Optically Enhanced Position-Locked Power Analysis

Sergei Skorobogatov

University of Cambridge, Computer Laboratory,
15 JJ Thomson Avenue, Cambridge CB3 0FD, United Kingdom
sps32@cl.cam.ac.uk

Abstract. This paper introduces a refinement of the power-analysis attack on integrated circuits. By using a laser to illuminate a specific area on the chip surface, the current through an individual transistor can be made visible in the circuit's power trace. The photovoltaic effect converts light into a current that flows through a closed transistor. This way, the contribution of a single transistor to the overall supply current can be modulated by light. Compared to normal power-analysis attacks, the semi-invasive position-locking technique presented here gives attackers not only access to Hamming weights, but to individual bits of processed data. This technique is demonstrated on the SRAM array of a PIC16F84 microcontroller and reveals both which memory locations are being accessed, as well as their contents.
Key words: side-channel attacks, power analysis, semi-invasive attacks, optical probing

1 Introduction

Power analysis, especially in the form of differential power analysis (DPA), became a serious concern since it was first announced in 1999 by Kocher et al. [1]. Since then, it proved to be a useful technique to breach security in many devices, including smartcards [2]. During the last six years, many attempts were made to improve protection against power analysis. This involved both hardware and software countermeasures [3,4,5]. On one hand, such improvements reduced the success of known techniques, on the other, they only pushed away poorly funded or less knowledgeable attackers, thereby creating the impression of an already solved problem.

Power analysis attacks had a huge impact on the industry since their effectiveness in helping to break many cryptographic algorithms was demonstrated in the late nineties [2]. However, in spite of the relatively simple setup necessary for power analysis (resistor in the ground line, digitizing oscilloscope and a computer) it is still not reliably and straightforwardly applicable to each situation. This is due to a number of reasons. Firstly, the power analysis technique is usually applied to a whole chip rather than to a small area of interest. As a result, power transitions in areas that are not relevant to the data processing also affect the power trace. Secondly, as the power fluctuations are affected by a number of bits being set or reset, only a Hamming weight of data (number of bits set) can be guessed, rather than the actual value. Finally, in order to get a reliable result from a power analysis, often hundreds or even thousands of traces have to be acquired and averaged. This is because the signal from a single transition is too small compared to the inevitable noise from the resistor in the power line and the noise from the A/D converter of the oscilloscope. Also, the number of transitions happening at a time might be so high that the signal from a single bit of information would be too small to be distinguished with 8-bit resolution. The easiest way to increase the resolution is averaging the signal. However, this requires precise triggering or post processing of the acquired power traces.

Recently introduced electro-magnetic analysis (EMA) [6] can increase the level of a useful signal by placing an antenna close to the area of interest, for example, above the internal RAM, CPU or ALU. However, this is still not enough to distinguish between values of data with the same Hamming weights, because the minimum size of the antenna is significantly larger than the data buffer or the memory cell.

In our laboratory, we have for many years tried to improve the effectiveness of power analysis during security evaluations of microcontrollers and smartcards. One idea was to combine optical probing attacks [7] with a standard power analysis setup. As such analysis will require partial decapsulation of the chip without direct connection to its internal wires, it should be considered to be a semi-invasive attack. If we could influence the power consumption of a certain area on the chip surface by exposing it to ionizing radiation, we would be able to see if the signal in the power trace came from this area or not. Thus, by moving from one location to another, we should be able to recognise which areas on the chip contribute to the power trace. Vice versa, if we know the point of interest, for example, an address of the variable which holds the security flag, we could point to the corresponding location inside the SRAM and find out the exact time when this memory address is accessed.

Lasers have been used in failure analysis for testing states of on-chip transistors for many years and the ability of laser radiation to ionize silicon substrate was studied long ago [8]. One of these techniques, called light-induced voltage alteration (LIVA) [9], uses the photovoltaic effect to distinguish between open and closed transistors. However, this technique assumes that the chip is in a static condition and the result of scanning cannot be updated faster than a few frames per second. Another technique, published in 1992 [10], was designed specifically to detect electrical signals at internal nodes in silicon ICs and uses the phenomenon that charge density affects the refractive index of silicon within the device. However, the setup necessary for detecting this change of refractive index in a tiny area is very difficult and expensive to implement. Therefore, methods which are less expensive and easier to implement are desirable.

Successful position-locked power analysis would be highly useful for failure analysis and security testing of secure microcontrollers as it would offer a faster and less expensive solution. It would also help in partial reverse engineering of a chip operation and help with the analysis of signals inside a chip. Of course, failure analysis techniques such as using a focused-ion beam (FIB) machine followed by microprobing [11] will with high probability give the required result, but at the cost of many hours of preparation work and a large number of analysed points. Optical probing can give a result in a significantly shorter time (normally minutes) and does not require expensive sample preparation techniques, which often irreversibly modify the die of an analysed chip.

In spite of the seeming simplicity of the proposed idea, it took me a long time until I managed to get a useful and reliable result. The main problem to solve was to find a reliable way of influencing the power consumption from a particular CMOS inverter, flip-flop or memory cell, without interfering with its operation.

2 Background

Most digital circuits built today are based on CMOS technology, using complementary transistors as basic elements. When a CMOS gate changes its state, it charges/discharges a parasitic capacitive load and causes a dynamic short circuit of the gate [12]. The more gates change their state, the more power is dissipated. The current consumed by a circuit can be measured by placing a 10–50 Ω resistor in the power supply line, usually a ground pin, because an ordinary oscilloscope probe has a ground connection.

Drivers on the address and data bus consist of many parallel inverters per bit, each driving a large capacitive load. During transition they cause a significant power surge, in the order of 0.5–1 mA per bit, which is sufficient to estimate the number of bus bits changing at a time using a 12-bit A/D converter [13]. By averaging the measurements of many repeated identical operations, smaller transitions can be identified. Of particular interest for attacking cryptographic algorithms would be observing the state change of a carry bit. Each type of instruction executed by a CPU causes different levels of activity in the instruction decoder and arithmetic unit, therefore instructions can be often quite clearly distinguished such that parts of algorithms can be reconstructed.

Memory inside a microcontroller or a smartcard, especially SRAM, is of particular interest to an attacker, because it may store sensitive variables, encryption keys, passwords and intermediate results of cryptographic operations. When accessing an SRAM memory cell, not only data bits are contributing to the power trace, but also the address being accessed, because of the different number of bits set inside the address latches. An SRAM cell consists of six transistors (Figure 1), four of which create a flip-flop while the other two are used for accessing the cell inside the memory array. An SRAM write operation often generates the strongest signal, because the output of the flip-flop is connected to the output of the bit lines, causing a current surge. However, still only bits which are changed during the write operation will contribute to the power trace.

In order to apply optical attacks, the surface of the chip must be accessible. Originally, optical attacks were demonstrated with light from a photoflash [7]. In order to influence each memory cell independently, a better light source should

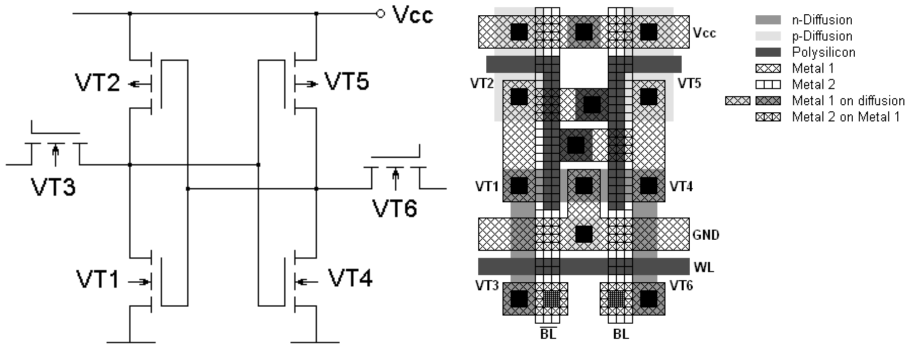

Fig. 1. The architecture and layout of an SRAM cell

be used, for example a laser beam [14]. As the target of my experiments was SRAM, we should look at the structure of such memory first. One example of the SRAM layout is shown in Figure 1. If it is possible to partially open one of the transistors forming the flip-flop, then the cell will behave differently when accessed, consuming more power, and this can be detected by comparing the acquired power trace with a reference trace. If the flip-flop switches, this will reduce the leakage current, because the leaking channel will be closed. However, if it were possible to influence both transistors of the flip-flop simultaneously, then any access to the cell will result in a change of the power consumption.

Laser radiation can ionize semiconductor regions in silicon chips if its photon energy exceeds the semiconductor band gap (> 1.1 eV or $\lambda < 1100$ nm). This results in free carriers (electrons and holes) being created that produce a photocurrent at p-n junctions and this can be detected, for example, by observing a voltage drop over a resistor inserted in the power supply line. The injected photocurrent can also influence the normal operation of the chip and this can be simulated [15]. From a practical point of view, it is more efficient to influence n-channel transistors, as they have higher doping concentrations and their carriers (electrons) have higher mobility. P-channel transistors can be influenced as well, but will require a higher level of ionizing radiation.

3 Experimental Method

For my experiments, I chose a common microcontroller, the Microchip PIC16F84 [16], which has 68 bytes of SRAM memory on chip. The allocation of data bits in the memory array and the mapping from the addresses to the corresponding physical location of each memory cell were already documented for this chip [7]. The microcontroller was decapsulated in a standard way [13] and placed in a computer-controlled test board with a ZIF socket under a special microscope for semi-invasive analysis (Figure 2).

As a light source, I chose a red laser, which can be easily focused down to a submicron point on a chip surface. The most difficult part was choosing the right equipment for my experiments. Firstly, precise control over the sample position with submicron precision was essential. Secondly, as any sort of fault injection was undesirable, precise control over the laser power was required. Finally, because the chip has a metal layer, the optical system must allow focusing the laser beam at any point within several micrometers distance from the focal plane of the microscope. Otherwise, most of the energy will be reflected or deflected by the metal wires. Optical fault injection equipment, such as industrial laser cutters [17], was unsuitable for my needs because they offer limited control over timing. I performed several tests and also found that the pulses emitted by such laser cutters have too much power variability and too short and uncontrollable duration.

Fig. 2. Test setup for semi-invasive analysis

After a long time of searching, I finally chose equipment from Semiconductors Research Ltd – a company specialising in security testing and evaluation of integrated circuits [18]. What I used in my experiments was a special semi-invasive diagnostic system that combines several laser sources with extended positioning control, mounted on a specialized optical microscope with long working distance high-magnification objectives and a CCD camera for imaging. The software control toolbox for this equipment allowed fully computerised control over all parameters of the laser sources in both manual and automatic modes (Figure 2). The last capability was very important as it allowed me to synchronize the supply of test signals with the photon sources. In addition, the system has a very useful high-resolution laser scanning capability, which helps to find active areas on the chip surface.

To acquire power traces with a sampling rate of 500 MHz, I used a Tektronix TDS7054 oscilloscope with a P6243 active probe (DC coupled) connected on the test board across a 10 Ω resistor. A metal-film resistor was used to minimize noise. The oscilloscope's built-in analogue 20 MHz low-pass filter was activated (anti-aliasing filter), along with the "Hi-Res" acquisition mode, in which a digital low-pass filter implemented in the oscilloscope further reduces noise and increases the effective A/D-converter resolution to slightly more than 8 bits per sample.

The images of the SRAM area and the image produced by a video camera during the experiment with a 100$\times$ objective are presented in Figure 3. The laser source (639 nm) was set to a safe reference mode (0.01 mW) in which the image can be taken with a camera and the laser can be directly observed without any danger to eyes.

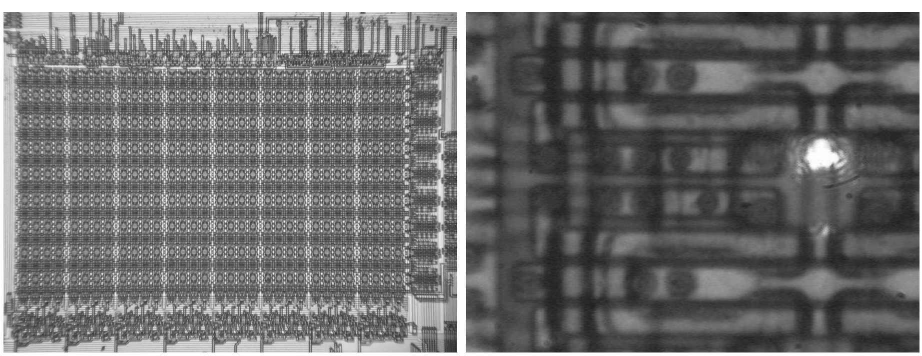

Fig. 3. Optical image of the SRAM area in the PIC16F84 microcontroller and the laser beam focused with a 100$\times$ objective

Although the circuit diagrams of most SRAM cells are identical, their layouts can differ. The layout of the SRAM cell presented in Figure 4 is very similar to the one found in the PIC16F84.

In order to locate active areas inside the memory cell, a passive laser scanning operation was applied to the sample. In failure analysis, this technique is called optical beam induced current (OBIC) and the image produced as location-dependent induced current. The result of scanning the SRAM cell with the laser is presented in Figure 4. Having such a reference helps in focusing the laser beam on any of the MOS transistors forming the flip-flop. The right bright areas correspond to light-sensitive areas of p-channel transistors VT2 and VT5, where the left grey lines correspond to n-channel transistors VT1 and VT4. The left grey areas correspond to light-sensitive areas of the select transistors VT3 and VT6.

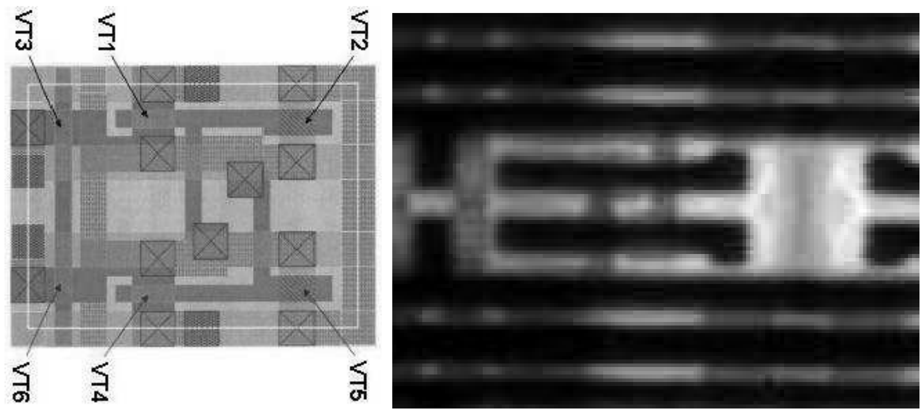

Fig. 4. Layout and laser scan of the SRAM cell

The PIC16F84 microcontroller was programmed with a simple test program which firstly initialised SRAM locations 0x10, 0x11, 0x20, 0x31 with value 0x00 and locations 0x21, 0x30, 0x40, 0x41 with 0xFF, and then executed the following code:

```
bsf PORTA,test ; generate pulse on PA0 for triggering
bcf PORTA,test
nop
movf 0x10, W   ; read location 0x10
nop
movwf 0x11     ; write to location 0x11
nop
movf 0x20, W   ; read location 0x20
nop
movwf 0x21     ; write to location 0x21
nop
movf 0x30, W   ; read location 0x30
nop
movwf 0x31     ; write to location 0x31
nop
movf 0x40, W   ; read location 0x40
nop
movwf 0x41     ; write to location 0x41
```

Finally, it outputs the contents of all memory locations to Port B.

I put NOP commands between each instruction to avoid the influence of instruction pipelining, so that the result from a previous instruction will not affect the next instruction. This was necessary only for the evaluation stage. In a power-analysis comparison, such an influence will be eliminated anyway, because we are not interested in the absolute values in the power traces, but in their changes. However, pipelining might pose problems for recognising particular instructions.

Previous experiments with power analysis of a similar microcontroller [19] showed that instructions can be distinguished, and that there is a correlation to the number of bits set or changed in the data during operations. My aim was to identify, which particular bits were set and which addresses in the memory array were accessed.

4 Results

Writing into an SRAM cell causes a significantly larger current response than a read operation, therefore my first experiment was performed on the SRAM memory locations being written by the test program. The aim was to check whether write operations performed on a particular memory location can be reliably identified.

In the test program, the write operation does not change the state of memory locations 0x11 and 0x41, which are 0x00 and 0xFF, respectively. Location 0x21 was changed from 0x00 to 0xFF and location 0x31 from 0xFF to 0x00. For each memory cell, I performed a series of tests with different focusing points and power settings for the laser. The optimum laser power I found to be between 1 mW and 3 mW. The laser was switched on in the beginning of the test program and switched off before sending the contents of the memory locations to Port B.

As predicted, the maximum response was received from areas close to n-channel transistors. I averaged the traces of 16 repeated program executions to reduce noise and the acquired waveform with the laser focused on transistor VT1 of memory location 0x31 is presented in Figure 5. The power trace is compared with a reference waveform acquired without laser light. The difference between the reference and the acquired waveforms is presented in enlarged scale. The trace difference is clearly noticeable, however, the signal is very close to the noise level. Any attempts to influence transistor VT1 at address 0x21 and transistor VT4 at 0x31 were unsuccessful. Also, for unchanged locations (0x11, 0x41), I was unable to see any noticeable change in the power consumption.

Any attempts to improve the signal-to-noise ratio by increasing the laser power caused the memory cell to change its state, resulting in noticeable changes in the power analysis traces (Figure 6). Similar waveforms, if the state of the memory cell was changed, were received for memory locations 0x11, 0x21 and 0x41. This was still a positive result, because it allowed detection of memory access events, however, from an attacker's point of view, it is always better to be unnoticeable.

Similar measurements were performed for memory locations which were read by the test program. Unfortunately, I received only a very small signal response, which was very hard to distinguish from noise. Again, increasing the laser power caused these memory locations to change their state and this was detectable in the power trace in a similar way as with the written locations.

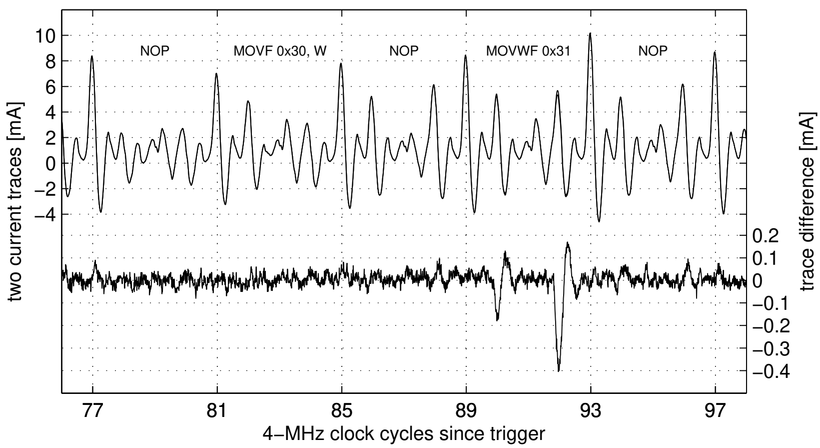

Fig. 5. Laser focused on VT1 of memory cell 0x31, write leaves state unchanged

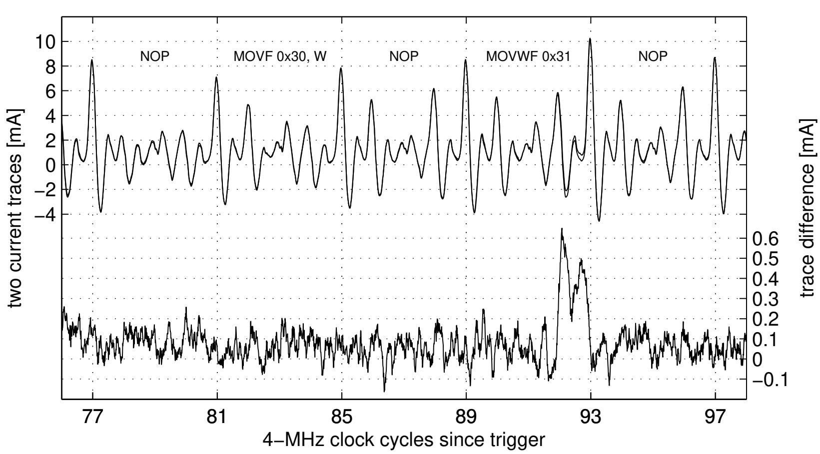

Fig. 6. Laser focused on VT1 of memory cell 0x31, write changes state

My next idea was to try focusing the laser at the area in between the two n-channel transistors, hoping that this will influence both CMOS inverters of the flip-flop and, therefore, might result in influencing the power consumption each time the memory cell was accessed (previously it was either VT1 or VT4 which influenced the signal). Again, I decided to start with the write operations as they always give a stronger signal in the power trace.

However, the result of the measurements surpassed my expectations. The difference signal had significantly increased, such that it became possible to see a clearly distinguishable difference between two traces, even without averaging the waveforms (Figure 7). Still, increasing the laser power resulted in the contents of the memory location to be changed (Figure 8). However, the difference in the waveforms is significantly easier to distinguish than before, when either VT1 or VT4 was influenced.

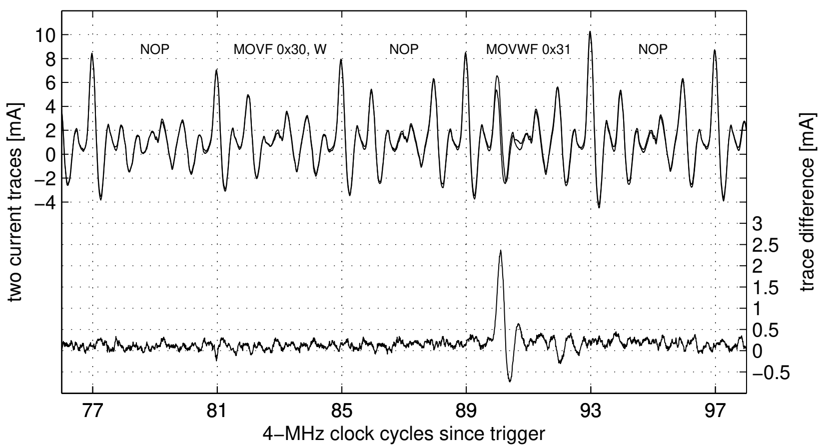

Fig. 7. Laser focused on VT1+VT4 of memory cell 0x31, write leaves state unchanged

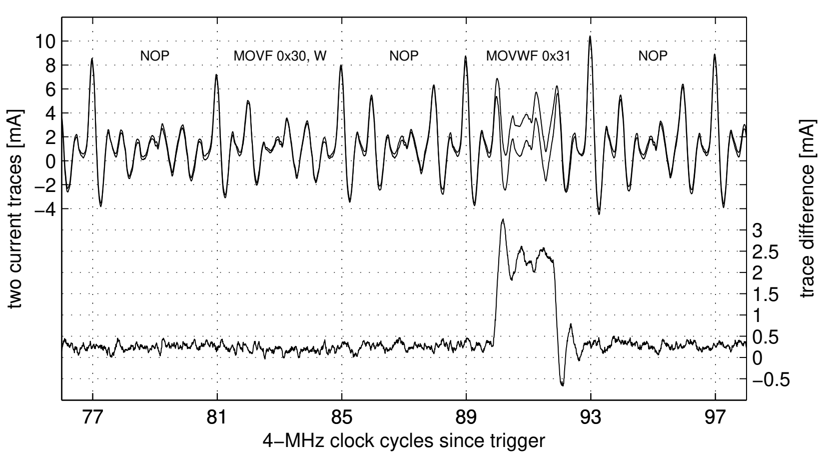

Fig. 8. Laser focused on VT1+VT4 of memory cell 0x31, write changes state

This is very likely an outcome of a short circuit created inside a memory cell if both n-channel transistors forming a flip-flop were opened for a short period of time. Such a situation happens because the ionizing radiation creates excessive carriers, which require additional time to recombine, keeping a transistor in the open state longer. I described the influence of laser radiation on microcontrollers in the form of laser pulses already in [20]. If the energy of the laser is too high, the memory cells become unstable and can spontaneously switch into the other state. This causes a surge in the power consumption.

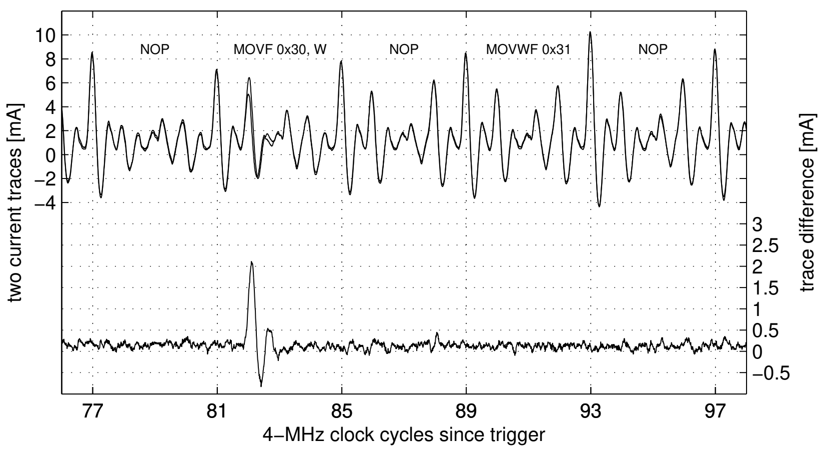

Fig. 9. Laser focused on VT1+VT4 of memory cell 0x30, read

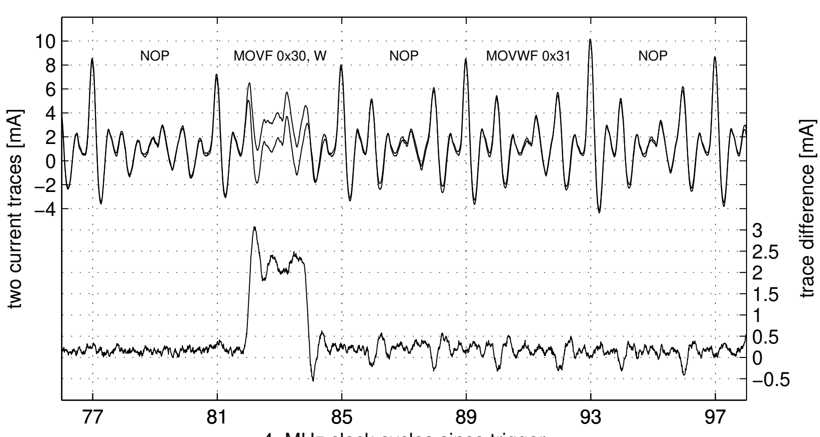

Fig. 10. Laser focused on VT1+VT4 of memory cell 0x30, read changes state

Applying the same approach to a memory addresses being read, the same level of current response was achieved when the state of a memory cell was not changed (Figure 9). However, higher laser power was destructive to the memory contents (Figure 10). Repeating the non-destructive operation of data analysis for each bit of the memory with the same address revealed the actual value of the byte.

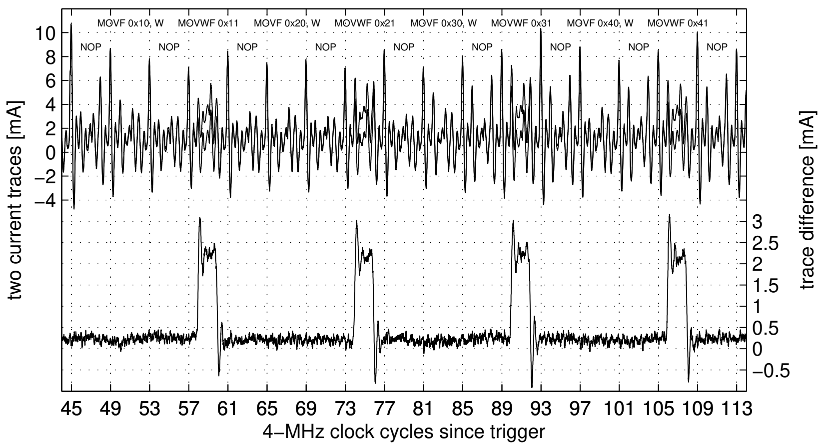

Fig. 11. Laser focused on VT3+VT6 of memory cell 0x31

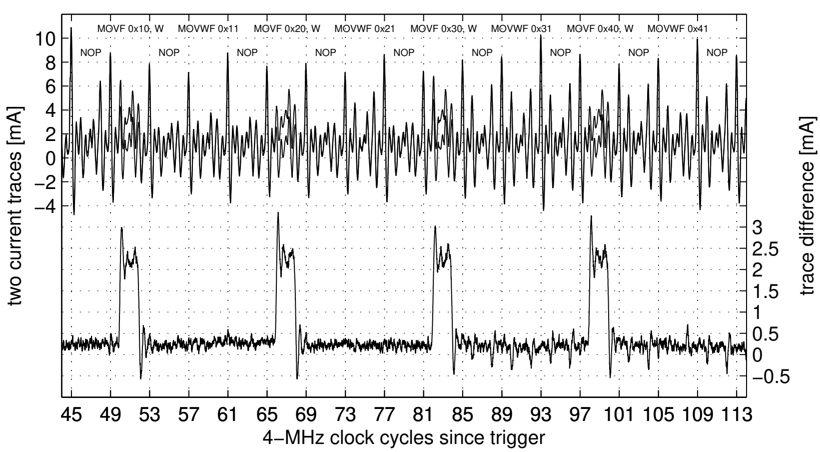

Fig. 12. Laser focused on VT3+VT6 of memory cell 0x30

Another surprise came at a point when a laser was focused on the area between cell select transistors VT3 and VT6. In this case, the whole column of memory cells was affected, independently of which cell in the row was influenced. In my first experiment, the laser beam was pointed between VT3 and VT6 of memory cell 0x31, which caused all cells from this row (addresses 0x31, 0x41, 0x11 and 0x21) to be detectable in the power trace (Figure 11). Similar, by pointing between VT3 and VT6 of memory location 0x30, responses were received if any of the addresses 0x30, 0x40, 0x10 and 0x20 were accessed (Figure 12). However, in both experiments the state of the selected memory locations always changed to zero.

5 Limitations and Further Improvements

My results were achieved on a relatively old microcontroller (PIC16F84) built with 0.9 μm technology. The majority of modern microcontrollers are built with 0.35 μm and 0.25 μm technology (three or four metal layers) and some high-end microcontrollers employ now 0.18 μm technology (up to six metal layers). This fact, in addition to interlayer polishing and gap filling techniques, significantly reduces the amount of laser radiation which can reach the underlying transistor gates.

One improvement could be to approach memory cells from the rear side of the chip. However, in this case, laser radiation with a longer wavelength must be used, which causes lower levels of ionization and also creates unnecessary carriers in the whole volume of the silicon substrate. In order to achieve similar results, it might be necessary to reduce the thickness of the substrate.

6 Conclusions

My experiments showed how combining optical probing techniques with power analysis methods can significantly improve the results. Using such techniques, partial reverse engineering to locate data bits and addresses being accessed in memory becomes easier and significantly faster compared to with other methods [20]. However, this technique has some limitations, especially for modern deep submicron technologies, where multiple metal layers and small transistor sizes prevent easy and precise analysis. Further improvements to these methods might involve approaching the die from its rear side, but this requires more expensive equipment.

Possible forms of protection against such 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 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.

7 Acknowledgements

I would like to thank Semiconductors Research Ltd for providing me with the special equipment necessary for optical analysis of semiconductors. I would also like to thank Markus Kuhn for his helpful discussions and Matlab programming.

References

1. Paul Kocher, Joshua Jaffe, Benjamin Jun: Differential Power Analysis. CRYPTO '99, LNCS, Vol. 1666, Springer-Verlag, 1999, pp. 388–397
2. Thomas Messerges, Ezzy Dabbish, Robert Sloan: Investigations of Power Analysis Attacks on Smartcards. USENIX Workshop on Smartcard Technology, Chicago, Illinois, USA, May 10–11, 1999
3. Jean-Sebastien Coron: Resistance against Differential Power Analysis for Elliptic Curve Cryptosystems. Cryptographic Hardware and Embedded Systems Workshop (CHES-1999), LNCS, Vol. 1717, Springer-Verlag, 1999, pp. 292–302
4. Simon Moore, Ross Anderson, Robert Mullins, George Taylor, Jacques Fournier: Balanced Self-Checking Asynchronous Logic for Smart Card Applications. Microprocessors and Microsystems Journal, Vol. 27, No. 9 (October 2003), pp 421–430
5. Thomas Popp, Stefan Mangard: Masked Dual-Rail Pre-charge Logic: DPA-Resistance Without Routing Constraints, Cryptographic Hardware and Embedded Systems Workshop (CHES-2005), LNCS, Vol. 3659, Springer-Verlag, 2005, pp. 172–186
6. Jean-Jacques Quisquater and David Samyde: ElectroMagnetic Analysis (EMA): Measures and Counter-Measures for Smard Cards. Smart Card Programming and Security (E-smart 2001), Cannes, France, LNCS Vol. 2140, Springer-Verlag, 2001, pp. 200–210
7. Sergei Skorobogatov, Ross Anderson: Optical Fault Induction Attacks, Cryptographic Hardware and Embedded Systems Workshop (CHES-2002), LNCS Vol. 2523, Springer-Verlag, 2002, pp. 2–12
8. D.H. Habing: Use of Laser to Simulate Radiation-induced Transients in Semiconductors and Circuits. IEEE Transactions on Nuclear Science, Vol. 12(6), December 1965, pp. 91–100
9. Cheryl Ajluni: Two New Imaging Techniques Promise to Improve IC Defect Identification. Electronic Design, Vol. 43(14), July 1995, pp. 37–38
10. H.K. Heinrich, N. Pakdaman, J.L. Prince, G. Jordy, M. Belaidi, R. Franch, D.C. Edelstein: Optical Detection of Multibit Logic Signals at Internal Nodes in a Flip-chip Mounted Silicon Static Random-Access Memory Integrated Circuit. Journal of Vacuum Science and Technology, Microelectronics and Nanometer Structures, Vol. 10(6), November 1992, pp. 3109–3111
11. Lawrence C. Wagner: Failure Analysis of Integrated Circuits: Tools and Techniques. Kluwer Academic Publishers, 1999
12. Manfred Aigner, Elisabeth Oswald: Power Analysis Tutorial `http://www.iaik.tugraz.at/aboutus/people/oswald/papers/dpa_tutorial.pdf`
13. Oliver Kömmerling, Markus G. Kuhn: Design Principles for Tamper-Resistant Smartcard Processors. USENIX Workshop on Smartcard Technology, Chicago, Illinois, USA, May 10–11, 1999
14. David Samyde, Sergei Skorobogatov, Ross Anderson, Jean-Jacques Quisquater: On a New Way to Read Data from Memory. SISW2002 First International IEEE Security in Storage Workshop, Greenbelt Marriott, Maryland, USA, December 11, 2002
15. Vladimir V. Belyakov, Alexander I. Chumakov, Alexander Y. Nikiforov, Vyacheslav S. Pershenkov, Peter K. Skorobogatov, A.V. Sogoyan: Prediction of Local and Global Ionization Effects on ICs: The Synergy between Numerical and Physical Simulation. Russian Microelectronics, Vol. 32(2), March 2003, pp. 105–118

16. Microchip PIC16F8X 18-pin Flash/EEPROM 8-Bit Microcontrollers
http://ww1.microchip.com/downloads/en/DeviceDoc/30430c.pdf
17. Hagai Bar-El, Hamid Choukri, David Naccache, Michael Tunstall, and Claire Whelan: Workshop on Fault Detection and Tolerance in Cryptography, Florence, Italy, June 30, 2004
18. Semiconductors Research Ltd: Special equipment for semi-invasive hardware security analysis of semiconductors
http://www.semiresearch.com/inc/equipment_for_sale.html
19. Rita Mayer-Sommer: Smartly Analyzing the Simplicity and the Power of Simple Power Analysis on Smart Cards. Cryptographic Hardware and Embedded Systems (CHES-2000), LNCS Vol. 1965, Springer-Verlag, 2000, pp. 78–92
20. Sergei Skorobogatov: Semi-invasive attacks – A new approach to hardware security analysis. Technical Report UCAM-CL-TR-630, University of Cambridge, Computer Laboratory, April 2005