On the Use of Forward Body Biasing to Decrease the Repeatability of Laser-Induced Faults

Marc Lacruche, Noemie Beringuier-Boher, Jean-Max Dutertre, Jean-Baptiste Rigaud, Edith Kussener

On the Use of Forward Body Biasing to Decrease the Repeatability of Laser-Induced Faults

Marc Lacruche*, Noemie Beringuier-Boher*, Jean-Max Dutertre*, Jean-Baptiste Rigaud*, Edith Kussener†
*École Nationale Supérieure des Mines de Saint-Étienne (email: firstname.name@emse.fr),
†IM2NP (email: firstname.name@im2np.fr)

***Abstract*—This paper presents a study on the effect of Forward Body Biasing on the laser fault sensitivity of a CMOS 90 nm microcontroller. Tests were performed on a register of this target, under several supply voltage and body bias settings, showing significant laser sensitivity variations. Based on these results, a method which aims at decreasing fault repeatability by using variable supply voltage and body bias settings is proposed. Finally, tests are performed on an implementation of this method on a temporally redundant AES and the results are presented.**

I. Introduction

With the constant push for more power-efficient circuits, it is essential to make sure that low-power design techniques do not compromise the secure elements of the devices they are used in. Fault attacks are a particular concern as low-power devices are often vastly deployed, giving easy physical access to the attacker.

Fault attacks aim to disturb a device and exploit the resulting computation errors to recover secure informations. Existing attacks include the Differential Fault Analysis (DFA) [1] that uses the differences between a faulty cipher and the correct one to recover the key of an algorithm such as the AES [2].

Many methods exist to perform fault attacks, including voltage or clock glitches [3], electro-magnetic pulses [4] and more recently body biasing injection [5]. Laser fault injection is one of these methods, it was introduced in the field of hardware security by S. Skorobogatov and R. Anderson in [6]. It uses the photoelectric effect to induce currents locally in a circuit to disturb its operation. [7]

This paper takes a look at the effect of Forward Body Biasing (FBB) [8] as a low-power design method on the vulnerability of a circuit to laser fault injection: first, FBB is briefly introduced. The next part presents the results of experiments where laser fault injection was performed on a register of a microcontroller embedding FBB capabilities. Then, a method based on the previous results is proposed, aiming at decreasing fault repeatability by using the sensitivity variations induced by the use of FBB and modified supply voltage values. Finally, in the last section, this method is applied to a temporally redundant hardware AES as a proof of concept and test results are presented.

II. Triple Well and Forward Body Biasing

In standard bulk CMOS architectures, NMOS transistors are located in the P-Substrate, and PMOS transistors are located in a N-Well itself located in the P-Substrate. Commonly, the P-Substrate is biased at ground voltage while the N-Well is biased at V_{dd}.

In a triple-well architecture (figure 1), NMOS may be located in a P-Well isolated from the P-Substrate by a deep N-Well layer underneath and an additional N-Well on the side.

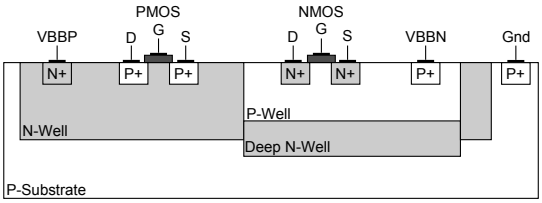

Fig. 1: Cross-sectional view of CMOS Triple-Well process

Having an isolated P-Well for NMOS transistors allows the use of body biasing on both NMOS and PMOS (as opposed to only the PMOS when using bulk technology). The NMOS body bias voltage is set at $V_{bbN} = V_{bb}$ and the PMOS body bias voltage at $V_{bbP} = V_{dd} - V_{bb}$. Using a positive V_{bb} value, is called Forward Body Biasing (FBB) while using a negative V_{bb} value is referred at as Reverse Body Biasing (RBB). Using FBB or RBB allows to adjust the leakage and frequency of the transistors by acting on their threshold voltage (V_T) [9]. FBB allows to increase frequency at the expense of leakage while RBB has the opposite effect.

This paper is focused on the use of FBB as a low-power design method. Indeed, FBB can be used to reduce power consumption at equivalent timing performance by lowering supply voltage and using FBB to offset the performance hit.

III. Preliminary Evaluations

The first evaluation step was to test whether V_{dd} variations and FBB had an effect on the sensitivity of a chip to laser fault injection. To do so, a 32-bit register located on a CMOS 90 nm microcontroller with V_{dd} scaling and FBB capabilities was targeted. Possible core V_{dd} values were 1.13 V, 1.26 V and 1.32 V and FBB could be either set to FBB ON (V_{bb}=400 mV) or FBB OFF (V_{bb}=0 mV).

A. Test Setup

The laser test bench used has its objective mounted on a motorized XYZ table, allowing the automation of the mapping process. The register covers a $70\,\mu m$ by $80\,\mu m$ area out of which a $25\,\mu m$ by $25\,\mu m$ square in the middle of it was targeted (due to mapping time constraints) and mapped with a $0.5\,\mu m$ step grid. For each position fault injection was

performed while using every possible V_{dd}/FBB combination and the register set to 0xFFFFFFFF and 0x00000000.

The injection was done through the backside of the chip using a 1064 nm wavelength laser (IR). Laser pulse duration was 50 ns and the size of the spot was 1 µm. The targeted chip was thinned to a substrate thickness of 140 µm.

B. Results

1) Laser-Sensitive Area: The first tested metric was the evolution of the laser-sensitive area depending on the V_{dd}/FBB configuration for a fixed laser power of 0.6 W. This power is just under the cell destruction threshold.

Since the mapped area has the same fixed number of points for every configuration, the evolution of the laser-sensitive area can be measured by looking at the number of faults that were recorded on the map for a given V_{dd}/FBB configuration. Figure 2 summarizes the sensitive area for each V_{dd}/FBB pair.

The first trend that we can extract from the histogram is that lowering V_{dd} increases the sensitive area and turning FBB ON results in another sharp increase in sensitivity. Overall, the fault count for the most sensitive configuration (V_{dd}=1.13 V and FBB ON) is about twice as important as the sensitive area of the least sensitive configuration (V_{dd}=1.32 V and FBB OFF).

2) Laser Power Influence: The second step was repeating the experiment with several laser power values. The goal was to find the laser power threshold from which faults start to appear as well as the rate at which sensitivity increases depending on power for each V_{bb}/FBB pair.

For each position of the map, tests were performed with laser power varying from 0.25 W to 0.40 W with a step of 0.02 W. Table I displays the obtained laser power thresholds. Similarly to the previous results, a lower V_{dd} is linked to a higher laser-sensitivity (ie. a lower fault injection threshold) and activating FBB further increases the sensitivity. The same trend can be observed on the graph showing the fault count versus the laser power (figure 3).

V_{dd}	FBB OFF	FBB ON (400mV)
1.13 V	0.34 W	0.30 W
1.26 V	0.38 W	0.32 W
1.32 V	0.38 W	0.34 W

TABLE I: Laser power fault injection threshold

C. Conclusion

While they will obviously vary depending on the technology used, these results give a good overview of the trends that can be expected when applying a reduced V_{dd} and using FBB. Lowering V_{dd} increases the laser sensitivity, and turning FBB ON introduces another sharp increase in sensitivity.

However, these results do not introduce any new major security vulnerability as the laser power threshold variations are still in the power range of common laser power sources. The main concern is that the sensitivity increase diminishes the efficiency of sensors-based countermeasures as lowering the fault threshold can push it under the detection threshold of the sensors.

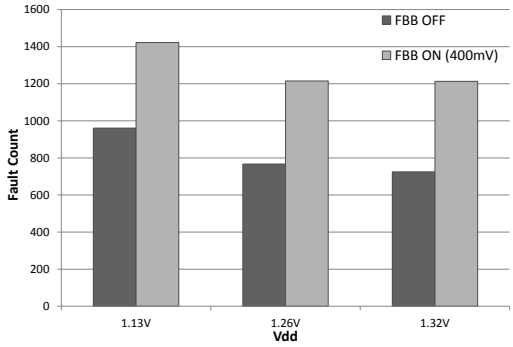

Fig. 2: Total faults per map for each tested power configuration

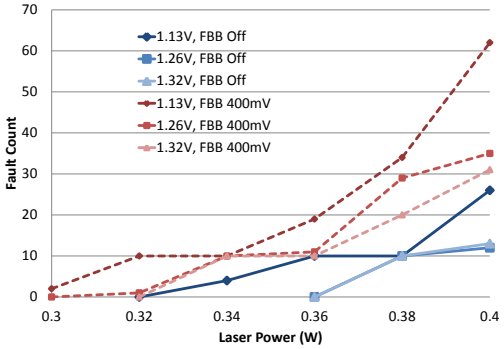

Fig. 3: Faults for different laser power values

IV. Exploiting Sensitivity Variations

When performing fault-attacks, fault repeatability and predictability are important factors for the success of many attacks such as the ones described in [10]. They also help breaking redundancy-based countermeasures. As such, it is the defender's best interest to minimize those parameters.

Although the previous results should have a limited impact on device vulnerability, the following section is going to show how the varying sensitivity when using different V_{dd}/FBB configurations can be used to harden a circuit against laser fault injection by reducing fault repeatability.

A. Principle

When performing laser fault injection with increasing laser power, it is possible to distinguish different laser power ranges for which different fault types are injected. For very low laser power values, no faults will be induced, then when attaining the fault threshold, limited faults will appear (single bit/byte faults). Further increasing the power will eventually induce multi bit/byte faults, and finally a threshold leading to permanent chip damage will be attained.

As shown in section III-B2, it is possible to use V_{dd}/FBB to modify the threshold at which faults appear and increase or decrease the laser sensitivity of a circuit.

Figure 4 depicts how the sensitivity of the cell will vary when using different V_{dd}/FBB settings, effectively moving the laser power thresholds up or down the laser power scale. Using this, it is possible to decrease the repeatability and predictability of laser induced faults by using varying V_{dd}/FBB settings over time.

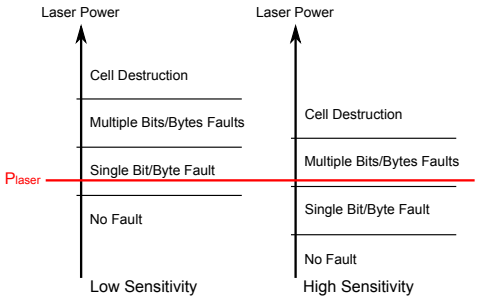

Fig. 4: Laser power fault thresholds variations

B. Proof of Concept Implementation

As a proof of concept, we applied this technique to temporal redundancy which consists in computing two (or more) identical encryptions one after the other using the same hardware and comparing the results. If the outputs of both encryptions are identical the result is returned, otherwise the error is detected and an action can be taken (block the output, return a random cipher, raise an alarm etc.). This means that an attacker would have to inject two identical faults consecutively to retrieve a faulty cipher, rendering fault attacks much more complex to implement. It also has the advantage of being simple to implement and is a quite immediate application of the proposed method. To do so, the same microcontroller as in the previous tests is used as it includes a hardware crypto-accelerator that can perform AES calculations.

Two power supply configurations are defined: a low-sensitivity configuration using V_{dd}=1.32 V and FBB OFF and a high-sensitivity configuration using V_{dd}=1.13 V and FBB ON. The first calculation of the redundancy is performed using the low-sensitivity configuration and the second one using the high-sensitivity configuration. The assumption is that two different faults will be induced, leading to the detection of the attack, even if the attacker is able to perform two consecutive laser shots at the exact same instant of the algorithm.

C. Test Setup and Procedure

Using the same laser test bench as previously described, the test is set up as follows:

- The map covers a region of 200 µm x 300 µm inside the hardware AES of the microntroller with a step of 5 µm.
- In order to speed up the mapping, one encryption is performed on each position, if a fault is injected, the mapping stops and performs 2000 encryptions on the current position (1000 times one low-sensitivity encryption followed by one high-sensitivity encryption) before continuing with the rest of the map.
- The crypto-accelerator clock runs at 50MHz.
- The microcontroller sends a trigger signal to the laser before each encryption, targeting the second to last round of the AES (the laser shoots after a fixed delay after receiving the trigger signal).
- Laser power is fixed at 0.7W and pulse duration at 50 ns (the minimum pulse duration of the laser source).

As described in figure 5, the results of consecutive encryptions are compared in order to obtain what would have been the output of different redundancy types: a hardened redundancy (one computation with each sensitivity setting), a simple redundancy with both computations in the high-sensitivity mode and a simple redundancy with both computations in the low-sensitivity mode.

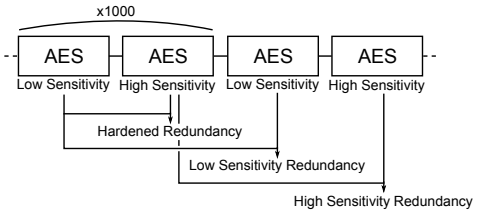

Fig. 5: Testing Procedure

D. Experimental Results

1) Overall Results: When looking at the results, three cases are distinguished:

- Double fault: a successful injection of the same fault on two consecutive AES computations. In which case the redundancy is broken and the faulty cipher is returned.
- Two different outputs: either two different faults or only one of the two computation is faulted. Here, the AES is safe thanks to the redundancy and the attack is detected.
- No fault: no faults were injected on either computations. Here, the system is safe but the attack is not detected.

Output Type	Hardened	High-sensitivity	Low-sensitivity
Double fault	35.6%	74.3%	51.7%
Different outputs	58.6%	13.6%	18.9%
No fault	5.8%	12.1%	29.4%

TABLE II: Occurrence rate of the different cases

Table II reports the results obtained over all the tested positions. When looking at double fault-rate, the hardened redundancy is more resistant to double faults than the other two configurations. The difference is especially important compared to the high-sensitivity configuration which is the FBB use-case that was considered early on in this paper. The hardened redundancy also has the highest rate of two different output results by a large margin which makes it the best at both preventing double faults and detecting unsuccessful attacks.

While the results of the two first rows from the table were foreseeable, the last one is a bit unsuspected. The low-sensitivity mode has the highest no-fault rate as expected, but the hardened redundancy is lower than the high-sensitivity mode, which means that there are a non-negligible number of cases where a fault appears on the low-sensitivity configuration and not on the high-sensitivity configuration.

Overall the proposed method looks quite effective at both limiting the ease of creating a double fault and increasing the ability to detect the attack.

2) Result Maps: We drew maps out of the results, but no particular spatial tendencies were observed. All the redundancies performed equally over the tested surface. The only interesting information was that for a few number of points faults would appear in low-sensitivity mode and not in high-sensitivity mode.

3) Positional Comparisons: The last evaluation parameter is the effectiveness of the counter-measure on a position by position basis. Figure 6 shows the comparison location by location between the hardened redundancy and the high-sensitivity redundancy. Each point represents a tested position from which the x coordinate is the hardened redundancy double fault rate and the y coordinate is the high-sensitivity redundancy double fault rate. For example, a point located in the top left corner of the chart represents a tested location for which the double fault rate was 100% using the high-sensitivity redundancy and 0% using the hardened redundancy (ie. the ideal case). Points on the diagonal are positions for which both redundancies are as effective as each other. Hence, points above the diagonal are the positions for which the hardened redundancy performs better than the high-sensitivity one and the points below are the positions for wich the hardened redundancy is worse.

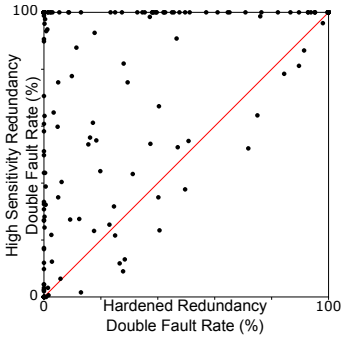

Fig. 6: Double fault sucess rates (hardened vs. high-sensitivity)

Figure 7 is the same graph but shows the detection rate instead. In this case, a higher rate is desired as it represents the ability to detect an unsuccessful attack.

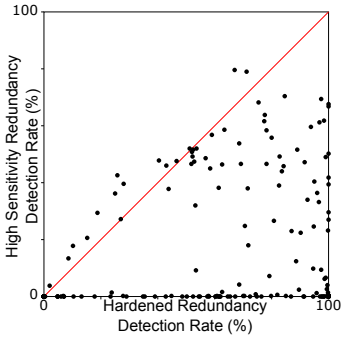

Fig. 7: Detection rates (hardened vs. high-sensitivity)

These two graphs show that there are actually a few points for which the hardened redundancy performs worse than the high-sensitivity redundancy by a small margin, which was not evident by looking at the results so far. Still the overall results are largely in favor of the hardened redundancy.

V. Conclusion

This paper started by providing an evaluation of the effects of FBB on the sensitivity of a register to laser fault injection showing that FBB and a lower V_{dd} value increased the sensitivity of the register. Based on these results a method using sensitivity variations to decrease fault repeatability by using variable V_{dd}/FBB configurations was proposed, implemented and tested. The obtained results showed that the method was effective both at decreasing double-fault injection rate as well as increasing attack detection rate.

All in all, these results show that low-power technologies can induce security risks if not carefully accounted for. But when planned for they can be used to enhance security.

We insist on the fact that temporal redundancy is only one application that we used as a proof of concept. However, this is something that is easily implemented on circuits that already FBB capabilities available and further increases the complexity of the attacker's work. Other applications of the method will be evaluated in the future.

Further works will also include evaluating the use of Reverse Body Biasing in order to increase the sensitivity range, as well as investigating the physical phenomenons and trying to understand the cases where faults appear in low sensitivity mode and not in high sensitivity mode.

References

[1] G. Piret and J.-J. Quisquater, "A differential fault attack technique against spn structures, with application to the aes and khazad," in Cryptographic Hardware and Embedded Systems-CHES 2003. Springer, 2003, pp. 77–88.

[2] J. Daemen and V. Rijmen, The design of Rijndael: AES-the advanced encryption standard. Springer Science & Business Media, 2013.

[3] M. Agoyan, J.-M. Dutertre, D. Naccache, B. Robisson, and A. Tria, "When clocks fail: On critical paths and clock faults," in Smart Card Research and Advanced Application. Springer, 2010, pp. 182–193.

[4] A. Dehbaoui, J.-M. Dutertre, B. Robisson, P. Orsatelli, P. Maurine, and A. Tria, "Injection of transient faults using electromagnetic pulses-practical results on a cryptographic system-." IACR Cryptology ePrint Archive, vol. 2012, p. 123, 2012.

[5] P. Maurine, K. Tobich, T. Ordas, and P. Y. Liardet, "Yet Another Fault Injection Technique : by Forward Body Biasing Injection," in YACC'2012: Yet Another Conference on Cryptography, Porquerolles Island, France, Sep. 2012. [Online]. Available: http://hal-lirmm.ccsd.cnrs.fr/lirmm-00762035

[6] S. P. Skorobogatov and R. J. Anderson, "Optical fault induction attacks," in Cryptographic Hardware and Embedded Systems-CHES 2002. Springer, 2003, pp. 2–12.

[7] S. Buchner, F. Miller, V. Pouget, and D. McMorrow, "Pulsed-laser testing for single-event effects investigations," Nuclear Science, IEEE Transactions on, vol. 60, no. 3, pp. 1852–1875, June 2013.

[8] V. De, A. Keshavarzi, S. Narendra, and S. Borkar, "Multiple well transistor circuits having forward body bias," Apr. 17 2001, uS Patent 6,218,895. [Online]. Available: https://www.google.com/patents/US6218895

[9] J. Tschanz, J. Kao, S. Narendra, R. Nair, D. Antoniadis, A. Chandrakasan, and V. De, "Adaptive body bias for reducing impacts of die-to-die and within-die parameter variations on microprocessor frequency and leakage," Solid-State Circuits, IEEE Journal of, vol. 37, no. 11, pp. 1396–1402, Nov 2002.

[10] T. Fuhr, E. Jaulmes, V. Lomné, and A. Thillard, "Fault attacks on aes with faulty ciphertexts only," in Fault Diagnosis and Tolerance in Cryptography (FDTC), 2013 Workshop on. IEEE, 2013, pp. 108–118.