Laser fault injection at the CMOS 28 nm technology node: an analysis of the fault model

To cite this version:

HAL Id: emse-01856008
https://hal-emse.ccsd.cnrs.fr/emse-01856008
Submitted on 11 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Laser Fault Injection at the CMOS 28 nm Technology Node: an Analysis of the Fault Model

Jean-Max Dutertre*, Vincent Beroulle†, Philippe Candelier‡, Stephan De Castro*§, Louis-Barthelemy Faber‡, Marie-Lise Flottes§, Philippe Gendrier†, David Hély§, Regis Leveugle†, Paolo Mastrì†, Giorgio Di Natale‡, Athanasios Papadimitriou‡, and Bruno Rouzeyre§,

*Mines Saint-Etienne, CEA Tech, Centre CMP, F-13541 Gardanne France, name@emse.fr
†Univ. Grenoble Alpes, CNRS, Grenoble INP, TIMA, 38000 Grenoble, France, surname.name@univ-grenoble-alpes.fr
‡STMicroelectronics, 850 rue Jean Monnet, 38926 Crolles, firstname.name@st.com
§LIRMM, University of Montpellier, CNRS, 161, rue Ada, 34095, Montpellier, France, name@lirmm.fr
¶Univ. Grenoble Alpes, Grenoble INP, LCIS, 26000 Valence, France, surname.name@esisar.grenoble-inp.fr

Abstract—S. Skorobogatov and R. Anderson identified laser illumination as an effective technique to conduct fault attacks in 2002. In these early days of laser-induced fault injection, it was proven to be possible to inject single-bit faults into integrated circuits. This corresponds to the more restrictive fault model found in the fault attack bibliography. The target area under laser illumination (a few micrometers, down to $\sim 1 \mu m$) broadly matched that of a single transistor. It was consistent with a single-bit fault model. However, since then the technology of secure devices has evolved. In current circuits even the smallest laser spots may illuminate several logic cells. This raises the question of the validity of the single-bit fault model: does it still hold? In this work, we report an assessment of its validity through experimental results obtained from circuits designed at the 28 nm CMOS technology node. We also describe the main properties of the corresponding fault model obtained from both static and dynamic experiments.

I. INTRODUCTION

A wide range of physical attacks targets secure circuits. Among these, Fault Attacks (FAs) are based on the alteration of the circuit environment in order to change its behavior or to induce faults into its computations. FA is an active attack technique that aim at inducing an information leakage from a targeted Integrated Circuit (IC). Many means exist to inject faults into an IC, mostly based on the distortion of the chip environmental conditions, such as, voltage or clock glitches, temperature increase, electromagnetic perturbations, or laser exposure. Therefore, research works have been done to understand and mitigate fault attacks.

The use of a laser beam to inject faults into the computations of a secure IC was first reported by S. Skorobogatov and R. Anderson in 2002 [1]. Since then, laser is considered as a very efficient tool to carry out FAs. It permits an accurate injection of faults both in space and time [2, 3, 4, 5]. Besides, despite the scaling down of IC’s technologies, it was considered to be a practical means to inject faults with a high resolution (at byte or even at bit level [6]), which is mandatory to apply most of the known FA schemes [2, 3]. However, the assumption that laser fault injection (LFI) is still able to induce bit level faults is regularly brought into question as CMOS technology has continued to scale down: at cutting edge technology nodes several logic gates may be simultaneously illuminated by the smallest achievable laser spot ($\sim 1 \mu m$ due to the laws of optic). Then, if several gates are simultaneously disturbed by a single laser shot, laser may proves unable to induce single-bit faults (as faults encompassing a single bit are called).

In this work, we report an assessment of the ability of laser-based fault injection to induce single-bit faults through experimental results obtained at the 28 nm CMOS technology node. We carried out both static and dynamic experiments, respectively on test patterns of D flip-flops and also on an implementation of the Advanced Encryption Standard (AES [7]). We also studied the corresponding fault model properties in terms of fault size (i.e. the ability to inject single-bit or single-byte faults) and of data-dependence.

This article is organized as follows: section II provides a reminder on the theory of laser-induced fault injection attacks. It also discusses the importance of the fault model properties. Section III reports the experimental results obtained on D flip-flop test patterns. Then, section IV describes the properties of the faults induced by laser into a running AES hardware block. Finally, our findings are summarized and discussed in the last section.

II. THEORY OF LASER-INDUCED FAULT INJECTION ATTACKS

A. Theory of fault attacks - Short reminder.

The concept of fault attacks was introduced by Boneh et al. in 1997. They described how the injection of a fault into the computations of the RSA encryption algorithm makes it possible to retrieve its secret key. This principle of information extraction from a running IC by means of fault injection was then extended to other encryption algorithms both asymmetric or symmetric (we refer the reader to [8, 2] for a throughout description of the existing FA techniques).

These techniques involved the observance of strong requirements related to the characteristics of the injected faults, such as, the moment of the fault injection w.r.t. the sequence of calculations of the algorithm, the size of the injected fault (i.e. the number of faulted bits or bytes), or the duration of the laser pulse. When building a Fault Model (FM), all these
characteristics are considered. Each FA scheme has its own fault model which may be more or less difficult to achieve in practice. In this work we study the ability of LFI to attain two restrictive features of the related fault model in terms of fault size and of data dependence.

The more restrictive fault models are generally associated with the injection of faults restricted to one bit or one byte of data: the so-called single-bit and single-byte fault models. These FMs may be uneasy to obtain in practice however they usually lead to very efficient FAS. This raises the question of their practicability as the technology of ICs scales down continuously.

Two FMs are related to the data dependence of the fault injection process: the bit-flip FM and the bit-set/reset FM (considering fault injection at bit level). A bit-flip corresponds to the injection of a fault irrespectively of the state of the faulted bit (either a logic 0 or 1). On the other hand, a bit-reset (respectively a bit-set) describes a fault injection that forces the target bit to 0 (resp. to 1). For a bit-reset (resp. bit-set), if the target bit value was already at 0 (resp. at 1), the fault has no effect. In other words, the bit-flip fault model is data-independent, while the bit-set/reset fault model is data dependent. Therefore, the latter provides additional information on the targeted bit, which can make an attack easier (e.g. by conducting a safe error attack 10).

B. Theory of laser-induced fault injection

Laser may be used to inject faults into ICs because of the photoelectric effect resulting from its interaction with silicon. When a laser beam with a wavelength corresponding to an energy level higher than the silicon bandgap passes through silicon, it creates electron-hole pairs along its path (the so-called photoelectric effect 11). These charge carriers may recombine without any noticeable effect. An exception exists when the laser beam passes through a transistor’s reverse biased PN junction (drain/bulk or source/bulk): a place where there exists a strong electric field. As a consequence, the charge carriers drift in opposite directions and a current pulse is induced. This photocurrent pulse vanishes as the charges are exhausted. It may last a few hundreds of picoseconds after the laser pulse ceased 3. This current pulse in turn creates a transient voltage pulse, which may propagate through the circuit’s logic and may induce a fault in its operations: a so-called Single-Event Transient (SET).

Laser faults may also be induced directly in memory elements, e.g. RAM or registers. This phenomenon is called a Single-Event Upset (SEU): it is exemplified in figure 1 for the case of a D latch.

Figure 1(a) depicts the schematic of a D latch for a basic implementation. Its core part is made of two cross-coupled inverters inv1 and inv2 (transistors Mn1/Mp1 and Mn2/Mp2 resp.) for data memorization. A pass-gate PG1 (transistors Mnt1/Mpt1) is used to access the latch. A second pass-gate PG2 (transistors Mnt2/Mpt2) is used to open or close the memorization loop. For EN = 0 (the latch enable input), the D latch is in write mode with PG1 ON (i.e. passing) and PG2 OFF (i.e. none-passing). On the other hand, for EN = 1, the D latch is in hold (or memorization) mode with PG1 OFF and PG2 ON. An SEU may arise when in hold mode. In this instance, the location of the laser-sensitive reverse biased PN junctions of the D latch is data dependent. They are highlighted in Fig. 1(a) with a color code: red (resp. blue) when the D latch stores a logic 1 as Q = 0 (resp. a logic 0 as Q = 1). As an illustration, consider the red-marked drain of Mn1: if hit by a laser pulse while Q = 0, the node Qb will undergo an SET hence passing from 1 to 0. In turn, the SET will propagate through inv2 and PG2 inducing a transition of node Q from 0 to 1. As a result, the D latch reaches an opposite steady state with Q = 1, it will not revert to its previous state and the stored value is altered (a so-called SEU).

The D latch laser sensitive areas are also highlighted in Fig. 1(b) which displays the latch layout (it was drawn according the assumption that the size of a transistor drain was close to that of a 1 µm laser spot, as an example 12 illustrates...
The laser source we used has the advantage of inducing a fault irrespectively of the target state, since the laser pulse is directed to an overlapping area. This corresponds to the single-bit fault models. If a laser attack is directed toward the bit-set/reset areas resp. highlighted in red and blue (courtesy of [14]).

Note that only two laser-sensitive areas are simultaneously activated depending of the D latch state. Fig. 1(b) is drawn according the assumption that every laser effect area does not extend significantly beyond the laser-sensitive diffusions of transistors. Which is consistent with both the bit-set/reset and single-bit fault models. If a laser attack is directed toward the red area of Mn1 in Fig. 1(b) a fault shall be induced provided that $Q = 0$ (consistency with the bit-set FM). Moreover, the assumption of a reduced effect area implies that the neighboring cells will not be affected (consistency with the single-bit FM). Whereas Fig. 1(c) illustrates the assumption that the laser sensitive areas extend far beyond the transistors diffusions. As a result, they may overlap, which is consistent with the ability of inducing a fault irrespectively of the target state, since the laser pulse is directed to an overlapping area. This corresponds to the bit-flip FM and also to a difficulty to meet the single-bit FM as a single laser shot may fault several neighboring cells.


In the early days of laser fault injection, the single-bit and bit-set/reset FMs were achieved and reported [1], [3], [6]. The most recent state-of-the-art was obtained at the 40 nm CMOS technology node on memory elements in static mode.

[14] reports LFI experiments carried out on a 40 nm custom designed D flip-flop (DFF). Fig. 2 (courtesy of [14]) displays the laser-sensitivity map drawn with a Near Infrared (NIR) laser ($1 \mu m$ laser spot diameter, picosecond range duration, $0.7 nJ$ laser energy). The $4 \mu m \times 2 \mu m$ DFF features the SEU sensitive areas of the two D latches found in every DFF (the master and slave latches). The shapes of their laser-sensitive areas are well defined and match those of Fig. 1(b) they are consistent with the bit-set/reset and single-bit FMs. Note that the laser sensitive areas of the master latch were obtained with the clock signal at 1 (master in memory mode, slave in transparent mode), and that of the slave latch with the clock signal set at 0.

The same year, the authors of [15] reported LFI experiments on a RAM memory of a CMOS 45 nm programmable device (the block RAM of this FPGA). Their experiments were also carried out through the backside of their target with a NIR laser ($4 \mu m$ laser spot diameter, picosecond range duration, $nJ$ range laser energy). They ascertained the possibility of inducing single-bit faults with these settings. The results they obtained were also consistent with the bit-set/reset FM.

III. EXPERIMENTAL RESULTS: STATIC LASER TESTING OF CMOS 28 NM D FLIP-FLOPS

A. Experimental Setup.

Fault injection setup: The laser source we used has the following characteristics: 1,030 nm wavelength (NIR), a laser pulse duration of $30 ps$ and an energy ranging from 0 to 100 $nJ$. The optical path outputs a laser spot diameter of $1 \mu m$, $5 \mu m$ or $20 \mu m$ depending on the chosen lens (diameters of the gaussian laser beams were measured using the knife-edge technique [16] and defined at FWHM as expressed in [3]). An infrared camera was used to adjust the focus of the spot. Fault injection was performed through the backside of the targeted chip (i.e. through its silicon substrate which was thinned to a ~ $100 \mu m$ thickness to lessen the attenuation of the laser beam energy as it travels through the substrate: note that the use of a NIR laser source is mandatory to access the laser-sensitive parts of an IC through its substrate [3]). The optical lens is attached to a motorized XYZ stage with a minimum displacement step of $0.1 \mu m$.

Laser-Sensitivity Map Drawing Process: using a PC to automate the process, we moved the laser over the area of the targeted cells by elementary displacement steps of various lengths. For each position, we shot the laser after writing the cell to 0, then shot again after writing it to 1 and read the stored value after each shot. This allowed us to draw XY maps of laser-sensitivity. For each position where a fault was recorded, we drew a dot colored according the obtained FM. Such laser-sensitivity maps were drawn at various laser energies.

Targets: The experiments reported in this section were carried out exclusively on DFFs we designed on purpose (their design was chosen close to those found in various design kits for the sake of representativity). They were assembled in two shift registers arranged either in line (10 DFFs) or in a matrix (64 DFFs) shapes of a CMOS 28 nm test chip. The reported experiments were carried out in static mode, i.e. the memory cells were in their memorizing mode (clock signal set at 0 or 1).

B. Experimental Results.

 Matrix-shaped shift register: the left part of Fig. 3 gives the arrangement of the matrix-shaped shift register; it is arranged in two blocks of 32 DFFs (DFF cell layout: $1.2 \mu m \times 4.5 \mu m$) with a short space between them. The right part of Fig. 3 displays the laser-sensitivity map of the DFF matrix obtained with the clock signal set at 0 and the DFFs initialized at 1. Hence, it shows the bit-set/reset sensitive areas of the DFFs slave latches. It was drawn for a 0.5 $nJ$ laser energy, a XY...
displacement step of 1 \( \mu \text{m} \), and a laser spot diameter of 1 \( \mu \text{m} \).

It illustrates the level of accuracy that may be achieved by the fault injection process at 28 nm: the slave latches sensitive areas of neighboring DFFs form horizontal patterns separated by voids corresponding to the master latches (which are in transparent none-sensitive mode). The space between the two blocks of DFFs is also clearly visible. From the 136 faults reported in Fig. 3, 96 were single-bit faults, the other 40 injected faults were 2-bit wide.

For its part, Fig. 4 highlights in 3D a similar laser-sensitivity map obtained with the same laser settings but a laser energy increased twofold to 1 nJ. The third dimension is used to give the rank of the faulted DFFs in the shift register. It proved possible to fault every of its DFFs. A color code indicates the number of DFFs simultaneously faulted at a given location. 215 laser shot locations were associated with fault injection, of which: 149 single-bit faults, 62 2-bit faults and 4 3-bit faults.

The rates of single-bit fault injection were almost equal at 0.5 nJ and 1 nJ (70 % and 69 % resp.).

We also carried out LFI experiments with the DFFs initialized at 1. The obtained results were similar. The fault injection threshold (i.e. the energy level corresponding to the first injection of faults) for bit-reset faults was found equal to 0.3 nJ, that for bit-set faults at 0.4 nJ.

Another series of experiments on the matrix-shaped shift register were conducted with a laser spot diameter of 5 \( \mu \text{m} \) and a XY displacement step of 5 \( \mu \text{m} \). The fault injection threshold was found at 0.4 nJ. The statistics of the injected faults are reported in Table I for an increasing laser energy. Even at 5 nJ (which is twelve times the injection threshold) the rate of single-bit faults was still close to 30 %.

In-line shift register: our last tests on DFFs were carried out on the in-line shift register with the purpose to study the data-dependence of the LFI mechanism. We used 0.2 \( \mu \text{m} \) XY displacement steps to draw the laser-sensitivity maps of two consecutive DFFs as reported in Fig. 5 for a 0.5 nJ laser energy. The left (resp. right) part of Fig. 5 was obtained with the clock signal set at 0 (resp. set at 1), it displays the laser-sensitive areas of the slave latches (resp. of the master latches). Bit-set areas are highlighted in blue, bit-reset areas in red, bit-flip areas in purple. The laser-sensitive areas of the master latches are well defined, they almost correspond to the four bit-set/reset areas described in the theory (Fig. 1(b)) and found for the 40 nm DFF (Fig. 2). They only overlap in two points of the upper DFF (leading in these instances to bit-flips). Those of the slave latches (left part of Fig. 5) are less reconcilable

<table>
<thead>
<tr>
<th>Energy [nJ]</th>
<th>0.4</th>
<th>0.5</th>
<th>0.8</th>
<th>1.0</th>
<th>1.5</th>
<th>2.0</th>
<th>3.0</th>
<th>4.0</th>
<th>5.0</th>
</tr>
</thead>
<tbody>
<tr>
<td># of faults</td>
<td>1</td>
<td>8</td>
<td>21</td>
<td>23</td>
<td>24</td>
<td>24</td>
<td>26</td>
<td>30</td>
<td>31</td>
</tr>
<tr>
<td># of 1-bit faults</td>
<td>1</td>
<td>8</td>
<td>15</td>
<td>17</td>
<td>10</td>
<td>7</td>
<td>7</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td># of 2-bit faults</td>
<td>-</td>
<td>-</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td># of 3-bit faults</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>4</td>
<td>7</td>
<td>8</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td># of 4-bit faults</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td># of 5-bit faults</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td># of 6-bit faults</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td># of 7-bit faults</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td># of 8-bit faults</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table I: Number of laser-induced faulted bits in the DFF matrix at various laser energies (30 ps, 5 \( \mu \text{m} \) laser spot diameter).
with the theory. Although, they do not overlap (except for one point of the upper DFF). Moreover, during these series of tests only single-bit faults were obtained. 2-bit and 3-bit faults were effectively always observed only for DFFs placed side by side as in the case of the matrix-shaped shift register.

Conclusion on static results: the experiments reported in this section demonstrate on experimental grounds that single-bit laser fault injection is achievable into DFFs at the CMOS 28 nm technology node. This result was ascertained even when using a laser spot with a 5 µm diameter. In addition, very few bit-flips were obtained in the same DFFs when studying the data-dependence of the LFI process. It shows that the bit-set/reset FM is still achievable. However, these experiments were carried out in static mode, which is not representative of fault injection into a running IC. This is why we report dynamic testing in the next section.

IV. EXPERIMENTAL RESULTS: DYNAMIC LASER TESTING OF A CMOS 28 NM HARDWARE AES

A. Experimental Setup.

Fault injection setup: for the experiments reported in this section we used a 1,064 nm wavelength (NIR) laser source. The duration of the emitted laser pulses was set to 10 ns and the laser spot diameter set to 5 µm. This laser pulse duration was chosen equal to that of the target clock period (it also complements the picosecond range tests carries out on DFFs). The other parameters of the setup were left unchanged (backside laser injection, etc.).

Target: We targeted a hardware AES encryption unit embedded on the same CMOS 28 nm test chip. This AES features a counter-measure against fault injection based on parity tests. However, this CM description is out of the scope of this paper and this does not change the main properties of the LFI process. The core power of the device was 1.2 V.

B. Experimental Results.

The whole area of the hardware AES (about 200 µm × 130 µm) was scanned for fault injection with a XY displacement step of 1 µm. For each location, LFI experiments were carried out during the last three rounds of the AES while running at a 100 MHz frequency with laser power settings ranging from 0.6 W to 1.0 W. A unique pair of plaintext and encryption key was used for all these experiments. Since the plaintext and key were known and that we retrieved the obtained ciphertexts (faulted or not), we were able to identify when and where a fault was injected by comparing the obtained ciphertext with the correct AES encryption. Then, running backward the AES algorithm we tried to recover what faults were injected.

Fig. 6 displays the laser-sensitivity map of the AES we obtained: 26,380 fault were injected. Due to the parity-based counter-measure we were not able to identify with certainty all the corresponding faults at the moment of their injection. These 6,574 unidentified faults are displayed in red in Fig. 6 (they were mostly faults encompassing 5 to 8 bytes). The other 19,806 faults (marked in blue in Fig. 6) were single-byte faults. Their statistics are reported in table II and also as a laser-sensitivity map in Fig. 7 with a color code to show the exact number of their faulted bits. 19,413 of the laser-induced faults were single-bit faults.

Thus, the single-bit fault injection rate for our LFI experiments on the AES is 73.6 %. The single-byte fault injection rate is very close at 75 %. The rate of single-bit LFI is higher than what we were expecting given the 5 µm spot size we used at this advanced technology node. Moreover, the laser-sensitive areas depicted in blue both in Fig. 6 (single-bit faults) and in Fig. 7 (known faults, and also avoiding were it overlap with

1Note that our nanosecond range laser source is tuned using the laser pulse power expressed in watts, while our picosecond range laser source is tuned using the laser pulse energy expressed in nano joules.
the red areas of unidentified faults) are large and well defined. It makes it possible for an attacker to find several locations where fault injection will lead with certainty to the injection of single-bit faults. These results demonstrate that the single-bit FM is still relevant for LFI at the 28 nm CMOS technology node.

V. Conclusion

In this paper, we investigated on experimental basis two aspects of laser fault injection on the CMOS 28 nm technology node: (1) the ability of an attacker to inject single-bit faults (the most restrictive and effective fault model), and (2) the data-dependence of the injected faults that may provide an attacker with additional information on the data handled by its target. We carried out static experiments (i.e. with the clock signal stuck either at 0 or 1) on DFFs and dynamic experiments on a hardware implementation of the AES running at a 100 MHz frequency.

Experiments on DFFs proved that single-bit fault injection is still achievable in 28 nm CMOS ICs with laser spots diameters of 1 µm or even 5 µm. Using an accurate displacement step (0.2 µm) we were able to draw the laser-sensitive areas of the DFFs depending on the data they were holding (see Fig. 3). It shows that the data-dependent bit-set/reset fault model is also achievable (although bit-flip fault were also induced). In addition, the well defined laser-sensitive areas of Fig. 3 and the use of a 5 µm laser spot during others experiments leading to single-bit fault injection are a strong indication that the single-bit fault model shall be as well achievable for more integrated technology nodes (see also [17]).

Laser fault injection experiments on the hardware AES were performed dynamically with a 5 µm spot size and a 10 ns laser pulse duration. Despite this relatively large laser spot size and the use of a nanosecond range duration, which has a lower spatial accuracy than picosecond range duration [13], we obtained a single-bit fault injection rate of 73.3 %.

These results (though obtained from a given unique test chip) show that the single-bit fault model (and obviously the single-byte fault model) and the bit-set/reset fault model still hold for modern CMOS technologies: they are still actual and practical fault models that shall be considered when designing a secure IC.

Acknowledgment

This work was supported by a research grant from the French Agence Nationale de la Recherche (LIESSE project, ANR-12-INS-0008-01).

References