Local Heating Attacks on Flash Memory Devices

Sergei Skorobogatov
Computer Laboratory
University of Cambridge
Cambridge, UK
sps32@cam.ac.uk

Abstract **This paper shows how lasers can be used to implement modification attacks on EEPROM and Flash memory devices. This was achieved with inexpensive laser-diode module mounted on a microscope. By locally heating up a memory cell inside a memory array, the contents of the memory can be altered. As a result, the security of a semiconductor chip can be compromised. Even if changing each individual bit is not possible due to the small size of a memory cell, cryptographic keys can still be recovered with brute force attacks. This paper also discusses the limits for the safe use of lasers in semi-invasive attacks without damaging the device under test**

Keywords— **hardware security, thermal attacks, semi-invasive attacks, optical attacks, data retention**

I. Introduction

Secure microcontrollers and smart cards are designed to protect both the integrity and confidentiality of information stored inside their memory. Very often, sensitive information, including encryption keys and passwords, is stored in non-volatile memory like EEPROM and Flash. Inducing memory errors could enable this sensitive information to be deduced. Memory-modification attacks were proposed as a serious threat to semiconductor devices in the late nineties [1], for example, leaking a signing key [2]. Since then several practical ways of implementing such attacks were announced, most notably semi-invasive attacks [3]. These do not require expensive and time-consuming preparation techniques as the passivation layer of the chip remains intact. Also, they do not cause any mechanical damage to the silicon of the device so they are reversible in most cases. So far, as a practical implementation, they were demonstrated in the form of UV attacks [4] and fault injection attacks [5]. However, despite the fact that such semi-invasive attacks using local heating were proposed in 2002, this field has hardly been explored and there are no publications on whether these attacks are possible at all. In this paper, I present the results of an investigation on what an attacker could possibly do by locally heating up a small area inside a semiconductor chip.

The ideal local heat source is laser radiation. My research demonstrates that a powerful enough laser, if focused on the embedded non-volatile memory, can cause its contents to be changed. This in turn could result in some security problems such as the leakage of a secret key [1]. Sections 2 and 4 discuss possible ways of estimating the temperatures reachable with focused lasers.

Lasers have been used for failure analysis in the semiconductor industry for many years. They have also proved their effectiveness in recently introduced semi-invasive attacks which represent a serious threat to many secure semiconductor devices, including smart cards. This paper also evaluates the damage risk to the silicon die during semi-invasive analysis techniques such as laser scanning and optical probing. These techniques evolved from failure-analysis methods widely used in the semiconductor industry [6]. However, there is little information on how damaging to the silicon such laser scanning techniques as optical beam induced current and light-induced voltage alteration could be [7]. Incorrectly selected lasers can cause permanent damage to a semiconductor device by overheating its transistors. Section 5 shows that lasers with more than 5 mW power can potentially cause permanent damage to the non-volatile memory inside a chip if focused on the memory cell for a long time.

II. Background

In order to implement optical attacks on CMOS transistors, the chip surface needs to be accessible. Early optical attacks were demonstrated with light from a photo flash [5]. To influence each memory cell independently, a more focusable source of ionizing radiation is preferable, such as a laser beam [8].

Targets of my experiments were EEPROM and Flash memories. Both of these use floating-gate transistors to store the information [9]. An example EEPROM structure is shown in Figure 1, an example Flash memory structure in Figure 2. EEPROM memory was introduced by Intel in 1980 and offers full electrical control over both the write and erase operations. Due to its high manufacturing cost and complexity, it was not widely used in microcontrollers until the early nineties. Today most microcontrollers and smart cards have either EEPROM or its successor, Flash memory on chip. Flash memory has a simpler structure, faster write and access time, but

2009 IEEE International Workshop on Hardware-Oriented Security and Trust (HOST-2009), 27 July 2009 Moscone Center, San Francisco, CA, USA

unfortunately it cannot be reprogrammed in single bytes. It can be erased only in blocks, which is not convenient for small data updates. Flash memory has many different layouts and each semiconductor manufacturer normally has its own design.

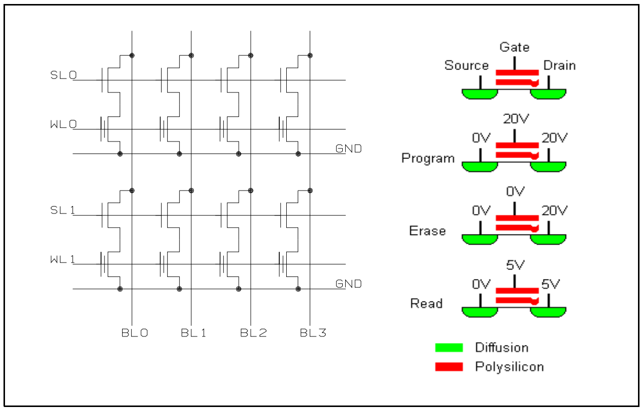


Fig.1. The structure of the EEPROM memory array

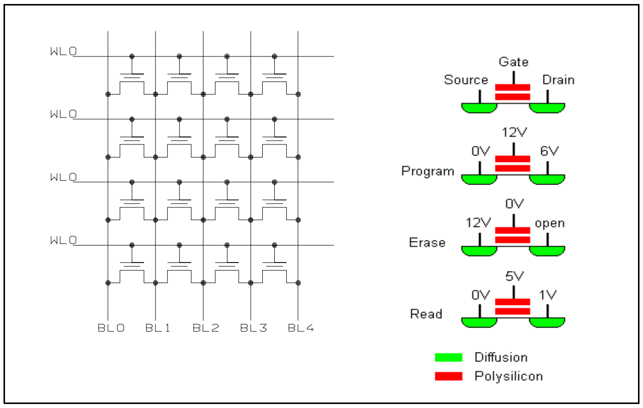


Fig.2. The structure of the Flash memory array

Programmed floating-gate memories cannot store information forever. Various processes, such as field-assisted electron emission and ionic contamination, cause the floating gate to lose its charge, and this increases at higher temperatures. Typical guaranteed data retention times for EEPROM and Flash memories are 40 and 100 years, respectively. However, there is no information on how fast the memory cell loses its state at temperatures above the specified operating range of +125 °C. If the memory transistor can be heated up locally, for example by laser radiation, this might cause it to lose its charge faster and hence change its state. In this case, by moving from one transistor to another in the array, the memory contents can be altered.

The memory storage transistor in the EEPROM array is large enough to precisely focus a laser beam down to each individual cell, as the cell size varies from ten micrometers in old devices to about one micrometer in modern microcontroller chips. Focusing a laser down to a single cell in a Flash array is harder due to the significantly higher density of this type of memory. However, this is still practical for most 8-bit and some 16-bit microcontrollers.

Lasers are also used in the semiconductor industry for crystallisation of a silicon thin film on a glass substrate [10]. However, in order to melt the silicon, which occurs at 1200 °C, a laser with a power of 2.7 W focused at a 170 µm diameter spot is required. For my experiments, a much smaller spot size is required, as well as a lower temperature. With a stationary 650 nm, 100 mW continuous wave laser focused on a 1 µm spot, up to 700 °C in the area under the beam can be expected. However, if the laser is scanning the surface, the heat-up drops proportional to the scanning speed.

III. Experimental Method

For my experiments I chose a common microcontroller, the Microchip PIC16F628 [11], with 2048 14-bit words of Flash and 128 bytes of EEPROM memory on chip. The allocation of data bits in the memory array and the mapping between addresses and the physical location of each memory cell were found using optical probing attacks [8]. To achieve this, the microcontroller was decapsulated in a standard way [12] and placed on a test board with a ZIF socket under a microscope (Figure 3).

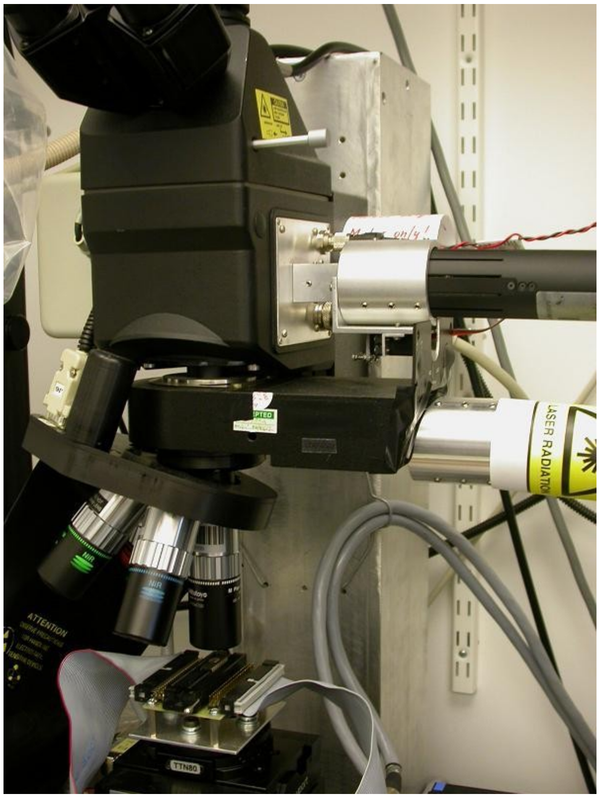


Fig.3. The setup for semi-invasive thermal analysis

The equipment used consisted of a XYZ motorised sample positioner, several laser-diode modules mounted on a Mitutoyo FS60Y optical microscope with long working distance high-magnification objectives and a CCD camera for imaging.

The image of the chip die is presented in Figure 4, while the Flash area is shown in Figure 5. Figure 6 presents the image produced by a video camera during the experiment with a 100× objective lens. For positioning over the die surface, the laser source was set to a safe reference mode (Class 1 laser, <1 mW) in which the image can be taken with a camera and the laser can be directly observed without danger to the eyes. In my first set of experiments, I used the laser with a wavelength of 650 nm. The laser was positioned with 0.1 µm accuracy using the motorised stages. The laser can be focused down to approximately half of its wavelength, which is about 0.35 µm spot size.

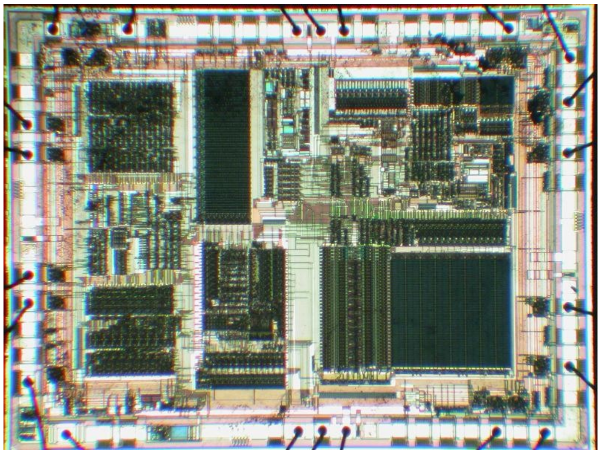


Fig.4. Image of the PIC16F628 microcontroller die

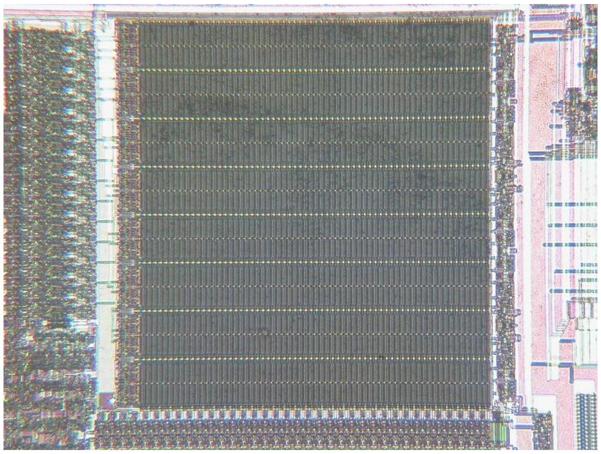


Fig.5. Image of the Flash memory inside the PIC16F628

The PIC16F628 microcontroller was initially programmed with a test pattern in its EEPROM and Flash areas. A specially built programmer (Figure 7) was used for the experiments, allowing easier integration into the test environment. The microcontroller was placed into a test socket and exposed to the laser with different power and duration settings. During the exposure to the laser, the power supply of the microcontroller was switched off in order to prevent any damage that a laser-injected current might cause. After the exposure, the microcontroller was tested in the programmer in order to observe any changes in its EEPROM and Flash memories.

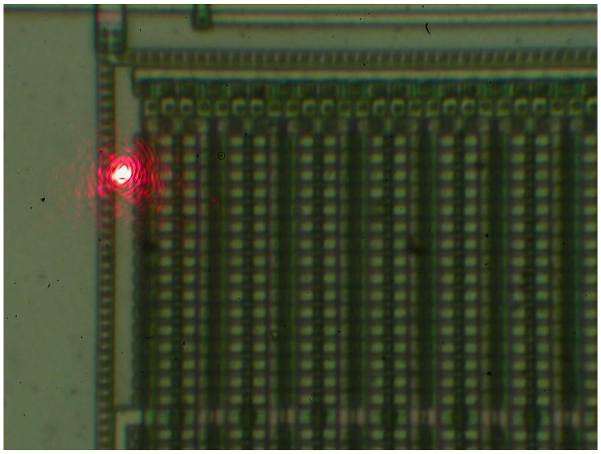


Fig.6. Laser focused with 100× objective near the EEPROM

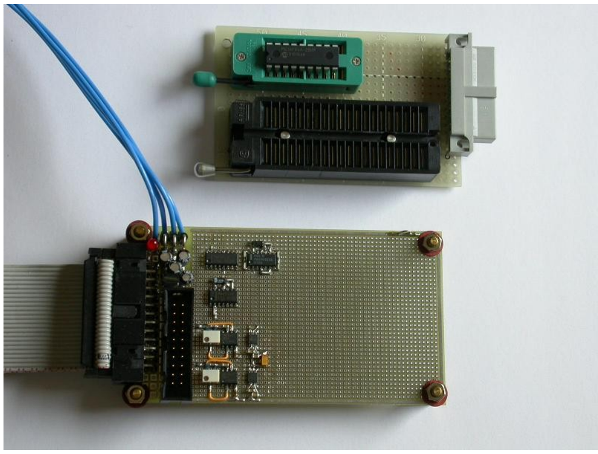
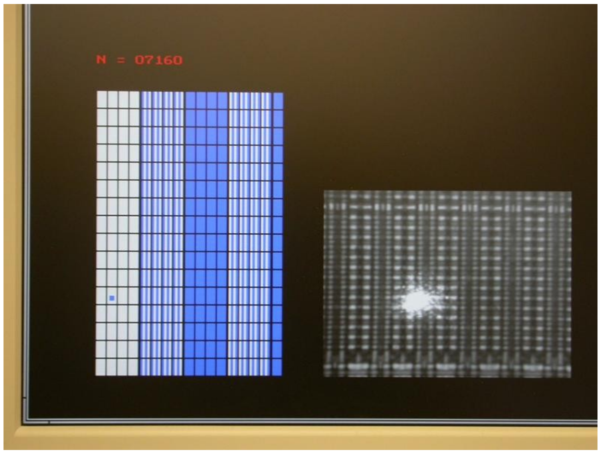


Fig.7. Board for testing the microcontroller



N = 07160

Fig.8. Screen shot of the control program

For easier observation of the results and navigating the laser over the chip surface, a special program was written for a PC.

The program read memory of the PIC microcontroller and displayed its contents according to physical location of each cell. The optical image of the target area was also visible on the display (Figure 8). The number of bits programmed to “0” (white pixels) was also displayed on the screen. That way any changes in the memory contents can easily be spotted.

For comparison with laser heating experiments the same microcontroller was heated up on a laboratory hotplate. This involved exposing it to 450 °C for different time periods, then cooling it down and testing it in order to correlate the results with the laser experiments. The power of the lasers used in my experiments is given for bare laser heads and should be adjusted for the microscope optics. The attenuation coefficients for the 650 nm and 1065 nm lasers were 1.7 and 2.3, respectively.

IV. Results

Both EEPROM and Flash memories inside the PIC16F628 microcontroller were found to be sensitive to local heating. For both memories, heating with a 650 nm, 50 mW laser caused the memory cells to lose their charge (Figure 9). It can be observed that more than one bit was set as the result of the local heating – this is because adjacent memory cells got affected as well. However, at higher temperatures an opposite effect was observed if the heating was performed for too long (Figure 10). This is very likely caused by the degradation of the memory transistor itself resulting in change of its threshold voltage. If the heating was performed for longer than 10 minutes with a 100 mW laser, the memory was permanently damaged and cannot be erased and reprogrammed anymore. This could be used as a permanent modification attack on a device.

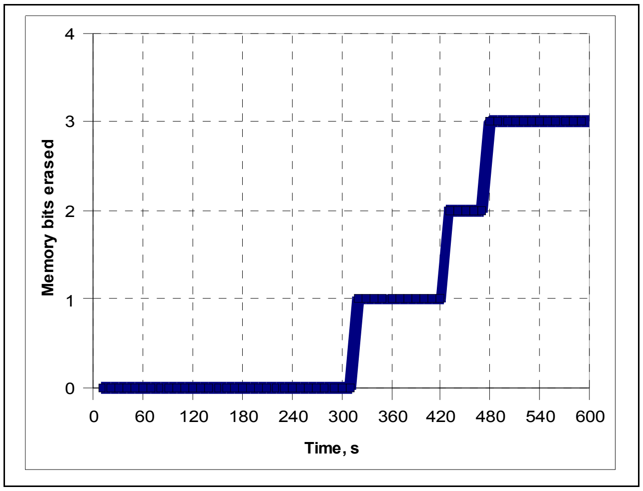


Fig.9. Changes in EEPROM contents during exposure to 50 mW laser

The Flash memory showed very similar results with the only difference being that it was virtually impossible to locally heat up each individual cell because of their smaller size. This resulted in the adjacent memory cells being influenced as well. However, that might not prevent an attacker from guessing an encryption key, because adjacent bits can be brute forced over all possible combinations [1]. If the number of bits changing at a time does not exceed about 30, such a brute force search will not take longer than a few minutes on a modern PC. Not only the key itself can be attacked, but also look-up tables used in encryption, and intermediate values [13].

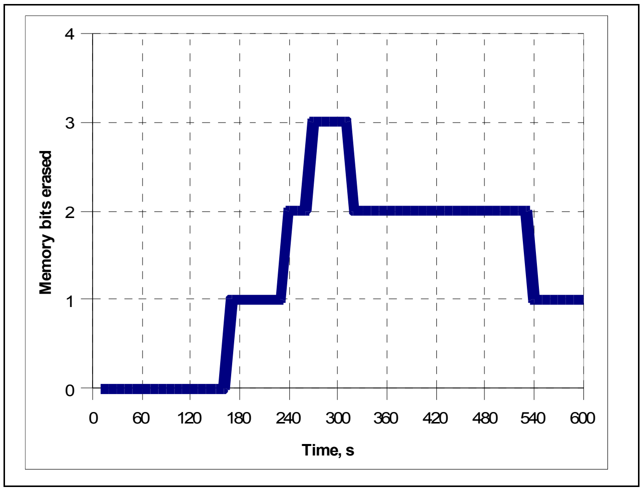


Fig.10. Changes in EEPROM contents during exposure to 100 mW laser

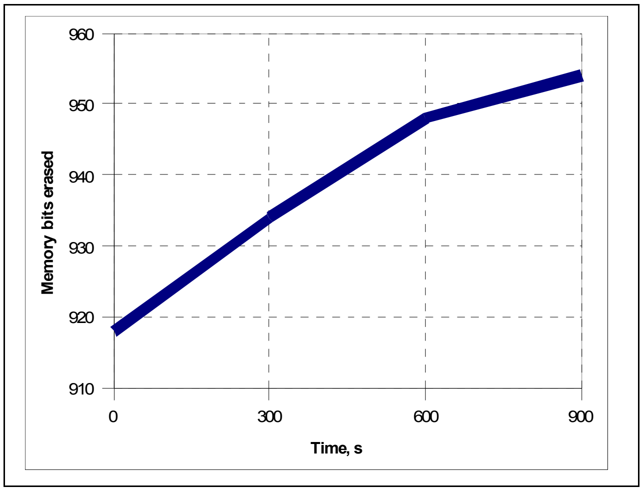
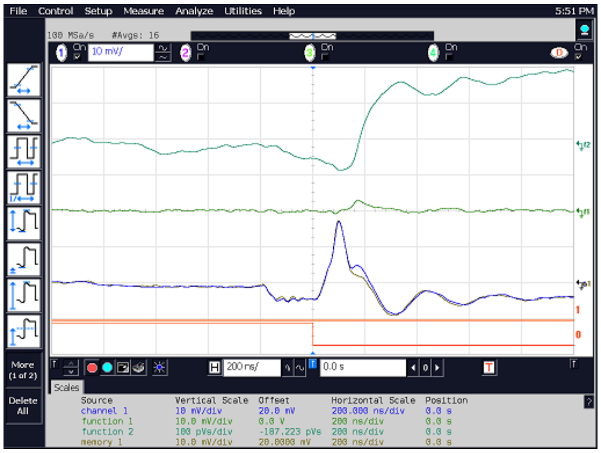


Fig.11. Changes in EEPROM contents during heating to 450 °C

When the same microcontroller was heated up to 450 °C on a hotplate for 2 hours and then tested, its memory did not show any sign of degradation. However, when the EEPROM and Flash of the chip were programmed and then partially erased by terminating the erase operation after 200 µs, they started showing dependency to the heating (Figure 11). The changes were very slow, leading to the conclusion that the local temperature to which the laser is heating up the surface was higher than 450 °C. Above 450 °C the plastic of the chip started degrading making applying higher temperatures infeasible.

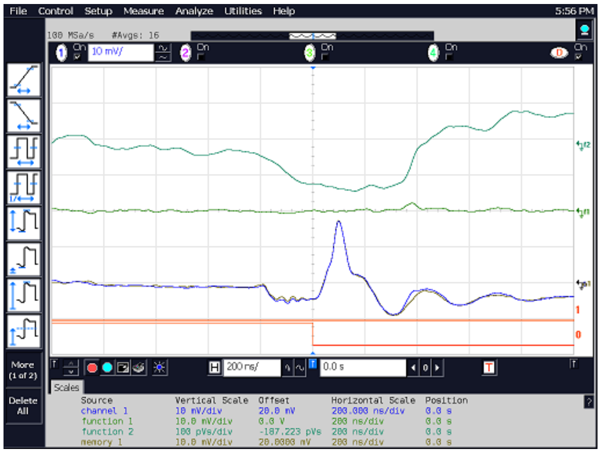
V. Limitations and Further Improvements

The PIC16F628 is a relatively old microcontroller built with 0.9 µm technology with two metal layers. The majority of modern microcontrollers is built with 0.35 µm technology (three or four metal layers) and some high-end microcontrollers employ 0.18 µm technology (up to six metal layers). This fact, together with interlayer polishing and gap filling techniques, significantly reduces the amount of laser radiation which could reach the underlying transistor gates.



Source	Vertical Scale	Offset	Horizontal Scale	Position
channel 1	10 mV/div	20.0 mV	200.000 ns/div	0.0 s
function 1	10.0 mV/div	0.0 V	200 ns/div	0.0 s
function 2	100 pVs/div	-187.223 pVs	200 ns/div	0.0 s
memory 1	10.0 mV/div	20.0000 mV	200 ns/div	0.0 s

Fig.12. Difference in the power trace caused by bit change



Source	Vertical Scale	Offset	Horizontal Scale	Position
channel 1	10 mV/div	20.0 mV	200.000 ns/div	0.0 s
function 1	10.0 mV/div	0.0 V	200 ns/div	0.0 s
function 2	100 pVs/div	-187.223 pVs	200 ns/div	0.0 s
memory 1	10.0 mV/div	20.0000 mV	200 ns/div	0.0 s

Fig.13. Difference in the power trace after heating

One way to improve the attack could be to approach memory cells from the rear side of a chip. However, in this case a laser with longer wavelength must be used, and higher power is required due to absorption of light in the silicon substrate. By using 1300 nm wavelength, which does not cause ionisation of the silicon, thermal attacks can be applied to a powered-up microcontroller. Some experiments were carried out using a 1065 nm, 50 mW laser focused from the rear side. However, no detectable changes were observed even after 30 minutes of exposure, suggesting that a laser with higher power is needed.

Another improvement could be detecting partially modified memory cells. This can be done through power-analysis observation during access to the memory cell [14]. Figure 12 presents an oscilloscope screen shot of the difference between power traces acquired from the chip with the same Flash memory location programmed to 0x3FFF and 0x3FFE (single bit difference). The two acquired power traces are at the bottom and the difference between them is in the middle of the screen shot. At the top of the screen the integral of the difference is shown. Figure 13 presents the difference between the location holding 0x3FFE and the same location after exposure to 650 nm, 10 mW laser for 30 seconds. Although the memory location still read as 0x3FFE, the difference between the partially erased and the untouched cell is visible in the power trace, however, delayed by about 200 ns.

The same experiments were repeated with a 1065 nm, 50 mW laser focused from the rear side. The changes were detectable in the power trace only after 5 minutes of exposure. This indicates that for rear-side approach more powerful lasers are required due to reflections and optical absorption of the light in bulk silicon.

Another approach for observing partial loss of charge in the memory cell is the same technique used to determine the threshold voltage of the memory cell [15]. This works well for the PIC16F628 microcontroller and allows precise detection of any changes of the charge inside the Flash and EEPROM memory cell in the nearly erased state. This way even small changes caused by the laser can be observed. I carried out a set of experiments and it turned out that the memory cell starts losing its charge even when using a 5 mW laser.

All the above mentioned observations apply for EEPROM memory as well. However, as EEPROM cells are larger than Flash, these attacks are easier to carry out and each memory cell can be individually attacked. Further research might involve applying similar techniques to other memory types, such as FRAM and MRAM.

VI. Conclusion

My experiments showed how semi-invasive thermal injection attacks can be used to modify the contents of EEPROM and Flash memories widely used in secure microcontrollers and smart cards, which could represent a serious threat to hardware security. These attacks can be used to cause permanent fault injection. Even when several bits of the data are flipped simultaneously, this can still help to recover a key by brute forcing over all possible combinations.

When the memory cell is modified in a way that maintains its binary value, the change in the charge on the floating gate can be monitored through power analysis techniques.

Possible forms of protection against these attacks could involve using tamper sensors to prevent direct access to the chip surface, as well as implementing light sensors. Top metal protection might help, but it is very likely to be overcome by approaching the sample from the rear side. Using modern deep submicron technologies will also eliminate most of these attacks, but can again be overcome by using the rear-side approach.

When lasers are used for semi-invasive attacks it is always important to estimate any danger from the heat produced by the lasers. Such memories as EEPROM and Flash are particularly sensitive to overheating and it might be possible to permanently damage the chip if high-power lasers are used for optical probing or semi-invasive attacks. I showed that lasers with output power larger than 5 mW can be dangerous to the internal memory if focused at a memory cell for long enough. The local temperature at the point where the laser is focused appears to reach over 450 ºC. However, in practice this depends on many factors and cannot be precisely simulated or directly measured.

References

[1] R.J. Anderson, M.G. Kuhn, “Low cost attacks on tamper resistant devices,” Security Protocols, 5th International Workshop, Paris, France, April 1997

[2] D. Boneh, R.A. DeMillo, R.J. Lipton, “On the importance of checking cryptographic protocols for faults,” Advances in Cryptology – Eurocrypt 97, LNCS 1233, pp. 37–51

[3] S. Skorobogatov, “Semi-invasive attacks - a new approach to hardware security analysis,” Technical Report UCAM-CL-TR-630, University of Cambridge, Computer Laboratory, April 2005

[4] R. J. Anderson, M.G. Kuhn, “Tamper resistance – a cautionary note,” The Second USENIX Workshop on Electronic Commerce, Oakland, California, November 1996

[5] S. Skorobogatov, R. Anderson, “Optical fault induction attacks,” Cryptographic Hardware and Embedded Systems Workshop (CHES-2002), LNCS 2523, pp. 2–12

[6] L.C. Wagner, “Failure analysis of integrated circuits: tools and techniques,” Kluwer Academic Publishers, 1999

[7] C. Ajluni, “Two new imaging techniques promise to improve IC defect identification,” Electronic Design, Vol. 43(14), July 1995, pp. 37–38

[8] D. Samyde, S. Skorobogatov, R. Anderson, J.-J. Quisquater, “On a new way to read data from memory,” SISW2002 First International IEEE Security in Storage Workshop

[9] W.D. Brown, J.E. Brewer, “Nonvolatile semiconductor memory technology: a comprehensive guide to understanding and using NVSM devices,” IEEE Press, 1997

[10] G. Andrä, J. Bergmann, F. Falk, “Laser crystallized multicrystalline silicon thin films on glass,” Thin Solid Films, Volume 487, Issues 1-2, September 2005, pp. 77–80

[11] PIC16F62X Data Sheet, Flash-Based 8-Bit CMOS Microcontroller. http://ww1.microchip.com/downloads/en/DeviceDoc/40300C.pdf

[12] O. Kömmerling, M.G. Kuhn, “Design principles for tamper-resistant smartcard processors,” USENIX Workshop on Smartcard Technology, Chicago, Illinois, USA, May 1999

[13] G. Piret, J.-J. Quisquater, “A differential fault attack technique against SPN structures, with application to the AES and Khazad,” C.D. Walter et al. (Eds.): CHES 2003, LNCS 2779, pp. 77–88

[14] P. Kocher, J. Jaffe, B. Jun, “Differential power analysis,” CRYPTO’99, LNCS 1666, pp. 388–397

[15] S. Skorobogatov, “Data remanence in flash memory devices,” Cryptographic Hardware and Embedded Systems Workshop (CHES-2005), LNCS 3659, pp. 339–353