually, we show that the detection of state-of-the-art commercial microprobes is possible even under extreme conditions and the margin with respect to false positives is sufficient.**

***Index Terms*— Data buses, digital integrated circuits, invasive attacks, microprobing, security, smart cards.**

I. Introduction

SEMICONDUCTORS have been used in security applications for more than 30 years. Their first applications were in public telephones where they served as payment cards, as well as in pay TV where they were required to decrypt video signals. As such, security relevant semiconductors were most frequently embedded into plastic cards, the term "Smart Card" was coined for such cards with an embedded semiconductor.

Three decades ago, when the first Smart Cards appeared, so did attacks against them. In the simplest case, their purpose could have been preventing to debit balance from phone cards, while more sophisticated attacks already aimed at full dumps to reveal algorithms and keys of cryptographic primitives. The methods used were quite simple in the beginning. Debiting balance could be prevented by disconnecting the programming voltage; read only memory dumps were possible, for example, using glitching [1].

In the meantime, a circle of novel attacks and countermeasures has significantly improved the attack resistance of today's security microcontrollers. Glitch detectors as well as temperature and light sensors were added to detect fault attacks. When side channel attacks came up, massive efforts were spent on modeling and reducing the leakage at different abstraction layers. Today, the most sophisticated attacks of this kind appear to be localized electromagnetic attacks [2]; recent publications [3], [4] have presented detectors of these attacks.

In 2010, Tarnovsky was able to carry out a full memory dump of a smart card controller by microprobing the bus [5]. This was successful in spite of its protective mesh. In the following years, industry and academia have been working on countermeasures against microprobing. Masking schemes were implemented to make probing single lines worthless [6]; circuits dedicated to the detection of microprobes based on their parasitic capacitance have been proposed [7], [8] and also other circuits proposed in academia that can detect microprobes as a side effect [9].

We have presented the concept of a low area probing detector (LAPD) [8] that only consists of a few gates and therefore keeps the area and power overhead low. In this paper, we demonstrate its reliability with respect to process variations and varying environmental conditions and give recommendations how to increase the reliability beyond its intrinsic limits.

Section II will describe microprobing and its countermeasures in more detail. Section III will give a brief description of the LAPD and show how it can be integrated into a bus system. The results and discussion of reliability are presented in Section IV; furthermore, this section explains how its reliability can be improved. Section V presents future work and this paper is concluded in Section VI.

II. Problem Statement

The bus of a smart card controller is a highly desirable attack target for microprobing. It concentrates the information transferred between CPU core, memory, and peripherals in a small area; this includes sensitive data such as the controller firmware or cryptographic keys. While it is physically difficult to probe all lines of a bus at the same time, adversaries have been iteratively probing one line of a bus after another; the acquired data are then accumulated in a later step.

Manuscript received April 4, 2017; revised August 18, 2017; accepted September 28, 2017. This work was supported by the Spanish TEC2013-J41209-P Government Project. *(Corresponding author: Michael Weiner.)*

M. Weiner is with the Chair of Security in Information Technology, Technical University of Munich, 80333 Munich, Germany and also with the Research and Development Department, SimonsVoss Technologies GmbH, 85774 Unterföhring, Germany (e-mail: m.weiner@tum.de).

S. Manich and R. Rodríguez-Montañés are with the Department of Electronic Engineering, Escola Tècnica Superior d'Enginyeria Industrial de Barcelona, Universitat Politècnica de Catalunya, 08028 Barcelona, Spain (e-mail: salvador.manich@upc.edu; rosa.rodriguez@upc.edu).

G. Sigl is with the Chair of Security in Information Technology, Department of Electrical and Computer Engineering, Technical University of Munich, 80333 Munich, Germany and also with the Fraunhofer Institute for Applied and Integrated Security, 85748 Garching, Germany (e-mail: sigl@tum.de).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVLSI.2017.2762630

1063-8210 © 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

A. State of the Art

State-of-the-art microprobing protection in smartcard controllers can be classified into three categories. One can either devalue the outcome of probing, e.g., by using bus encryption or masking, one can obstruct physical access to target lines, or one can detect inherent effects of probes.

Ishai *et al.* [10] apply multiparty computation techniques to mask signals; however, the circuit complexity increases by $O(n^2)$ in the general case for protecting against probing n lines simultaneously. The authors themselves put the practicability of their approach in question. Furthermore, protection against fault injection would require additional complexity.

On the industry side, redundant cores combined with bus encryption can provide protection against targeted fault injection and void the value of probed signals. While this approach provides a generic protection against faults and information leakage and is implemented by a major semiconductor manufacturer in their flagship smartcard controllers, the large hardware overhead might not be suitable for low-cost high-volume products such as subscriber identity module cards. In addition, the introduced delay and power consumption may complicate their use in low-latency and ultralow power applications.

Obstructing access to target lines can be done in a passive way, e.g., by metal fillings or passive shields, or by active shields. Passive shields can be removed by focused ion beam (FIB) machines. Active shields usually drive test patterns through a mesh on the top layer and verify that the patterns reach the other ends of the mesh lines. Cioranesco *et al.* suggested to use cryptographic PRNGs to provide a large number of unpredictable test signals [11]. However, this comes with an increased hardware cost, and it can likely be circumvented by adding bypass lines on top of the passivation layer using an FIB.

Other approaches try to bury security critical signals underneath other functional, but noncritical lines. Shi *et al.* present an algorithm to determine the exposure of critical lines [12]. Still, this does not appear as the overall solution: the zero exposure of target lines is hard to reach, especially if designers want to avoid multiple layout iterations, which is critical for fast time-to-market. Also, bypassing cut lines above the top layer is still feasible for an attacker.

All of the described countermeasures do not protect against probing attacks from the backside. This vulnerability can be avoided if the inherent effects of invasive attacks such as probing are detected, as it can be done by observing the capacitive load of a probe. That way, probing can be detected no matter whether extensive FIB editing was used to uncover target lines, or whether a probe was connected on the back side. The only approach that detects such attacks and that has been evaluated with respect to process, voltage, and temperature variation is the probe attempt detector (PAD) by Manich *et al.* [7] whose principle of operation is briefly described in Section II-B. However, the necessity of a large tank capacitor that needs to be charged and discharged still comes with an area, power, and timing overhead that prevents it from being used in ultralow resource applications.

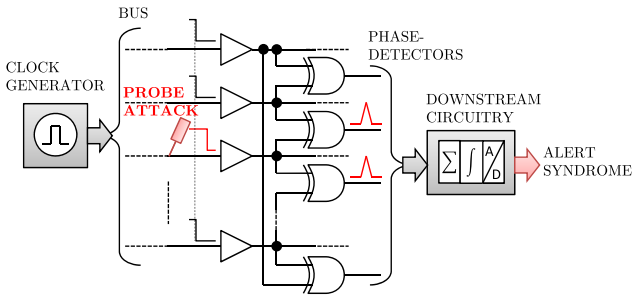

Fig. 1. PAD overview.

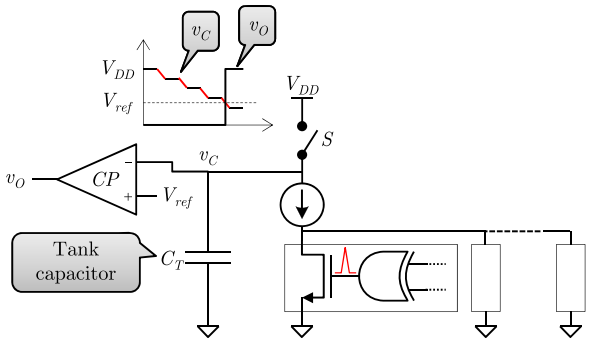

Fig. 2. PAD detector circuit.

Also, it does not allow its implementation in programmable logic platforms like field-programmable gate arrays.

Please note that in addition to invasive attacks, there exist semi-invasive attacks that do not require electrical contacts to the chip, as classified by Skorobogatov [13]. Localized electromagnetic attacks [2] are an example for semi-invasive attacks. Such attacks can be detected by other types of detectors, as for example presented by Homma *et al.* [3], [4]. However, these attacks are specialized to a certain target, e.g., to the extraction of cryptographic keys by the means of correlation attacks [14], and they are usually the wrong tool for an attacker to create a complete memory dump. Therefore, we consider protection against semi-invasive attacks as orthogonal to protection against invasive attacks, and we do not consider them here any further.

B. Probe Attempt Detector

The PAD was proposed in [7] and is the first technique detecting physical attacks in buses by measuring the modification of parasitics provoked in the lines.

In Fig. 1, an overview of the detector is shown. The PAD runs in off-line mode and when started, a periodic signal is sent simultaneously through all the lines. At the outputs, XOR gates compare the state of the lines and if transitions arrive with different propagation delays, and they generate pulses of a width proportional to the delay difference. A downstream circuitry adds all these pulses, integrates over time, and generates a digital alert symptom. Because of the differential mode, the response of the PAD does not depend on the number of buffers inserted in the bus lines.

In Fig. 2, a simplified model of the downstream circuitry is shown. A tank capacitor C_T with the initial charge $C_T V_{\mathrm{DD}}$ is gradually discharged by the pulses coming from the XOR gates.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

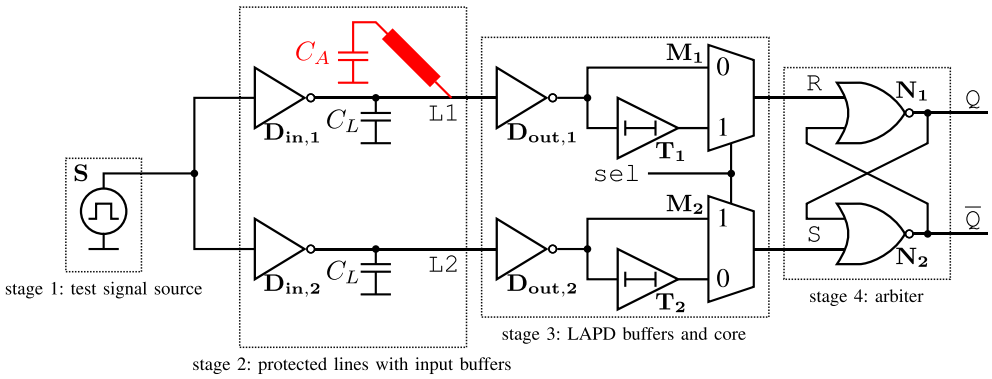

Fig. 3. Schematic of the LAPD.

When the pulses arrive, they switch ON nMOS transistors, which in turn extract some charge from C_T through a current source; therefore, the amount of charge discharged from the capacitor is proportional to the "active" time of the nMOS transistors. Initially, when the detector starts, C_T is charged to the maximum voltage V_{DD} through a switch S. Then, the switch is opened and the XOR gates start comparing signals coming from the bus during a given integration time. If the arrival times of the XOR inputs are mutually delayed by a probe, the XOR gates generate pulses accordingly, which in turn gradually discharge the capacitor. A comparator CP raises its output when the voltage v_C goes below the threshold V_{ref}. A probing attack alert is activated when this signal is raised earlier than normal.

C. Limitations of Previous Work

What we consider missing is a low-cost detection circuitry for state-of-the-art invasive attacks that can, on the one hand, support the security of high-end smartcard controllers, but on the other hand is also able to increase the attack barrier for mass-produced low-cost devices. The concept of an LAPD by Weiner *et al.* [8] fills this gap; here, we provide a detailed analysis of the LAPD including a reliability analysis with respect to manufacturing variations and varying environmental conditions.

III. Low Area Probing Detector

A microprobe attached to a line on a semiconductor acts as a small parasitic capacitance; this increases the rise and fall times of the transmitted signals. Considering a set of lines that are symmetric with respect to dimensions and timing, probing *one* of the lines introduces a small timing asymmetry between the probed line and the unprobed lines. The LAPD [8] can measure such timing differences and raise an alarm if they are beyond normal noise or manufacturing variations. This increases the complexity of a microprobing attack. If n lines are protected by the LAPD, $n-1$ microprobes can be detected such that the adversary would need to attach the same capacitive load to all n protected lines. We assume this to be an effective countermeasure against practical probing attacks, as the space for micropositioners on a probe station is limited and the measurement setup becomes more and more fragile with each additional probe. Tarnovsky [5], for example, preferred using only two probes for a successful attack, even though this implied a significant postprocessing overhead.

The LAPD performs pairwise comparisons, so Sections III-A and III-B will focus on the case of two lines. Section III-C will then show how a set of n lines can be protected.

A. Principle of Operation

The LAPD compares the delays of two lines by alternatingly introducing an intentional delay t_D to each one of the lines and then verifying that the delayed line is effectively slower than the line without intentional delay.

In Fig. 3, the full circuit is shown. In Fig. 3, bold letters represent gate instances, typewriter letters represent line names, and italic letters represent capacitances. The different stages of the LAPD are indicated by dotted squares.

The signal source **S** in stage 1 generates test pulses that are fed to the lines under tests `L1` and `L2`. In stage 3, a combination of multiplexers **M** and delay elements **T** of delay t_D allows alternately delaying one of the lines at a time through signal `sel`. Finally, the arbiter in stage 4, which consists of gates **N**, decides who "wins" the race. Under normal conditions, both lines "win" alternately; however, one line is always winning if an imbalance of more than t_D is introduced by a probe. For signal values `sel` $= 0/1$, the output `Q` produces Q_1/Q_2, respectively, as described in Section III-B.

In our case, an NOR RS latch is used as an arbiter. In one test cycle, both latch inputs are first set to the active state (1); after that, both inputs change to the inactive state (0). After the transition, the output is determined by the input signal that had been active for longer. If the `R` input remains active longer than the `S` input, `Q` becomes 0 and vice versa.

The complete LAPD timing is shown in Fig. 4. It shows two test cycles. In the first cycle, `L2` is delayed by element $\mathbf{T_2}$, while `L1` is directly passed through; in the second cycle, the delay introduced by element $\mathbf{T_1}$ is applied to `L1`, while `L2` is directly passed through. Fig. 4(a) shows the default case in which both lines have an equal capacitance. In this case, the latch output `Q` alternates between $Q_1 = 1$ and $Q_2 = 0$ at the two sampling times shown (red vertical lines). The case in which an attached probe introduces an additional delay to `L1`

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

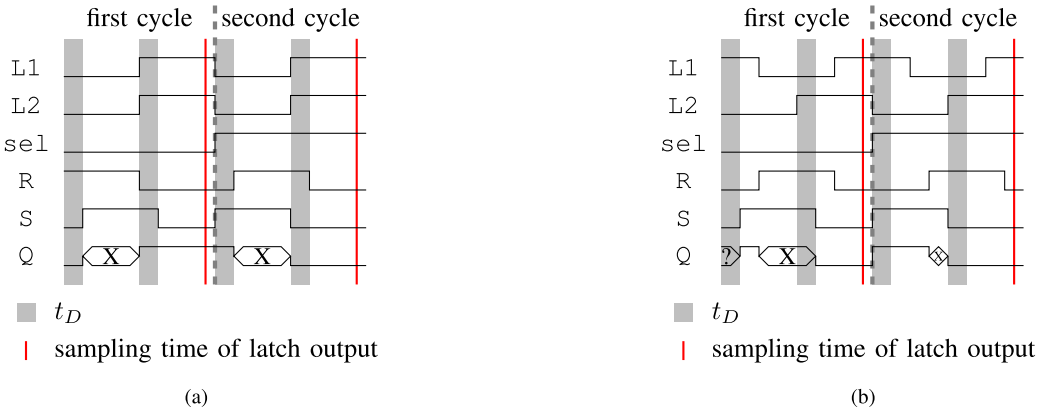

Fig. 4. LAPD timing. (a) Without probing. (b) With probing of L1.

is shown in Fig. 4(b). Here, the R input dominates the latch at both sampling times and the output Q is $Q_1 = 0$ and $Q_2 = 0$ in both the cases. Note that when both latch inputs R and S are simultaneously high, both latch outputs Q and $\overline{\text{Q}}$ are low, and thus the state of the latch is invalid. This is denoted by X in the timing diagrams.

B. LAPD Model

In order to compare the delay between two lines, we assume that both have an intrinsic parasitic capacitance of C_L. In addition, a microprobe with a parasitic capacitance C_A is attached to one of the lines. As a result, the effective capacitances of the lines during the attack are

$$\text{probed: } C_1 = C_L + C_A \tag{1}$$

$$\text{unprobed: } C_2 = C_L. \tag{2}$$

The line driver delay can be estimated using the alpha-power model for the transistors [15], [16]

$$d_i = \tilde{k}\frac{C_i\ V_{\text{DD}}}{(V_{\text{DD}} - V_t)^\alpha} \tag{3}$$

where C_i is the capacitive load at the buffer output and V_{DD} denotes the supply voltage. V_t is the threshold voltage of the transistors, α represents the velocity saturation coefficient of the carriers, and $\tilde{k}$ is called the trans-resistance that summarizes the remaining transistor parameters [17]. We assume that the technological parameters between nMOS and pMOS transistors are balanced with respect to the output transition times. Furthermore, we assume that the signals in the lines exhibit the full swing between *GND* and V_{DD} for (3); otherwise, the approximation would significantly deviate from the real behavior. This last assumption is quite reasonable, since an attack will always try to disturb the observed signals as little as possible.

Then, the delay difference between the probed and the unprobed lines is

$$\Delta t_{L1,L2} = d_2 - d_1 = \tilde{k}\frac{(C_2 - C_1)V_{\text{DD}}}{(V_{\text{DD}} - V_t)^\alpha} = -\Omega\ C_A \tag{4}$$

with the technological parameter

$$\Omega = \tilde{k}\frac{V_{\text{DD}}}{(V_{\text{DD}} - V_t)^\alpha}. \tag{5}$$

In a first approximation, the delay difference is proportional to the attack capacitance C_A, as shown in (4). The alpha model approach works better for small values of C_A. State-of-the-art semiconductor microprobes, as, for example, offered by GGB Industries, Inc. [18], [19], have parasitic capacitances in the range of tens of femtofarads and therefore can be assumed to be small enough for the approximation. Microprobes with a larger C_A may disturb regular operation of the circuit and thus not be suitable for successful microprobing attacks; furthermore, the delay function is also monotonic outside the boundaries of the small-value approximation of (4), and therefore a reliable LAPD operation can be expected.

After the bus, the $\mathbf{D_{out}}$ inverters increase the slew rate to minimize the effects of different switching thresholds of the multiplexers **M**. $\mathbf{D_{out}}$ also scales the delay difference

$$\Delta t_{\overline{L1},\overline{L2}} = k_{\mathbf{D_{out}}} \cdot \Delta t_{L1,L2} \tag{6}$$

where $\Delta t_{\overline{L1},\overline{L2}}$ is the delay difference observed after the $\mathbf{D_{out}}$ inverters.

After the $\mathbf{D_{out}}$ inverters, the transitions pass through **T** and **M** before they reach the RS latch; therefore, the delay difference at the latch inputs can be expressed as follows:

$$\Delta t_{RS} = \Delta t_{\overline{L1},\overline{L2}} \pm t_D + (t_{\mathbf{M2}} - t_{\mathbf{M1}}) \tag{7}$$

t_D is the delay introduced by the delay element **T**. In the two cycles shown in Fig. 4, it is alternated between the R and S inputs of the latch. As (1) and (2) assume that the attack capacitance C_A is attached to L1, i.e., the R path of the latch, it is sufficient to concentrate on the case where t_D only affects the S path, i.e., the first cycle of Fig. 4.

The difference $(t_{\mathbf{M2}} - t_{\mathbf{M1}})$ models the imbalances of the multiplexers **M** due to different slew rates at the input.

Inserting (4) and (6) into (7) and focusing on the first cycle, it follows:

$$\Delta t_{\text{RS}} = -k_{\mathbf{D_{out}}}\Omega \cdot C_A + t_D + t_{\mathbf{M2}} - t_{\mathbf{M1}}. \tag{8}$$

The latch needs to have a minimum distance between the falling edges to produce a reliable output; this distance can be compared with the hold time of a flipflop. Therefore

$$|\Delta t_{\text{RS}}| > t_H \tag{9}$$

holds, where t_H is the "hold time" of the latch.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

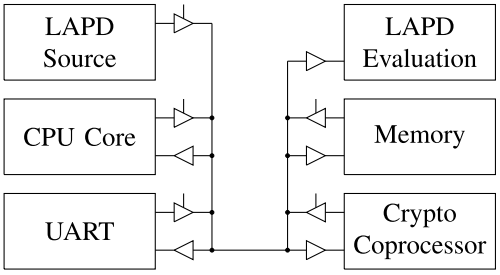

Fig. 5. LAPD integration into tristate bus.

Inserting (8) into (9), one can get the two cases

$$C_A < \frac{t_D + t_{\mathbf{M2}} - t_{\mathbf{M1}} - t_H}{\Omega \; k_{\mathbf{D_{out}}}} \tag{10}$$

$$C_A > \frac{t_D + t_{\mathbf{M2}} - t_{\mathbf{M1}} + t_H}{\Omega \; k_{\mathbf{D_{out}}}} \tag{11}$$

where (10) refers to the case that reliably does not raise an alarm, and (11) denotes the case that does raise an alarm reliably.

C. System Integration

The LAPD by itself can only protect two lines. In order to extend its protection to buses with n lines, one can use switching elements to connect two lines to the LAPD at a time and then cycle through different pairs. Alternatively, the low area of the LAPD allows us to insert n LAPD instances, which avoids additional noise introduced by switching elements and allows performing parallel evaluation of multiple lines.

To protect a bus, the LAPD can be surrounded by a state machine that requests low level bus access applying direct memory accesslike concepts; in this case, it can be connected like a peripheral component. The LAPD can be split up into a "source" part consisting of components **S** and $\mathbf{D_{in}}$ that acts as a bus master and an evaluation part consisting of $\mathbf{D_{out}}$, **M**, **T**, and **N**.

As an example, the LAPD can be integrated into a tristate bus with multiplexed address and data lines, as shown in Fig. 5. Note that for its basic operation, the LAPD source only needs an output driver to drive the test signals, and the evaluation part only needs an input driver to evaluate them.

The result evaluation can be performed in firmware. this provides most flexibility with respect to cancelling out noise by repeating the detection cycles; also, it is most flexible with respect to what to do in the case of an alarm. This can be a chip erase in the case of highly sensitive data stored on the chip, but it can also just be a reset trigger. Just as any security critical embedded software, this firmware should be protected against fault attacks either by hardware or software redundancy [20].

D. Error Compensation

Manufacturing variations as well as varying environmental conditions lead to intersample variation of the threshold capacitance value that decides between "alarm" and "no alarm." In this context, the following two types of errors should be considered.

1) Errors upon which an alarm is raised when the circuit is in fact not being attacked. These errors are called false positives or type I errors.
2) Errors upon which no alarm is raised when the circuit is in fact being attacked. These errors are called false negatives or type II errors.

These types of errors will be analyzed Section IV.

In this section, we have described the concept of an LAPD. Its simple construction allows it to be implemented in a very lightweight manner. The advantages of the LAPD over other protection concepts are qualitatively summarized in Table I.

TABLE I
QUALITATIVE ADVANTAGES OF THE LAPD AGAINST OTHER STATE-OF-THE-ART PROBING PROTECTION

alternative contermeasure	advantages of the LAPD
meshes	no additional layer protection against backside attacks
bus encryption	no latency
PAD [7]	no large capacitor
general	only few gates hardware overhead

IV. SIMULATIONS AND RESULTS

We implemented the LAPD in a 65-nm STMicroelectronics technology with a core voltage of 1.2 V. For the gates, we used the low-power standard threshold voltage transistors `psvtlp` and `nsvtlp`. Simulations were performed using Cadence spectre 11.1 and SALVADOR [21] on a machine with four AMD Opteron 6274 CPUs and 256-GB RAM.

A. Nominal Simulation

In a first run of nominal simulations, we assumed an ambient temperature of 27 °C. All transistors in our design had an aspect ratio $(W/L) = 10$. The intrinsic line capacitance was assumed as $C_L = 100$ fF. This corresponds to a line length of approximately 1.3 mm on the top metal layer in the technology we used, assuming an adjacent `GND` line with minimum distance.

In the case that the delay elements **T** are implemented as chains of two inverters, the minimum detected attack capacitance is $C_A^* = 10.3$ fF. For the case of four inverters, the minimum value becomes $C_A^* = 23.4$ fF.

B. Effects of Local Variations

As the LAPD works in differential mode, we next performed a Monte Carlo analysis of the mismatch variations using $N = 2000$ samples. Fig. 6 shows the result for the two implementations, again with delay elements **T** consisting of two and four inverters, respectively. The x-axis represents the attack capacitance and the y axis denotes the relative frequency of alarms. It is defined as $f_A(C_A) = (A/N)$. A is the number of Monte Carlo instances rising an alarm, as exemplarily

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

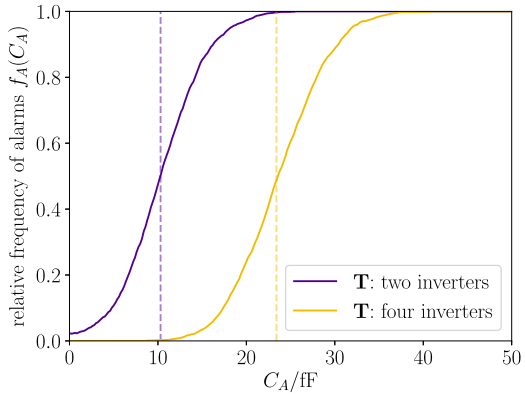

Fig. 6. Relative alarm frequency for different implementations of delay elements **T** (nominal transition values are dashed).

shown in Fig. 4(b), and N is the total number of Monte Carlo simulations.

The following qualitative observations can be made from Fig. 4.

1) The two-inverter delay implementation exhibits a nonzero alarm frequency for $C_A = 0$.
2) Both implementations exhibit an uncertainty region $\Delta C_A^U = C_{\text{reliableAlarm}} - C_{\text{reliableNoAlarm}}$ in which the behavior of the circuit is not well predictable. This region has an approximate size of $\Delta C_A^U = 25$ fF.

We note that state-of-the-art microprobes by the commercial supplier GGB Industries can all be detected by the four-inverter implementation. While the *Picoprobe Model 18C/19C* [18] is declared to have a minimum input capacitance of 20 fF, the datasheet constrains this property to signals with a transition time lower than 3 ns, while its input capacitance is 60 fF for transition times below 1 ns. Further analysis of our simulation results shows that the maximum transition time for $C_A = 50$ fF is smaller than 0.6 ns. The second best microprobes with respect to input capacitance are called *Picoprobe Model 28/29* [19], which exhibit $C_A = 40$ fF regardless of the slew rates of the probed signals.

While the two-inverter **T** implementation can marginally detect all probes according to their specification, its uncertainty region $\Delta C_A^U > 25$ fF is still significant. The size of this region determines both the likelihood of false positives and false negatives, and hence the reliability of the circuit. As conversations with industry representatives have suggested that reliability is one of the most important design goals, we want to evaluate how much the uncertainty region ΔC_A^U can be narrowed by optimizing the LAPD. For the sake of reliability, we chose the four inverter implementation as a starting point as it does not show a nonzero alarm rate at $C_A = 0$.

To get a better understanding about the effects of variations, we were first interested in how strongly the variation of each LAPD stage affects the alarm threshold. In a first set of simulations, we observed the variance of the delay difference between the two latch inputs $\text{Var}(\Delta t_{\text{RS}})$ at $C_A = 0$ to quantify this variation. It is Δt_{RS} that determines the latch output state `Q`, and furthermore, only few simulations are necessary to obtain Fig. 4, as only one C_A sweep point needs to be considered.

TABLE II
VARIANCE OF TIMING DIFFERENCES AT LATCH INPUTS OF FOUR-INVERTER IMPLEMENTATION

	$\text{Var}(\Delta t_{RS,sel=0})$
all variations enabled	$8.15 \times 10^{-23}\,\text{s}^2$
w/o $\mathbf{D_{in}}$ variation	$5.32 \times 10^{-23}\,\text{s}^2$
w/o $\mathbf{D_{out}}$ variation	$2.43 \times 10^{-23}\,\text{s}^2$
w/o **M** variation	$8.00 \times 10^{-23}\,\text{s}^2$
w/o **T** variation	$7.89 \times 10^{-23}\,\text{s}^2$

A technology feature allows to selectively switch OFF variations for single transistors—we used this feature to selectively disable variations stage by stage and quantify the influence on the variance. We captured the results of both the cases `sel` $= 0$ and `sel` $= 1$, but we only noticed minor differences, so our explanations are focused on the first case `sel` $= 0$ for simplicity. The results for the four inverter implementation of the delay element **T** are shown in Table II; the bold letters in Table II refer to the gates in Fig. 3. Notice that disabling the variations in the buffer stage $\mathbf{D_{out}}$ significantly reduces the variance of Δt_{RS}. Therefore, we assume this stage to have the highest influence on reliability at our design point. This is pointed out in more detail in the Appendix.

C. Reliability Metric

Prior to dimensioning the LAPD, we introduce a reliability metric that allows us to compare the quality of different LAPD implementations. We define this metric q as the area between the ideal curve of an LAPD having a detection threshold C_A^* according to the definition in III-B and the curve of an actual implementation.

This approach incorporates all tested C_A points of an implementation and thus minimizes numerical noise. For reasons of computational complexity, the boundary of this area is chosen to be $[0; C_{\text{max}}]$

$$q = \int_0^{C_A^*} f_A(C_A)dC_A + \int_{C_A^*}^{C_{\text{max}}} (1 - f_A(C_A))dC_A \quad (12)$$

with

$$f_A\left(C_A^*\right) = 0.5. \quad (13)$$

Equation (13) centers the threshold C_A^* between the two integrals in (12) around the intrinsic 50% alarm frequency of a circuit. With this, the metric effectively prefers a low uncertainty range over a predefined alarm threshold. We define the condition $f_A(0) < \epsilon$ as an additional filter criterion to sort out false positives. Fig. 7 illustrates the metric for the four-inverter implementation. The plot was generated using Matplotlib [22].

D. LAPD Dimensioning

We conducted a simple optimization on the four-inverter implementation to estimate the minimum C_A that can be detected reliably. As shown before, variations in stage $\mathbf{D_{out}}$ appear to have the strongest influence on the delay difference variations. For this reason, we analyzed how much the overall reliability can be improved by fine-tuning the transistor

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

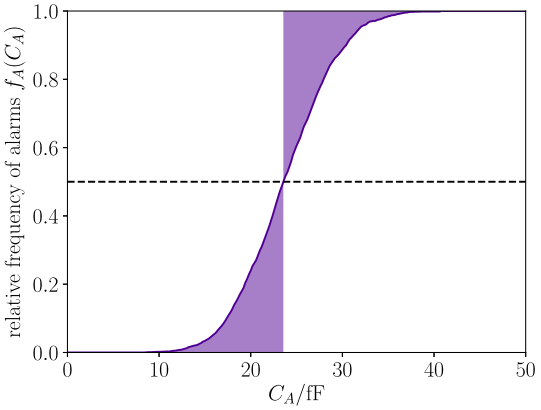

Fig. 7. Illustration of the reliability metric of the four-inverter implementation ($(q/fF) = 4.02$).

TABLE III

THRESHOLD CAPACITANCE AND RELIABILITY METRIC FOR DIFFERENT $\mathbf{D_{out}}$ DIMENSIONS ($\Delta C_A = 5$ fF, 200 MONTE CARLO RUNS)

$\frac{W}{L}$	$\frac{L}{L_{min}}$	$\frac{C_A^*}{\text{fF}}$	$\frac{q}{\text{fF}}$
10	1	23.5	4.46
10	**2**	**23.6**	**3.58**
10	**5**	**24.1**	**3.48**
10	10	23.3	4.80
20	1	25.3	3.67
20	**2**	**24.9**	**3.32**
20	5	24.8	3.85
20	10	24.8	6.72
50	**1**	**26.6**	**3.22**
50	**2**	**26.3**	**3.38**
50	5	25.7	5.43
50	10	24.5	12.47
100	**1**	**26.9**	**3.45**
100	2	26.5	3.89
100	5	25.5	8.15
100	10	17.5	18.53

dimensions of this stage. In a new series of simulations, we performed a coarse sweep over C_A as well as the aspect ratio and the channel length to select good candidates for a further finer analysis

$$\frac{W}{L} \in \{10, 20, 50, 100\}$$
$$L \in \{1, 2, 5, 10\} \cdot L_{\text{min}}$$

$L_{\text{min}} = 0.06\mu$m is the minimum channel length as needed to be specified in the simulator. The results of this set of simulations, for which we used a step size of $\Delta C_A = 5$ fF and 200 Monte Carlo iterations at each point, are shown in Table III.

The "50% alarm capacitance" C_A^*, which is defined in (13), has been estimated by linear interpolation. The best six rows (in bold) with respect to the reliability metric have been selected for a more detailed analysis with 2000 Monte Carlo iterations and a step size of $\Delta C_A = 0.2$ fF.

The results of the finer analysis are shown in Table IV; the separated row represents the initial design, the following lines show the results after optimization. The best case is highlighted in bold.

TABLE IV

THRESHOLD CAPACITANCE AND RELIABILITY METRIC FOR DIFFERENT $\mathbf{D_{out}}$ DIMENSIONS ($\Delta C_A = 0.2$ fF, 2000 MONTE CARLO RUNS)

$\frac{W}{L}$	$\frac{L}{L_{min}}$	$\frac{C_A^*}{\text{fF}}$	$\frac{q}{\text{fF}}$	$\frac{C_{0.01}}{\text{fF}}$	$\frac{C_{0.99}}{\text{fF}}$
10	1	23.6	4.02	11.2	37.2
10	2	23.8	2.95	14.2	33.4
10	5	24.3	2.92	15.0	34.2
20	**2**	**25.3**	**2.77**	**16.4**	**34.4**
50	1	26.8	2.82	17.4	36.0
50	2	26.4	2.93	17.0	36.6
100	1	27.3	2.83	18.2	37.0

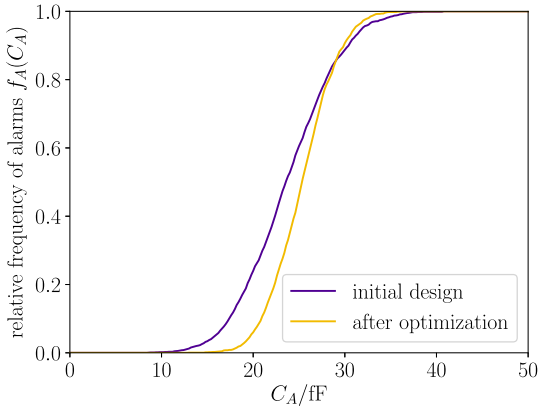

Fig. 8. Relative frequency of alarms of the best circuit, compared with the initial design.

We want to use these results to estimate the real alarm probability $p_A(C_A)$ based on the absolute number of alarms A, the number of Monte Carlo simulations N as well as the desired confidence level α that we assume as $\alpha = 0.01$. We used the Wilson method [23] to estimate the confidence intervals. We used these computed intervals to estimate a "1% alarm threshold" $C_{0.01}$ for which $p_A(C_{0.01}) < 0.01$ holds, as well as a "99% alarm threshold" $C_{0.99}$ for which $p_A(C_{0.99}) > 0.99$ holds.

Compared with the initial design, the quality metric of the best case has improved by more than 40%. If we define the uncertainty region ΔC_A^U as $\Delta C_A^U = C_{0.99} - C_{0.01}$, then the reduction of this region is also a little more than 40%. In other words, we have 40% more margin with respect to timing jitter or C_L imbalance, and we can also more effectively tune the delay elements $\mathbf{T}$ toward a lower C_A^* without increasing the number of false positives too much.

Fig. 8 shows the curve of the optimized implementation next to the initial design. The reduction of the uncertainty region is also clearly visible in Fig. 8.

E. Corners

We have analyzed the behavior of the LAPD after dimensioning with respect to process, voltage, and temperature corners. The process corner points used have been `SS` (slow-slow), `TT` (typical-typical), and `FF` (fast-fast); for the temperature, $\vartheta \in \{0\ °\text{C}, 27\ °\text{C}, 85\ °\text{C}\}$ were used; for the voltage, we used $V_{DD} \in \{1.08\ \text{V}, 1.2\ \text{V}, 1.32\ \text{V}\}$.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE V
ANALYSIS OF CORNERS

		typical	worst case	worst case at corners
initial	$C_{0.01}$	11.2 fF	7.4 fF	FF, 0 °C, 1.08 V
	$C_{0.99}$	37.2 fF	41.6 fF	SS, 85 °C, 1.08 V
optimized	$C_{0.01}$	16.4 fF	12.0 fF	FF, 0 °C, 1.08 V
	$C_{0.99}$	34.4 fF	39.4 fF	SS, 85 °C, 1.08 V

The analysis of corners was used to determine the worst case values of $C_{0.01}$ and $C_{0.99}$. The results are shown in Table V. The two rows labeled "optimized" represent the corner cases of the optimized design [$(W/L) = 20$ and $(L/L_{\min}) = 2$]. For reference, the values of the initial design [$(W/L) = 10, (L/L_{\min}) = 1$] are also given.

We can see that the initial design fails to detect a 40-fF probe in the worst case, while the optimized design has a worst case $C_{0.99}$ slightly below this value. Also, it can be stated that the worst case $C_{0.01}$ keeps away far enough from $C_A = 0$. The worst case uncertainty region ΔC_A^U has reduced from 34.2 to 27.4 fF, which is an improvement by about 20%. This shows that it is possible to achieve reliable operation of the LAPD even without dedicated optimization tools.

F. Model Fit

To verify the model stated in (10) and (11), we first characterized the latch **N** to find out t_H before we analyzed the complete circuit to determine the product $k_{\mathbf{D_{out}}}\Omega$.

We performed a sweep over Δt_{RS} at the latch inputs to estimate the minimum "hold time" t_H for which the output is reliable, that is

$$\Delta t_{\mathrm{RS}} < 0 \;\Rightarrow\; p(\mathrm{Q} = 0) > 0.99 \tag{14}$$

$$\Delta t_{\mathrm{RS}} > 0 \;\Rightarrow\; p(\mathrm{Q} = 1) > 0.99 \tag{15}$$

hold. We used $N = 2000$ Monte Carlo simulations and the Wilson score interval [23] with a confidence level of $\alpha = 0.01$ to estimate the bounds of t_H and obtained a value of $t_H = 1.40$ ps. This value concentrates the mismatch variation of the latch.

To continue the analysis, we solved (8) for

$$k_{\mathrm{D_{out}}}\Omega = \frac{t_D + t_{M2} - t_{M1} - \Delta t_{\mathrm{RS}}}{C'_A} \tag{16}$$

and then determined the mean and variance of this value at different C'_A sweep points by simulation of the complete LAPD circuit; the values of t_D, $t_{\mathbf{M1}}$, and $t_{\mathbf{M2}}$ were captured as well. We then used three sigma distances from the mean $E(k_{\mathbf{D_{out}}}\Omega)$ to estimate the reliability bounds

$$C^*_{0.01} = \frac{t_D + t_{\mathbf{M2}} - t_{\mathbf{M1}} - t_H}{E(\Omega\, k_{\mathbf{D_{out}}}) + 3\sqrt{\mathrm{Var}(\Omega\, k_{\mathbf{D_{out}}})}} \tag{17}$$

$$C^*_{0.99} = \frac{t_D + t_{\mathbf{M2}} - t_{\mathbf{M1}} + t_H}{E(\Omega\, k_{\mathbf{D_{out}}}) - 3\sqrt{\mathrm{Var}(\Omega\, k_{\mathbf{D_{out}}})}}. \tag{18}$$

Table VI shows the approximated threshold capacitances using $N = 2000$ Monte Carlo simulations at different sweep points. The reference values are listed in the "typical" column of Table V. One can see that with increasing C'_A, the difference

TABLE VI
APPROXIMATED THRESHOLD CAPACITANCES

		C'_A sweep point			
		10 fF	20 fF	30 fF	40 fF
initial design	$\frac{C^*_{0.01}}{\mathrm{fF}}$	10.0	13.7	15.4	16.3
	$\frac{C^*_{0.99}}{\mathrm{fF}}$	< 0	75.6	44.4	36.3
optimized design	$\frac{C^*_{0.01}}{\mathrm{fF}}$	13.0	16.8	18.4	19.3
	$\frac{C^*_{0.99}}{\mathrm{fF}}$	258.5	47.3	36.6	32.7

TABLE VII
AREA, TIMING, AND ENERGY COMPARISON OF CRYPTOGRAPHICALLY SECURE SHIELDS, PAD, AND LAPD

	CSS [11]	PAD [7]	LAPD
area (Gate Equivalents)	8081	549	48
number of cycles	*(runs continuously)*	50-100	2
energy consumption [fJ]	$7.01 \times 10^{12}\,\mathrm{s}^{-1}$	n/a	895

$C^*_{0.99} - C^*_{0.01}$ shrinks. Also, the error of $C^*_{0.01}$ increases; in absolute terms, the maximum error occurs with the initial design at $C'_A = 40$ fF is $\Delta C_{0.01} = 16.3$ fF -11.2 fF $= 5.1$ fF. On the other hand, an increasing C'_A also leads to a reduction of error for the $C_{0.99}$ estimation: At $C'_A = 40$ fF, the worst case occurs at the optimized design at $\Delta C_{0.99} = 32.7$ fF-34.4 fF $= -1.7$ fF. In relative terms, the approximation is more accurate for $C_{0.99}$ at higher values of C'_A than for $C_{0.01}$ at low values of C'_A. Therefore, it seems recommendable to focus on these values when using the model.

The proposed linear model appears sufficient for qualitative comparisons between different LAPD implementation variants and helps to significantly reduce the number of required simulations; for precise quantitative analyses, a more elaborate model seems advisable.

G. Resource Usage

A quantitative area, timing, and energy consumption comparison between the cryptographically secure shields by Cioranesco *et al.* [11], the PAD [7], and the LAPD is shown in Table VII. All three implementations are available in the same STMicroelectronics 65-nm technology (the PAD implementation in this technology is currently not published). The LAPD dimensions in terms of gate equivalents were determined by normalizing to the sum of transistor dimensions of the smallest size standard cell NAND gate `HS65_LS_NAND2X2`. The dimensions of the cryptographically secure shields and the PAD, both after layout, were normalized to the layout area of the same NAND gate.

It can be seen that the LAPD is one order of magnitude smaller than the PAD and more than two orders of magnitude smaller than the cryptographically secure shields.

The cryptographically secure shields are designed to run continuously, while PAD and LAPD shall only be used prior to security critical operations. For this reason, the energy consumption of the cryptographically secure shields is given per second. Compared with the PAD, the LAPD is faster by a factor of 25–50. The energy consumption of the LAPD was simulated for one test run at $C_A = 0$. Even assuming that the

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

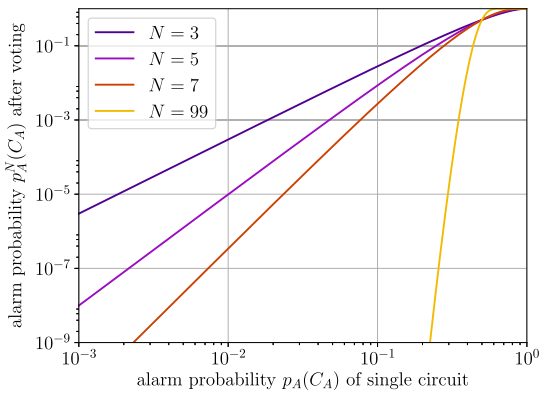

Fig. 9. Alarm probability after majority voting.

cryptographically secure shields would, for example, only run for 1 s, its energy consumption is larger than the one of the LAPD by several orders of magnitude.

H. Error Compensation

We have provided error probability bounds based on simulations of the aforementioned variations. However, the computational complexity of the simulations only allows us to provide error probability bounds in the magnitude of 10^{-2}. Assuming the statistical independence, voting schemes can be used to significantly improve the error probability, for example as proposed by Parhami [24].

1) Local variations can be compensated by providing k LAPD instances.
2) Timing jitter can be compensated by repeating the evaluation k times.

Note that the assumption of statistical independence is not true for global variations as well as voltage and temperature variations; however, focusing on the local variations can already lead to a significant improvement due to the differential mode of operation of the LAPD.

As an example, majority voting can be used as voting scheme. In this case, k should be odd such that at least $((k+1)/2)$ alarm votes are required to raise an alarm. If we assume that the alarm probability of a single LAPD instance evaluation at a certain operating point is $p_A(C_A)$, the alarm probability after voting follows a binomial distribution:

$$p_A^k(C_A) = P\left(X \geq \frac{k+1}{2}\right) = \sum_{i=\frac{k+1}{2}}^{k} \binom{k}{i} p_A(C_A)^i \left(1 - p_A(C_A)\right)^{N-i}.$$

This distribution shows a tendency toward its extremes

$$\lim_{k\to\infty} p_A^k(C_A) = \begin{cases} 0 & p_A(C_A) < 0.5 \\ 0.5 & p_A(C_A) = 0.5 \\ 1 & p_A(C_A) > 0.5. \end{cases}$$

Assuming that an alarm for a specific C_A for which $p_A(C_A) < 0.5$ holds is called false positive, Fig. 9 allows us to quantify the reduction of false positives. For example, voting with $k = 5$ for a single-instance alarm probability, $p_A(C_A) = 10^{-2}$ leads to an overall alarm probability of $p_A^k(C_A) = 10^{-5}$. Likewise, this approach also reduces false negatives.

V. Further Work

We have shown that the LAPD is able to work reliably using a simple manual optimization approach. As a next step, the optimization can be improved, for example by using gradient based optimization tools.

Also, the LAPD shall be implemented in silicon for practical results.

Still, a desirable yet unavailable feature of the LAPD is the ability to compensate manufacturing variations and line length imbalances. A next generation probing detector shall have these features.

VI. Conclusion

We have presented an LAPD that can detect the presence of microprobes by comparing delays introduced by the capacitive loads of bus lines to those introduced by delay elements. The circuit only consists of a few gates and has a significantly lower area than other protection mechanisms, such as the PAD [7] or bus encryption.

We have analyzed the reliability of such a detector with respect to local variations as well as process, voltage, and temperature corners using Monte Carlo simulations on a 65-nm technology. The results of these simulations have been used to estimate the regions of probe capacitances in which the circuit gives reliable results. These results show that an initial LAPD implementation can detect state-of-the-art commercial microprobes under typical conditions, but possibly not in worst case scenarios.

We have performed a simple optimization of the LAPD with the goal of reducing the capacity threshold for undetectable probes. With optimizing only one single stage of the LAPD, the uncertainty region could be reduced by 40% under nominal conditions as well as 20% for the corners. After optimization, the previously mentioned microprobes can also be detected in the worst case scenario.

Appendix

The LAPD principle of operation is based on the detection of the delay difference (Δt_{RS}) arriving at the RS latch inputs (the arbiter). The transitions arriving at the latch are delayed by chains d_1 and d_2 and t_D and t_D being switched during the two operating cycles in each one of the chains. Process variations alter the propagating delay of these three chains in such a way that the magnitude of Δt_{RS} becomes unstable at a certain degree and therefore less predictable.

These effects cannot be avoided completely but diminished at a certain degree as it is seen in Section IV-D. In particular, after a first assessment, it is observed that inverter $\mathbf{D_{out}}$ has a significantly larger influence on the variability than the rest of stages and therefore the optimization is further concentrated on this inverter.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

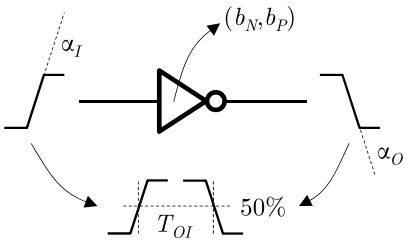

Fig. 10. Input/output slew-rates and delay of an inverter.

To understand why inverter $\mathbf{D_{out}}$ has a larger influence on the variability than the rest of stages, we can focus on the delay propagating model of a single inverter that was presented by Shoji [25], and is summarized in the following paragraphs.

In Fig. 10, a simple inverter is shown with the corresponding input/output transitions. Input/output slew rates (V s^{-1}) are α_I and α_O, respectively. The delay of the gate is calculated at 50% of the signal levels and is symbolized by T_{OI}. Internally, the pMOS and nMOS transistors have transconductances (A V^{-1}) that are represented by b_N and b_P, respectively, and have a big contribution to the switching speed of the output, together with the load capacitance.

Now, we will analyze the two possible scenarios in which the variability can disturb the propagating delay.

A. Inverter Delay Is Independent of the Input Slew-Rate

Intuitively, we know that if the input slew-rate is extremely fast ($\alpha_I \to \infty$), the propagating delay ($T_{OI} \to t_\infty$) will exclusively depend on the inverter load capacitance and the (dis-) charging transistor transconductance, b_N for the case in Fig. 10. In each inverter, the delay be affected by the variations of the (dis-)charging transistor transconductance and the load capacitance dimensions (typically the input of the next stage). These two elements will generate variability in the propagating delay (t_∞), but it will be independent of the input slew-rate variability produced by the previous stage. Therefore, the total delay variability of a chain of inverters will be the sum of the independent variabilities of each inverter stage, and it will typically become a normal random distributed variable whose variance will be the sum of the variances of each inverter delay.

This scenario is the most favorable in terms of reducing the effects of process variabilities. Minimizing tactics are fundamentally based on placing strategies and enlarging the dimensions of transistors in order to reduce the percentage of the variability over the total physical dimensions. However, this is at the cost of more area and is only partially applied until the range of the circuit tolerance is achieved.

B. Inverter Delay Is Dependent of the Input Slew-Rate

When the input slew-rate (α_I) is significantly smaller than the output slew-rate (α_O), the delay of gate (T_{OI}) becomes sensitive to it too. Particularly, the degree of sensitivity follows a hyperbolic function whose growing degree depends on the ratio between pMOS and nMOS transconductances. Therefore, if we consider the previous inverter controlling the input slew-rate, its variability will propagate to the next inverter through the variability in the slew-rate, and at the same time, it will affect the next stage delay with a contribution much stronger than the simpler addition seen in our previous scenario.

In the LAPD circuit in Fig. 3, this effect is clearly observed in the inverter $\mathbf{D_{out}}$, because it receives the input from heavily loaded bus-lines, and the output drives smaller gates like the internal delay chain **T** and the multiplexer **M**.

The reduction of the variability effects in this scenario is achieved by doing a proper balance of the pMOS and nMOS transconductances as it will be clear from the Shoji delay model presented in the following.

C. Delay Model Under Variable Input Slew-Rate Conditions

Let us first define the transconductance ratio $\beta = b_N/b_P$, the normalized input slew-rate $S_l = \alpha_I/\alpha_O$, and the normalized inverter delay $T_{\text{inv}} = T_{OI}/t_\infty$.

According to Shoji [25], the delay can be approximated by the following closed expressions if the transistor models are linearized and fixed to a zero threshold voltage

$$T_{\text{inv}} = 1 + \frac{1}{2(1+\beta)}\frac{1}{S_l}; \quad S_l \geq \frac{\beta}{2(1+\beta)}$$
$$T_{\text{inv}} = 2\sqrt{\frac{\beta}{2(1+\beta)}\frac{1}{S_l}} + \frac{1-\beta}{2(1+\beta)}\frac{1}{S_l}; \quad S_l < \frac{\beta}{2(1+\beta)}. \tag{19}$$

Interestingly, for more realistic transistor models (including nonzero threshold voltages), Shoji shows that the dependence of T_{inv} from S_l will follow the same law despite a closed expression could not be found.

Equation (19) is plotted in Fig. 11. At the x-axis, the normalized input slew-rate is fixed and it has two main regions (separated by a dotted line): larger and smaller than 1. For values larger than 1, the input transition is faster than the output, while for values smaller than 1, the input transition becomes slower than the output one. At the y-axis, the normalized delay is presented. When it is 1, the delay of the inverter is exactly t_∞ and is equal to the delay when the inverter input switches very fast. Each one of the curves represents a different β ratio going from 0.25 to 4.

When S_l is higher than 1, all the curves closely coincide and are almost equal to 1. This shows that the delay of the inverter is almost independent of the input slew-rate S_l and that the transconductance ratio β does not have any importance with respect to the process variations.

When S_l is smaller than one, curves diverge and thus the sensitivity of the inverter delay becomes stronger to the input slew-rate S_l. This is clearly seen in the curve $\beta = 0.25$, that for a variation of S_l from 0.05 to 0.042 (a relative change of 13%), the normalized inverter delay changes from 6 to 8 approximately (a relative change of 29%). This strong dependence can be reduced by tuning the β ratio at the proper value. In the plot, a dotted rectangle indicates the region of the best design. The transconductance ratio should be adjusted such that the normalized delay is kept inside this region.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

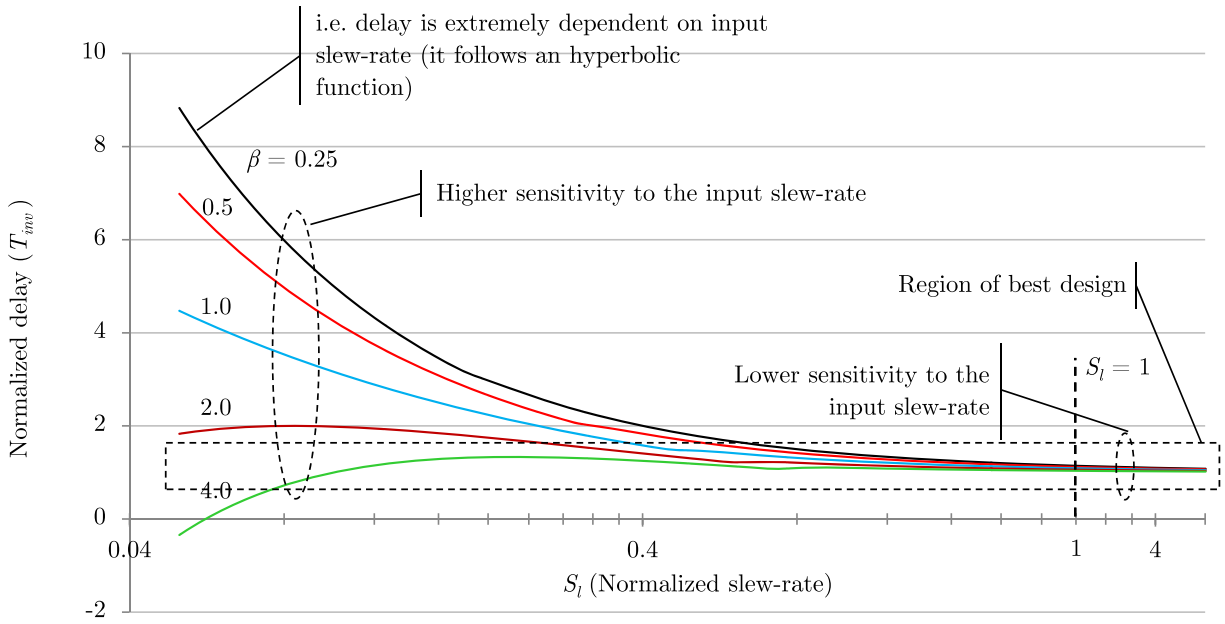

Fig. 11. Variation of the normalized inverter delay as a function of the normalized input slew-rate [25].

While this process cannot be done analytically, given the complexity of the transistor models, simulations can be used to find the best transconductance ratio β, i.e., like for the critical inverter $\mathbf{D_{out}}$.

REFERENCES

[1] R. Anderson and M. Kuhn, "Tamper resistance—A cautionary note," in *Proc. 2nd Conf. Proc. 2nd USENIX Workshop Electron. Commerce (WOEC)*, vol. 2. Berkeley, CA, USA, 1996. [Online]. Available: http://dl.acm.org/citation.cfm?id=1267167.1267168

[2] J. Heyszl, S. Mangard, B. Heinz, F. Stumpf, and G. Sigl, *Localized Electromagnetic Analysis of Cryptographic Implementations*. Berlin, Germany: Springer-Verlag, 2012, pp. 231–244. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-27954-6_15

[3] N. Homma *et al.*, *EM Attack Is Non-invasive?—Design Methodology and Validity Verification of EM Attack Sensor*. Berlin, Germany: Springer-Verlag, 2014, pp. 1–16. [Online]. Available: http://dx.doi.org/10.1007/978-3-662-44709-3_1

[4] N. Homma, Y.-I. Hayashi, N. Miura, D. Fujimoto, M. Nagata, and T. Aoki, "Design methodology and validity verification for a reactive countermeasure against EM attacks," *J. Cryptol.*, vol. 30, no. 2, pp. 373–391, 2015. [Online]. Available: http://dx.doi.org/10.1007/s00145-015-9223-3

[5] C. Tarnovsky, "Deconstructing a 'secure' processor," presented at the Conf. BlackHat DC 2010, Washington, DC, USA, 2012. [Online]. Available: http://www.blackhat.com/html/bh-dc-10/bh-dc-10-archives.html

[6] "Integrity guard—The newest generation of digital security technology," Infineon Technol. AG, Neubiberg, Germany, White Paper 04_12, Sep. 2012. Accessed: Nov. 7, 2016. [Online]. Available: https://www.infineon.com/cms/en/applications/smart-card-and-security/integrity-guard/#!documents

[7] S. Manich, M. S. Wamser, and G. Sigl, "Detection of probing attempts in secure ICs," in *Proc. IEEE Int. Symp. Hardw.-Oriented Secur. Trust (HOST)*, Jun. 2012, pp. 134–139.

[8] M. Weiner, S. Manich, and G. Sigl, "A low area probing detector for power efficient security ICs," in *Radio Frequency Identification: Security and Privacy Issues*, vol. 8651. Berlin, Germany: Springer, Jul. 2014.

[9] M. Wan, Z. He, S. Han, K. Dai, and X. Zou, "An invasive-attack-resistant PUF based on switched-capacitor circuit," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 62, no. 8, pp. 2024–2034, Aug. 2015.

[10] Y. Ishai, A. Sahai, and D. Wagner, *Private Circuits: Securing Hardware against Probing Attacks*. Berlin, Germany: Springer-Verlag, 2003, pp. 463–481. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-45146-4_27

[11] J.-M. Cioranesco *et al.*, "Cryptographically secure shields," in *Proc. IEEE Int. Symp. Hardw.-Oriented Secur. Trust (HOST)*, May 2014, pp. 25–31.

[12] Q. Shi, N. Asadizanjani, D. Forte, and M. M. Tehranipoor, "A layout-driven framework to assess vulnerability of ICs to microprobing attacks," in *Proc. IEEE Int. Symp. Hardw. Oriented Secur. Trust (HOST)*, May 2016, pp. 155–160.

[13] S. P. Skorobogatov, "Semi-invasive attacks—A new approach to hardware security analysis," Comput. Lab., Univ. Cambridge, Cambridge, U.K., Tech. Rep. UCAM-CL-TR-630, Apr. 2005. [Online]. Available: http://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-630.pdf

[14] E. Brier, C. Clavier, and F. Olivier, "Correlation power analysis with a leakage model," in *Cryptographic Hardware and Embedded Systems—CHES* (Lecture Notes in Computer Science), vol. 3156, M. Joye and J.-J. Quisquater, Eds. Berlin, Germany: Springer-Verlag, 2004, pp. 16–29. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-28632-5_2

[15] T. Sakurai and A. R. Newton, "Alpha-power law MOSFET model and its applications to CMOS inverter delay and other formulas," *IEEE J. Solid-State Circuits*, vol. 25, no. 2, pp. 584–594, Apr. 1990.

[16] K. A. Bowman, B. L. Austin, J. C. Eble, X. Tang, and J. D. Meindl, "A physical alpha-power law MOSFET model," in *Proc. ACM Int. Symp. Low Power Electron. Design (ISLPED)*, New York, NY, USA, 1999, pp. 218–222. [Online]. Available: http://doi.acm.org/10.1145/313817.313930

[17] A. Balankutty, T. C. Chih, C. Y. Chen, and P. Kinget, "Mismatch characterization of ring oscillators," in *Proc. IEEE Custom Integr. Circuits Conf. (CICC)*, Sep. 2007, pp. 515–518.

[18] GGB Industries, Inc. *Picoprobe Model 18C & Picoprobe Model 19C, Datasheet*. Accessed: Nov. 7, 2016. [Online]. Available: http://www.ggb.com/PdfIndex_files/mod18c.pdf

[19] GGB Industries, Inc. *Picoprobe Models 28 & 29, Datasheet*. Accessed: Nov. 7, 2016. [Online]. Available: http://www.ggb.com/PdfIndex_files/mod28.pdf

[20] H. Bar-El, H. Choukri, D. Naccache, M. Tunstall, and C. Whelan, "The sorcerer's apprentice guide to fault attacks," *Proc. IEEE*, vol. 94, no. 2, pp. 370–382, Feb. 2006.

[21] M. Weiner and S. M. Bou, "The salvador simulation framework," in *Proc. TRUDEVICE Workshop*, Nov. 2016, pp. 1–2.

[22] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 90–95, May 2007.

[23] E. B. Wilson, "Probable inference, the law of succession, and statistical inference," *J. Amer. Stat. Assoc.*, vol. 22, no. 158, pp. 209–212, 1927. [Online]. Available: http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1927.10502953

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[24] B. Parhami, "Voting algorithms," *IEEE Trans. Rel.*, vol. 43, no. 4, pp. 617–629, Dec. 1994.
[25] M. Shōji, *CMOS Digital Circuit Technology*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1988.

Michael Weiner received the B.Eng. degree in electrical engineering from the Baden-Württemberg Cooperative State University, Stuttgart, Germany, and the M.Sc. degree in electrical engineering from the Technical University of Munich, Munich, Germany, where he is currently working toward the Ph.D. degree.

He is also a Firmware Engineer with SimonsVoss Technologies GmbH, Unterföhring, Germany, where he is involved in the security of electronic locking systems. His current research interest is embedded systems security, including detectors of invasive attacks and analyzing real-life products.

Salvador Manich received the M.S. and Ph.D. degrees in industrial engineering from the Universitat Politècnica de Catalunya, Barcelona, Spain, in 1992 and 1998, respectively.

He has been an Associate Professor with the School of Industrial Engineering, Barcelona, since 2001 and a member of the Electronic Engineering Department, Barcelona. He is also with the Quality in Electronics Group, Barcelona, where he develops his research activity, and he is also a member of the Center for Research in Nanoengineering, Barcelona. He was an Invited Researcher with Instituto Superior Técnico, Lisbon, Portugal, and Technical University of Munich, Munich, Germany. His current research interests include low-power design, test of digital systems, and security in hardware structures.

Rosa Rodríguez-Montañés received the M.S. degree from the Universitat de Barcelona, Barcelona, Spain, in 1988 and the Ph.D. degree in physical science from the Universitat Politècnica de Catalunya (UPC), Barcelona, Spain, in 1992.

Since 1994, she has been an Associate Professor with the Department of Electronic Engineering, UPC. In 2002, she spent her sabbatical leave with the Test Group, Philips Research, Eindhoven, The Netherlands. Her current research interests include fault models, defect characterization, defect diagnosis, and hardware security of nanometric CMOS technologies.

Georg Sigl received the Ph.D. degree in electrical engineering from the Technical University of Munich, Munich, Germany, in 1992, with a focus on the area of layout synthesis.

Afterward, he introduced new design-for-testability concepts in telecommunication ASICs at Siemens, Munich, Germany. In 1996, he joined the Automotive Microcontroller Department, Infineon, Munich, Germany, to develop a universal library for peripherals to be used in 16- and 32-bit microcontrollers. Since 2000, he has been responsible for the development of new secure microcontroller platforms in the Chip Card and Security Division. Under his responsibility, two award winning platforms have been designed. In 2010, he founded the new Chair of Security in Information Technology at Technical University of Munich. In parallel, he drives embedded security research as the Director of the Fraunhofer Research Institute for Applied and Integrated Security AISEC Munich, Garching b. Munich, Germany.