On the Complexity Reduction of Laser Fault Injection Campaigns using OBIC Measurements

Falk Schellenberg*, Markus Finkeldey†, Bastian Richter*, Maximilian Schäpers†, Nils Gerhardt†, Martin Hofmann† and Christof Paar*

*Horst Görtz Institute for IT-Security
†Photonics and Terahertz-Technology
Bochum, Germany
{firstname.secondname}@rub.de

Abstract—Laser Fault Injection (LFI) is one of the most powerful methods of inducing a fault as it allows targeting only specific areas down to single transistors. The downside compared to non-invasive methods like introducing clock glitches is the largely increased search space. An exhaustive search through all parameters including dimensions for correct timing, intensity, or length might not be not feasible. Existing solutions to this problem are either not directly applicable to the fault location or require additional device preparation and access to expensive equipment. Our method utilizes measuring the Optical Beam Induced Current (OBIC) as imaging technique to find target areas like flip-flops and thus, reducing the search space drastically. This measurement is possible with existing laser scanning microscopes or well-equipped LFI setups. We provide experimental results targeting the Advanced Encryption Standard (AES) hardware accelerator of an Atmel ATXMega microcontroller.

Keywords—fault injection, laser fault injection, optical beam induced current, laser scanning, AES, ATX Mega

I. INTRODUCTION

Since the first introduction of fault attacks in [1], there is a large research interest in both, finding mathematical attacks that exploit the faulty output of a cipher and finding physical methods to introduce actual faults on a target device [2]. Like most implementation attacks, fault attacks are independent of the mathematical security of the employed cryptographic algorithm. From an attacker’s point of view, these attacks intend to operate the Device Under Test (DUT) outside its specification. Such include the maximum clock frequency, the minimum or maximum operation voltage, the temperature, etc. For example, introducing short glitches into the clock signal might prevent the device of completing a full clock cycle, causing computational errors in the critical path. For Laser Fault Injection (LFI) (cf. [3]), the silicon die has to be decapsulated and is usually thinned from the backside. Then, a Near Infrared (NIR) laser beam is focused through the silicon substrate on the active region to induce a current caused by the photoelectric effect. Depending on the spot size and the feature size of the target, multiple or even single transistors can be targeted causing multiple or single bit errors on the device. Having this full spatial control over the fault location renders LFI one of the most powerful methods for fault injection. In contrast, non-invasive attacks like voltage or clock glitches usually affect the whole device. This advantage of LFI, however, raises some important practical difficulties as well.

Even non-invasive attacks have a very large parameter dimensionality. Parameters common to nearly all fault injection methods include the timing, i.e., a specific clock cycle or a specific point in time within a clock cycle, the duration of a certain effect, and its physical intensity. Only if all parameters are correct, a useful fault may occur. Depending on the search space of each parameter, an exhaustive search through all dimensions might not be feasible in an
acceptable amount of time. This problem is strongly increased in the case of laser fault injection: Besides the parameters mentioned above, spatial parameters are added by the x- and y-coordinate of the laser spot and, to some extent for an entirely unknown chip, the z-coordinate as well. Further, it is usually desired to choose a very small spot size for laser fault injection, i.e., to target single transistors for a deterministic fault behavior and/or to bypass local sensors. When no precise layout information about the target is available, a very small step size has thus to be chosen to cover the whole area in consequence of the small spot size. Finally, many commercially available microscopes designed for fault injection offer two independently movable laser spots. The goal is to be able to target two areas individually on the silicon die (cf. [4], [5]) to circumvent, e.g., duplication countermeasures. Here, the laser-specific parameter dimensions are doubled likewise.

We propose a method similar to [6], i.e., using clever imaging to find target locations for fault injection independently of all other parameters like timing, length, or energy. In contrast to [6], our method does not require additional device preparation or expensive equipment: By measuring the Optical Beam Induced Current (OBIC), a current generated at every pn-junction on the target device when illuminated with a laser (cf. Sect. II), we are able to use the setup for laser fault injection for device profiling as well. Like [6], we show that it is possible to find flip-flops by cross-correlation. Instead of an exhaustive search over the whole chip, the found positions can then be used for fault injection, saving a tremendous amount of time. We show the effectiveness of our approach by performing our experiments on an ATX Mega16A4U, targeting the hardware Advanced Encryption Standard (AES) (cf. [7]) accelerator, and successfully extract the secret key by means of Differential Fault Analysis (DFA).

Note that targeting flip-flops is sufficient at least for the often desired single bit faults. When inducing a fault into combinatorial logic, the fault is only effective when the faulty value is latched into a register. In contrast, when the register is not transparent at the time the faulty value is at its input (e.g., no rising edge on the clock input), the fault will not be sustained. However, targeting the memory cell of the flip-flop directly enables fault injection independent of the exact clock edge. Of course, not all possible fault patterns can be reproduced only targeting the flip-flops. For example, a fault in a combinatorial path might result in faulty values latched by more than one flip-flop.

Ignoring rare cases and analog designs, all chips nowadays are based on the standard cell design principle. Here, the manufacturer of the chip has a library of layouts for many basic gates, e.g., NAND, and more complex cells like flip-flops. Those are individually optimized and only copies are instantiated on the silicon die when needed and are connected on the upper layers. Consequentially, at least on the lower layers and especially the dopant/polysilicon layer, all instances of a cell look identical. Of course, one has to account for mirrored versions in both directions: (1) In a standard layout the lines for $V_{DD}$ and $V_{SS}$ alternate and the cells have to be mirrored, and (2), mirroring within the Complementary Metal Oxide Semiconductor (CMOS) lanes is possible as well without changing its functionality.

In the following, the z-axis always refers to the focal plane, i.e., the distance from the DUT to the microscope objective. The x/y-plane corresponds to locations on the DUT.

A. Related Work

Depending on the desired method of fault injection, an exhaustive search through all the parameters can be impractical. Consequentially, many recent publications deal with methods to reduce the search space of the parameters of fault injection campaigns. The approaches [8], [9], [10] employ machine learning algorithms to find optimal parameters for voltage level and glitch length when considering power glitches. The prerequisite for such methods is that the parameters are (steadily) continuous and bounded, e.g., depending on the voltage level of the power glitch, the device is either unaffected (lower bound) or stops responding at all (upper bound). It is sound to assume that the optimal value here is somewhere in between those bounds. However, there are many parameters for which this approach cannot succeed, e.g., the timing (or delay) of the fault or the position on the silicon die for a successful LFI.

Courbon et al. present in [6] a way to identify areas of interest for LFI: As an initial profiling
step, the device gets thinned down to the dopant layer from the frontside. Then, a high resolution image is taken using Scanning Electron Microscopy (SEM); after a single flip-flop is found, the characteristic pattern of its dopant layer is used to find all other instances of this flip-flop on the die. The found locations can then be used as target for LFI, drastically reducing the number of required shots compared to an exhaustive search. The downside is the requirement of an expensive SEM and an additional sample for profiling, which is destroyed afterwards.

In [11], Kizhvatov showed that the AES hardware acceleration used on an ATX Mega128A1 is vulnerable to side-channel analysis measuring the power consumption and using CPA as distinguisher. The attack requires less than 3000 traces. Our DUT, an ATX Mega16A4U, presumably implements an identical AES core and is vulnerable as well. Yet, we use this device to show the concept of our attack targeting an easy-accessible and well-documented hardware implementation of AES.

Measuring the OBIC, a well-known technique in semiconductor failure analysis (cf. [12], [13]), is only rarely mentioned within a security context: the authors of [14] present an OBIC image (with a very low resolution) for backside navigation, and no additional discussion or interpretation is provided. Sergei Skorobogatov describes in [15] that it is possible to read out masked ROM by measuring the OBIC of a nowadays rather outdated Motorola (now Freescale) MC68HC705P6A. In [16], this technique is used from the frontside of a Microchip PIC16F84 (0.9 µm technology node) to identify sensitive areas in SRAM cells and to deduce the transistor layout. OBIC is used as well in [17] to locate the sensitive areas of the Flash control logic of a NEC PD78F9116 microcontroller (0.35 µm technology node). The found areas where then targeted with LFI to extract the firmware by “bumping” during the verify-operation. The same attack was carried out against an Actel ProASIC3 A3P250 Field Programmable Gate Array (FPGA) as well (0.13 µm technology node). For the FPGA, however, the author used an exhaustive search over the whole chip area with a larger laser spot in a 20 µm grid to find areas sensitive to bumping attacks during the verify-operation. Note that even though the vulnerable areas were found quite efficiently, the latter approach still requires correct timing and could trigger potential reactive countermeasures.

B. Our Contribution

We present a method to reduce the location-dependent search space for laser fault injection, i.e., reducing the points of interest to shoot at. Our method measures the OBIC to create a high-resolution image from the backside of the chip. We show that it is possible to identify flip-flops similar to the method in [6], however, without the requirement of an SEM, simply by using an existing laser scanning microscope. The main benefits of our proposed method are:

- Finding the correct z-value for LFI and points of interest (x/y) for LFI independent of other parameters like laser energy or pulse width.
- The device is not powered. Consequentially, there is no clock signal and potential areas can be found independent of the correct timing. Further, potential reactive countermeasures against (laser) fault injection are shut off as well.
- Minimal equipment overhead: In order to measure the OBIC, a shunt resistor and a low-bandwidth Analog to Digital Converter (ADC) are fully sufficient. A more sophisticated but still low-cost setup can use a transimpedance amplifier.
- No additional preparation step is required: For LFI, backside thinning is usually done anyway to account for absorption of the laser beam in silicon.
- Measuring the OBIC provides a considerably better resolution even when compared to a laser scanning reflective image (cf. [18], [19]). Further, no additional (expensive) imaging equipment is required.

After identifying all flip-flops, we show that we can extract the secret key of the AES by means of classical DFA.

C. Organization

The rest of this paper is structured as follows. Section II provides an introduction into the physical effects behind OBIC. In Sect. III we describe the exemplarily chosen DUT, the used optical setups throughout this work, and equipment required for
measuring the OBIC. Section IV presents and discusses our proposed method. Finally, the conclusion and an outlook for future work are provided in Sect.VI.

II. PHYSICAL BACKGROUND

Induced faults in semiconductors were originally studied to understand the effects caused by cosmic radiation in high altitudes for aviation and space technologies. To simulate these effects which are usually caused by high energy particles, laser fault injection was used.

When a semiconductor is radiated by light, electron-hole pairs are generated along the path by the inner photoelectric effect. Normally, these pairs will recombine directly but if separated by an additional force, i.e. an electric field, a photo current can be measured. In an Integrated Circuit (IC), an electric field occurs at the junctions of differently doped areas. Roscian et al. model the induced current $I_{\text{laser}}$ in [20] with Equations 1-4.

$$I_{\text{laser}} = (a \cdot V + b) \cdot \Omega_{\text{laser}} \cdot S$$

$$a = p \cdot P_{\text{laser}}^2 + q \cdot P_{\text{laser}}$$

$$b = s \cdot P_{\text{laser}}$$

$$\Omega_{\text{laser}} = \beta \cdot e^{-\frac{d^2}{c_1}} + \gamma \cdot e^{-\frac{d^2}{c_2}}$$

$\Omega_{\text{laser}}$ is of special interest for OBIC since it describes the spatial dependency of the induced current. It forms the Gaussian distribution of the laser spot and shows an exponential dependency on the distance to the junction $d$. Also, the surface of the sensitive zone $S$ is an influencing factor. Since we measure the OBIC with the device shut off and no voltage bias, we have $V = 0$ and the current we observe does not depend on Equation 2. $\beta$, $\gamma$, $c_1$, and $c_2$ are fitting parameters which depend on the optical setup and lens used, $b$ is a function of the laser power $P_{\text{laser}}$ using the fitting parameter $s$.

While for OBIC we measure the induced current for imaging purposes, in a running device this current can change the state of the logic. When the laser spot hits the drain or gate of a transistor realized in CMOS, it can charge/discharge the node of the circuit it is connected to. This effect can be used to alter calculations by targeting flip-flops and alter their saved state. Also, it is possible to target the logic and hold the pulse until the next clock edge so that the changed logic state gets saved, since it will most likely recover into its original state when the laser pulse ends.

Note that besides OBIC, there is a multitude of (optical beam based) imaging techniques known in failure analysis (cf. [13]). We specifically chose OBIC measurements because of the ease to be integrated into a laser fault injection setup: Simply a shunt or a transimpedance amplifier and a device to measure the voltage is fully sufficient. In contrast, as described in [13], Light-Induced Voltage Alteration (LIVA) requires a constant current supply with nano ampere precision and range. Optical Beam Induced Resistance Change (OBIRCH), Thermally-Induced Voltage Alteration (TIVA), and Seebeck Effect Imaging (SEI) use localized heating caused by a laser with a wavelength above 1100 nm. Such a wavelength is usually not found in LFI setups (cf. [5], [14]) since here, the photoelectric effect is desired instead of the thermoelectric one.

III. SETUP

In the following, the DUT, the used microscopes and laser sources, as well as the measurement setup are described.

A. DEVICE UNDER TEST (DUT)

All our experiments were conducted using an Atmel ATX Mega16A4U in a TQFP44 package [21]. The technology node of the DUT is approximately 250 nm (based on the approximate minimum feature size in the SRAM, measured using a SEM solely for this purpose). The DUT includes 2 KB SRAM, 1KB EEPROM, 20KB of flash, up to 32MHz clock speed, and, most prominently from a security perspective, Data Encryption Standard (DES) and AES acceleration in hardware. The DES is attached to the Central Processing Unit (CPU) by a special instruction executing a single round of DES and its key schedule using the regular CPU registers. In contrast, the AES engine is attached to the data bus and provides the STATE, KEY, CTRL, and STATUS registers to control its function. It takes 375 clock cycles to complete one encryption. The
way the AES core is attached to the CPU favors it as target for a controlled experiment. For fault attacks on DES, one might not be able to distinguish faults on the actual DES hardware between general faults on the CPU, e.g., disrupting the instruction fetch. In contrast, when the AES is running on its own and the CPU is simply executing NOP operations, one can be certain that a fault in the final result, i.e., the ciphertext, was actually induced into the AES core.

The DUT was placed on a custom circuit board especially designed for laser fault injection and side-channel measurements: to recover from possible latch-ups caused by LFI, the power supply can be detached automatically via Universal Serial Bus (USB). To prevent the DUT from powering itself through its IO-ports, the programming pins and a serial interface are connected through an electrical isolation IC.

To reduce absorption of the laser beam in silicon, the package was opened from the backside and the exposed silicon was thinned down close to the active layer and polished. The DUT used in our experiments has a silicon substrate thickness of around 20 \( \mu \)m (i.e., the distance from the backside surface to the doped layer). This is a typical approach for laser fault injection.

B. Experimental Setup

Two different optical setups were used throughout the experiments: Setup 1 for creating the OBIC image and Setup 2 for fault injection\(^2\). The main difference is that Setup 1 is equipped with a continuous scanning stage with position output and thus, was used for faster imaging. A common coordinate system for both setups was established by measuring different reference points on the DUT using OBIC measurement. Both setups stand in an air-conditioned class 4 laser laboratory on a vibration isolated optical table.

1) Setup 1 (Laser Scanning Microscope): Setup 1 is a modified self-built laser scanning confocal microscope [23], [24], which is used for OBIC and reflective mode imaging with high resolution and precision. A temperature stabilized, fiber coupled laser diode module from Lumics at 1064 nm was used as light source. The Single Mode Fiber (SMF) output was collimated by a large beam reflective collimator and propagated through a non-polarizing beam splitter (BS). The light was focused onto the DUT through a custom made Leica objective (Numerical Aperture (NA) 0.75, magnification 100x), optimized for use in the NIR range. The diffraction limited spot size for the given combination of objective and laser was calculated to be 1.7 \( \mu \)m. The DUT was scanned by three motorized stages (x, y: LTA-HS with M-462; z: LTA-HL with M-MVN80) from Newport Corp. with a measured precision better than 100 nm. For reflection mode imaging an additional achromatic lens (\( f=35 \) mm) was used to focus the backscattered light from the sample through a custom 500 \( \mu \)m aperture onto a Si photodiode. The OBIC was amplified and converted to a voltage using a FEMTO DLPCA-200 variable gain transimpedance amplifier and a Stanford Research SR560 low noise amplifier. The low-pass filter of the SR560 was set to 10 kHz to match the maximum ADC bandwidth in the Newport XPS motion controller unit. A PC with Matlab was used to communicate over Ethernet with the XPS controller for operation, data collection, and creating the point cloud. Figure 1 shows the electrical part of the setup and Fig.2 shows the optical part.

2) Setup 2 (Modified Laser-Fault-Injection Microscope from Opto GmbH): Setup 2 is a commercially available microscope from Opto GmbH especially designed for dual spot laser fault injection (cf. Fig.3). It was modified to reach the required mechanical precision and stability by changing the z-stage to a precision z-stage (M-501.1DG) from Physik Instrumente (PI). In addition, a new sample holder was designed and integrated. A 976 nm

\(^1\) Note that this separation has solely organizational reasons during our experiments as the main difference, the scanning stage, can be easily implemented in both setups.
SMF-coupled on-demand diode laser module from ALPhANOV [25] was used as light source, focused through a Mitutoyo Plan Apo NIR HR objective (NA 0.65, magnification 50x). We calculated the diffraction limited spot size to 1.83µm.

IV. Results

In the following, a fault injection campaign on the hardware AES implementation of an ATXmega16A4U using our proposed method of reducing the search space is described. We start by using the OBIC measurements to identify flip-flops as candidates. Then, we target the found positions and show that it is possible to extract the secret AES key using LFI. Although not within the scope of this work, we utilized side-channel measurements two times during our experiments: first, to obtain a rough estimation where the AES might be located based on its electromagnetic (EM) emanation, and second, to pinpoint a timespan for fault injection based on Correlation Power Analysis (CPA) with a known key. However, when the attacker has no such control over the target device, these measurements can be skipped.

A. Estimating the Rough Location of AES

For initially reducing the search space, we measured the localized EM emanation of the DUT to obtain a rough location of the AES core. Note that this approach might not work for every target, as it heavily depends on the characteristic emanation and implementation. We used an open-source probe available at [26]. During the computation of the AES core a trigger signal was pulled high and the probe was moved across the surface manually. While the core was running, the CPU simply executed NOP-operations. Figure 4 depicts the captured emanation at one side of the DUT, showing a strong signal not found elsewhere. A rectangle within this area was used for measuring the OBIC.

B. OBIC Measurements

In the following, we describe how OBIC measurement can be used to determine the correct z-position and how to find candidates (x/y) for laser fault injection. Note that this profiling step has to be done only once per target. The chip is not powered and thus, those steps are independent of other parameters like correct timing, pulse length, or (to some extent) pulse energy. Further, no responsive countermeasure against (laser) fault injection is active.
1) Establishing the correct z-position: The first step is to find the correct z-position for the following measurements of the OBIC on the x/y-plane. We sweep through the z-axis while holding the laser energy constant. The measurable current caused by the laser will get maximized exactly when the photon density on the sensitive area is maximized, i.e., the focus of the laser spot lays directly at the pn-junction. Consequently, the diameter of the laser spot is minimal as well and this z-value can be used for imaging, provided the optimal resolution. Figure 5 depicts the measured OBIC signal while sweeping through the z-axis. The found position can be used for fault injection as well. Repeating this process at multiple x/y-positions, e.g., at each border, further enables a very precise measurement of the angle of the DUT to the objective. In case the chip is slightly tilted, the z-axis can be adjusted automatically depending on the current x/y-position for targeting a very large area, both for OBIC and for fault injection.

2) x/y: Once the correct z-position is found, the OBIC can be measured in xy-direction, providing a detailed image of the active area of the DUT. Figure 6 depicts an OBIC measurement over an area of around 225 \( \mu \text{m} \times 150 \mu \text{m} \) with a step size of 100 nm taken in approximately 45 min. The location was roughly estimated using EM measurements as described above. One can clearly identify the 18 vertical lanes implementing CMOS-logic and the supply lines in between. In more detail, Fig. 7 has a width of two full “CMOS”-lanes, each having a width of around 12 \( \mu \text{m} \). The horizontal bars visible in the image directly correspond to the drain and source regions of the individual transistors. Given that PMOS transistors usually have larger channel widths than NMOS transistors in CMOS designs\(^2\), one can assume that the inner supply lane corresponds to \( V_{\text{DD}} \) and the two lines at the outside are connected to \( V_{\text{SS}} \). Most importantly, one can easily identify repeated cell instances at the provided resolution. For example in Fig. 7, there are four very large identical standard cells visible, two of which are mirrored vertically. A quick manual inspection revealed that the marked pattern is repeated multiple times within the area and we assumed that the pattern corresponds to a flip-flop. Flip-flops are usually one of the largest cells in a standard cell library. We argue that even without the exact knowledge of the layout of a flip-flop, we still can identify large repeated cells. Even if we obtain multiple potential candidate layouts, we do not care about a slight overhead caused by some false positives. Still, the time benefit will be very large compared to an exhaustive search over the whole area.

To verify whether the pattern in our case implements a flip-flop, we performed a fault injection test on one of its occurrences: During the laser pulse we stopped the clock of the ATXmega. Only faults which directly affect flip-flops will persist and produce a fault in the calculation. Faults in

\(^2\)The electrons forming the channel for n-type MOSFETs have a higher mobility than the holes for p-type MOSFETs. For a symmetric switching voltage, the widths are usually chosen so that the width of the p-type is 2.5 times larger than the width of the n-type MOSFET [27].
combinatorial logic will not be saved to the flip-flops due to the missing clock. Since we were indeed able to induce faults using this procedure, we are certain that the used pattern corresponds to flip-flops. Thus, we used this pattern for automatically finding other cell instances as described in the following. Note that devices specifically designed for security-critical applications will not provide external access to the clock signal. Thus, halting the clock is not possible there.

Figure 6: OBIC x/y, the Gray scale corresponds to the OBIC amplitude in arbitrary units. We assume that the wave around $Y = 100\,\mu m$ is a result of mechanical stress caused by the thinning of the silicon die.

C. Correlation-Based Pattern Recognition

Using the large characteristic pattern assumed to implement a flip-flop, we searched for all other instances in the area visible in Fig. 6 using correlation. As discussed above, the instantiated standard cells might be mirrored because the vertical lines connecting to $V_{DD}$ and $V_{SS}$ alternate. In order to optimize routing, the instances further might be horizontally mirrored, i.e., so that the output of the cell is closer to the input of the following cell. Figure 8 depicts the two-dimensional correlation between the pattern (without its mirrored version) and the selected area. This resulted in multiple spikes with a correlation ranging from 0.6 up to 0.8.

Note that for finding flip-flops based on correlation in a larger area, one might want to limit the search space to the individual lanes between the supply lines. We skipped this step because the computation in our example finished in a matter of seconds on a standard PC. Figure 9 depicts 34 found positions for all four mirrored versions of the pattern. In order to obtain the expected at least 128 positions, the search area has to be increased further.
The colors (or line types) in the figure differentiate each mirrored version. Note that the distribution is sound with respect to the vertical lanes: Only red and yellow boxes or blue and green boxes appear in one lane. Further, both types alternate caused by the alternating $V_{DD}$ and $V_{SS}$ connections. Now, instead of targeting the whole area with LFI, we can only focus on the found locations, reducing the search space. Usually, not the whole area within the box will have sensitivity towards fault injection. Instead, specific areas or transistors of the cell will have an effect while others not. Thus, after finding those sensitivity zones for one pattern, the search space can be decreased even further only targeting the correct zones with a few laser shots.

![Figure 9: Found flip-flops marked in OBIC image. Colors represent mirrored versions.](image)

**D. Finding the Correct Timing**

Before performing fault attacks on the isolated locations, we performed a CPA attack with a known key to narrow down the point in time. If no profiling device with known key is available, this step has to be skipped. For profiling, we chose the last round to avoid the MixColumns step of AES. Thus, for a successful fault injection, we expect either single-bit faults (fault occurred after the SBox) or single byte faults (fault before SBox). We started by calculating intermediate values at the beginning of the last round, i.e., after the addition of the round key but before the S-Box computation in ShiftRows order. The most suitable power model was found to be the Hamming Distance between each byte of the state: $HD(s_i, s_{i+1})$ for state byte $s_i$. Figure 10 shows the correlation with 15 clearly visible peaks corresponding to the 15 Hamming Distances using 300 traces. Therefore, at exactly this point in time the value is processed or overwritten respectively, providing an exact time reference.

![Figure 10: Correlation for Hamming Distance between consecutive state bytes at the input of the last round after KeyAddition, the red lines represent the $\pm \frac{4}{\sqrt{\#traces}}$ bound for uncorrelated noise (cf. [28, p.150]).](image)

**E. Targeting found Locations**

Having found potential areas and the potential point in time; we performed actual fault attacks on the last round of AES. Note that the profiling steps above only have to be done once per target. We chose to energize the laser at exactly 181 µs for a duration of 600 ns, i.e., shortly before the rising edge of a clock cycle until shortly after the rising edge of the following cycle (cf. Fig. 10). This way, a full clock cycle is covered. The point in time is specifically chosen to affect after and before an addition with the secret key. To ensure an even distribution at the input of the last round, we tested each position for 256 different plaintexts. Figure 11 depicts an overlay of the found sensitivity zones within the previously found locations: Green dots correspond to Bit-Set and red dots to Bit-Reset faults. Note that not all found cells show sensitivity, presumably because their content gets overwritten later or the registers are not active.

Figure 12 provides a more detailed view of four cells and their sensitivity zones for LFI. Given that both red boxes correspond to an identical orientation

---

3Note that the previous attack in [11] required 3000 traces using the HD model after key addition in the first round.
of the instantiated standard cell, the obtained sensitivity zones are very similar as well. Further, since the blue boxes are a horizontally mirrored version of the red ones, the sensitivity zones are mirrored as well. Consistent to the expected behavior as discussed in Sect.II, the faults occur where the OBIC measurement shows high amplitude. However, investigating the sensitive zones differentiated by their exact behavior, i.e., the green and the red areas, reveals interesting effects. There is always one complementary pattern visible, cf. Fig.12 at each box in the upper part. The third sensitivity zone, however, changes its fault behavior: In the upper two boxes of Fig.12, the third zone at the bottom triggers Bit-resets. In contrast, the zones in the lower two boxes of the same figure correspond to Bit-Sets, although the other zones are unchanged. We assume that the complementary zones implement the storage part of the flip-flop. Hence, we obtain specific areas for Set and Reset, similar as reported in [29]. Because the third sensitivity zone appears uncorrelated to the other two in each cell, we assume that it is related to the input signal or the clock signal. Note that the orientation of the upper complementary zones is inverted for some instances as well. Since we are only able to determine the behavior based on the ciphertext and have no additional information about the implementation, we assume that those flip-flops store their content inverted.

Figure 11: Locations found sensitive to LFI within the previously isolated areas. Green dots correspond to Bit-Set and Red dots to Bit-Reset faults.

Figure 12: Locations found sensitive to LFI within the previously isolated areas, detail of four cells.

F. Differential Fault Attack (DFA)

Comparing the genuine ciphertexts and the faulted ones, we noticed that we only obtained single-bit faults. The last operation in AES is an exclusive-or between the current state and the last round key. Thus, for an unknown key, we cannot determine whether the ciphertext was altered or the key. Thus, we changed the timing so that the laser is energized exactly one round earlier, i.e., directly within the ninth round of AES. This way, we received identical sensitivity zones but this time resulting in single-byte and four-byte faults in the ciphertext. Concurrent with the considerations above (cf.Sect.IV-D), this corresponds to some faults occurring before the MixColumns-Step and some faults afterwards. Picking only the single-byte faults, we performed a well-known, straight-forward DFA [30]: For given pairs of ciphertext and faulty ciphertext, we test whether the difference between hypothetical state values at the input of the SBox in the last round resolves to a single-bit fault (at an identical position for all pairs). The key hypotheses that satisfy this equation are potential key candidates. As mentioned in [30], usually only a few key candi-
dates remain for one pair of ciphertext and faulty ciphertext. Indeed, we were able to calculate a byte of the correct round key using two respective pairs.

V. Discussion

A. Reduction of Points of Interest

The time required to perform a fault injection campaign linearly depends on the number of positions to test: The setup first has to move to these positions and then, can perform a test with constant runtime. The time required to move to a certain position depends solely on the way the setup is constructed (moving stage or scanning within the field of view of the objective). Further, the runtime per position heavily depends on the execution time on the target and the communication overhead. Thus, we only consider the total number of positions in our calculation. For the given example, $255 \cdot 150 = 38250$ positions need to be tested if scanning the whole range in $1 \mu m$ steps. Using our method, the number of positions can be reduced to $34 \cdot 17 \cdot 10 = 5780$ positions within the borders of the 34 found flip-flops. This is an improvement by a factor of 6.6 which is a considerable improvement given the high density of flip-flops in the targeted area. This number can be reduced further if the sensitivity zones are known. For example, one could theoretically target each of the sensitivity zones only once with a slightly larger laser spot. Then, the number of positions is reduced to $34 \cdot 3 = 102$ or by a factor of 375 for the given area and 34 found instances.

B. OBIC versus Reflective Imaging

Based on the provided figures, one might be tempted to think that there is no significant difference between OBIC measurements and capturing the NIR reflection using a camera for the NIR range. In order to compare both approaches, we captured the reflection of the laser beam using the laser scanning microscope (cf. Sect. III.B). Note that laser scanning microscopy with reflection measurement already provides a higher resolution compared to using a camera (cf. [31], [32]). Figure 13 depicts a direct comparison between OBIC and reflective measurements. As discussed in [18], [19] as well, measuring the OBIC provides more detail and improved contrast compared to laser scanning. Anyhow, OBIC measurement certainly enables a higher resolution than commercially available NIR-lightning/camera solutions for fault injection setups, e.g., shown in [33] for an ATMEGA163.

![Figure 13: Comparison between OBIC and reflective measurements. Jet color scale corresponds to arbitrary units of the respective signal.](image)

C. Influence of the Technology Node

To decide, whether the described method could be applied, one has to consider the following parameters:

- Technology node,
- characteristic cell layout of the DUT, and
- the (effective) spot size of the laser.

Given the SMF-coupling of the used laser diodes, the photon (or energy) density of the laser spot will follow a Gaussian shape. The spatial resolution $d$, as defined by the Rayleigh criterion, is given by $d = \frac{\lambda}{2NA}$. In case of the Setup 1 (LSM) these leads to $d \approx 710 \mu m$.

Comparing this value to the measured technology node of the ATMEGA16A4U of 250nm might seem conflicting. However, the technology node refers to the minimal feature size incorporated into the silicon die, which is usually the channel length of the smallest n-type MOSFET. In the other direction, the channel width is linked to the current handling capability of the transistor and is in CMOS designs usually much larger than the channel length. For example, the ratio between channel width and length of a NAND-gate of an ATMEGA32 is approximately $\frac{W}{L} = 8$ (cf. Fig. 14). Including the drain and source regions, connecting vias, and signals to the layout thus results in a characteristic structure much larger than the technology node. This especially holds for flip-flops consisting of a multitude of transistors. In fact, the found flip-flops in Sect.IV-B2 require a space of around $17 \mu m$ times $12 \mu m$, providing plenty of structural detail for the given resolution.
VI. Conclusion and Future Work

We proposed a method to drastically reduce the location-dependent search space for laser fault injection with minimal equipment overhead. Target locations for LFI can be found while the device is not running, and thus, independently of the correct timing, pulse length, etc. Further, potential reactive countermeasures are shut off as well during profiling. Similar to [6], we identified flip-flops using correlation, but instead of using a SEM and additional device preparation, we show that measuring the OBIC offers enough resolution (for the given DUT). Even if the exact layout of a flip-flop is not known, we discuss that once a candidate is manually found to be repeated multiple times, we can verify its behavior by inducing a fault while the clock is not running. Using our method, i.e., measuring the OBIC for precise imaging, we successfully attacked the AES hardware implementation of an ATXmega16A4U and extracted the secret key. Even though our exemplary chosen target area contains a large amount of flip-flops, having this precise location information results in our case to a reduction factor of 6.6 when covering the full flip-flop area, or a factor of 375 when targeting each sensitivity zone with a single shot. It can be assumed that for a less dense distribution of flips-flops, the improvement factor will be much higher.

For future work, it should be answered down to which minimum feature size measuring the OBIC is providing enough resolution to find repeated patterns based on correlation. Based on our measurements, the feature size of the ATXMega16A4U is certainly not the border. As long as single bit faults in very dense structures like SRAM can be induced, OBIC measurements should be possible likewise. Considering countermeasures, it should be investigated whether it is possible to isolate the power supply of the DUT in a way that the OBIC cannot be measured. However, then it might still be possible to probe the silicon substrate directly.

Acknowledgment

The authors would like to thank Enrico Dietz and Sven Frohmann from Technische Universität Berlin for providing multiple thinned ATXmega microcontrollers and Christian Kison from RUB/BKA for providing the SEM images.

References


