Optical Fault Induction Attacks

Sergei P. Skorobogatov and Ross J. Anderson

University of Cambridge, Computer Laboratory,
15 JJ Thomson Avenue, Cambridge CB3 0FD, United Kingdom
{sps32, rja14}@cl.cam.ac.uk

Abstract. We describe a new class of attacks on secure microcontrollers and smartcards. Illumination of a target transistor causes it to conduct, thereby inducing a transient fault. Such attacks are practical; they do not even require expensive laser equipment. We have carried them out using a flashgun bought second-hand from a camera store for $30 and with an $8 laser pointer. As an illustration of the power of this attack, we developed techniques to set or reset any individual bit of SRAM in a microcontroller. Unless suitable countermeasures are taken, optical probing may also be used to induce errors in cryptographic computations or protocols, and to disrupt the processor's control flow. It thus provides a powerful extension of existing glitching and fault analysis techniques. This vulnerability may pose a big problem for the industry, similar to those resulting from probing attacks in the mid-1990s and power analysis attacks in the late 1990s.

We have therefore developed a technology to block these attacks. We use self-timed dual-rail circuit design techniques whereby a logical 1 or 0 is not encoded by a high or low voltage on a single line, but by (HL) or (LH) on a pair of lines. The combination (HH) signals an alarm, which will typically reset the processor. Circuits can be designed so that single-transistor failures do not lead to security failure. This technology may also make power analysis attacks very much harder too.

1 Introduction

Secure microcontrollers and smartcards are designed to protect both the confidentiality and the integrity of sensitive information. It is not sufficient to prevent an attacker from finding out the value of a stored cryptographic key; she must also be unable to set part of the key to a known value, or to induce errors in the computation that enable sensitive information to be deduced. These errors may be data errors, such as an incorrect digital signature that leaks the value of the signing key [3], or errors in the code, such as a missed conditional jump that reduces the number of rounds in a block cipher [1]. Until now, the most widely known technique for inducing such errors was glitching – the introduction of voltage transients into the power or clock line of the target chip. Many chips are now designed to resist glitch attacks.

A review of the tamper-resistance of smartcard and secure microcontroller chips may be found in [2]. Attacks tend to be either invasive, using chip testing equipment such as probing stations and focused ion beam workstations to

B.S. Kaliski Jr. et al. (Eds.): CHES 2002, LNCS 2523, pp. 2–12, 2003.

extract data from the chip directly, or else non-invasive processes involving the exploitation of unintentional electromagnetic emissions, protocol design flaws, and other vulnerabilities that manifest themselves externally. Either type of attack may be passive or active. The standard passive invasive attack involves using microprobes to monitor a smartcard's bus while a program is executing; in an active attack, signals may be also injected, the classic example being the use of a grounded microprobe needle on the clock line to the instruction latch to disable jump instructions. A passive non-invasive attack is analyzing the electromagnetic field in the neighborhood of the device under test [10], while glitching is the classic example of an active attack.

Until now, invasive attacks involved a relatively high capital investment for lab equipment plus a moderate investment of effort for each individual chip attacked. Non-invasive attacks such as power analysis require only a moderate capital investment, plus a moderate investment of effort in designing an attack on a particular type of device; thereafter the cost per device attacked is low. Non-invasive attacks are thus particularly attractive where they exist.

Unfortunately for the attacker, many chipmakers have now implemented defenses against the most obvious non-invasive attacks. These defenses include random clock jitter to make power analysis harder, and circuits that react to glitches by resetting the processor. Meanwhile invasive attacks are becoming constantly more demanding and expensive, as feature sizes shrink and device complexity increases, We therefore set out to find new, more powerful, ways of attacking chips.

We describe our new class of attacks as 'semi-invasive'. By this, we mean that, like invasive attacks, they require depackaging the chip to get access to the chip surface. But the passivation layer of the chip remains intact – semi-invasive methods do not require electrical contact to the metal surface so there is no mechanical damage to the silicon.

Semi-invasive attacks are not entirely new. The electromagnetic analysis of [10] is best performed on a naked chip, and the old EPROM-hacking trick of exposing the write protect bit of a microcontroller to UV light usually entails depackaging it. Semi-invasive attacks could in theory be performed using such tools as UV light, X-rays, lasers, electromagnetic fields and local heating. They could be used individually or in conjunction with each other. However, this field has hardly been explored.

We will now show that extremely powerful attacks can be carried out quickly using very cheap and simple equipment.

2 Background

Once the semiconductor transistor had been invented, it was found to be more sensitive to ionizing radiation – whether caused by nuclear explosions, radioactive isotopes, X-rays or cosmic rays – than the thermionic valves (vacuum tubes) used previously. In the middle sixties, during experiments with pulsed lasers, it was found that intensive light causes some similar phenomena. Lasers started to be used to simulate the effects of ionizing radiation on semiconductors [4].

Since then the technology has been improved dramatically. Expensive inert-gas-based lasers and solid-state lasers have been replaced with low-cost semiconductor lasers. As a result, the technology has moved from the laboratory all the way down to consumer electronics.

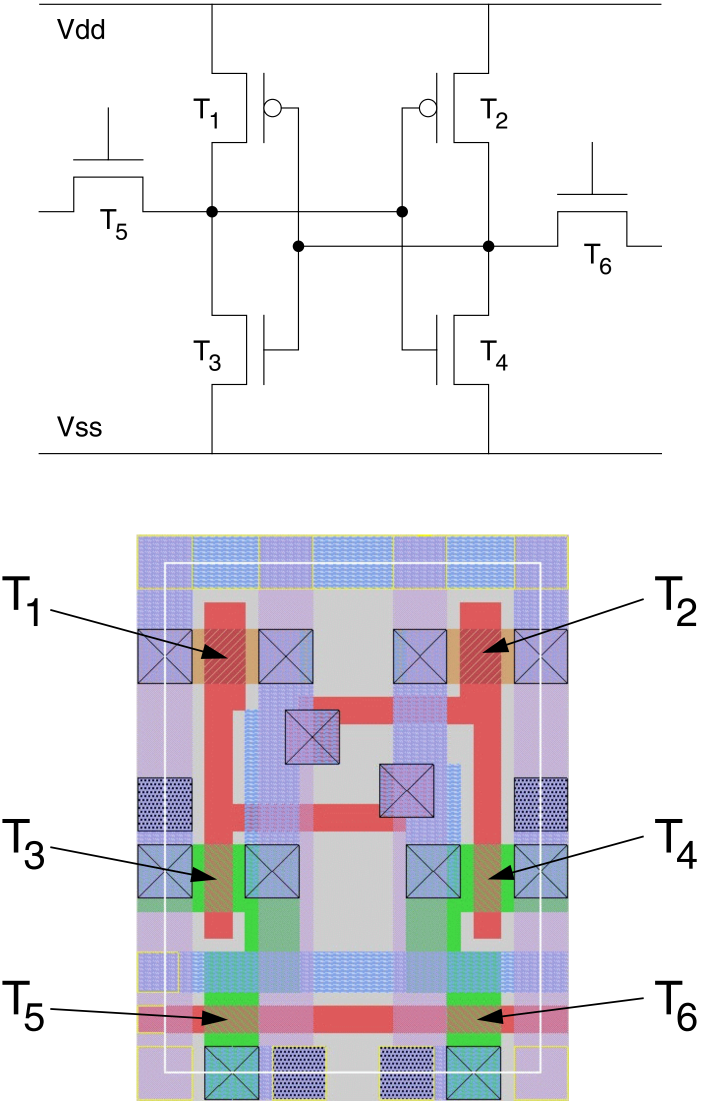

Fig. 1. Circuit structure and layout of a six-transistor SRAM cell

Laser radiation can ionize an IC's semiconductor regions if its photon energy exceeds the semiconductor band gap. Laser radiation with 1.06 μm wavelength (1.17 eV photon energy) used in [5] has a penetration depth of about 700 μm and provides good spatial ionization uniformity for silicon devices. However, its focusing is restricted by dispersion to several micrometers, and this is not precise enough for modern semiconductor devices. However, when moving from infrared to visible light, photon absorption dramatically increases [7], and it has become possible to use red and green lasers as the transistors in modern chips became thinner. Smaller devices also mean that less energy is required to achieve the same level of ionization.

In the case of CMOS devices, there is a danger of latching up the circuit, causing a short circuit that can result in permanent damage. So the use of radiation with CMOS structures must be done with appropriate precautions.

Although there are many publications about using pulsed lasers to simulate ionizing radiation, we could find no published information about using them to control or change the behavior of integrated circuits. So we decided to apply an intense light source to a semiconductor chip, and particularly to CMOS logic, to see whether it would be possible to change the state of a memory cell and how easy, or difficult, it might be.

Our first experiments targeted SRAM. The structure of a standard six-transistor SRAM cell is shown in Fig. 1 [8].

Two pairs of p- and n-channel transistors create a flip-flop, while two other n-channel transistors are used to read its state and write new values into it. The layout of the cell is shown on the right of Fig. 1 [9]. The transistors T_1 and T_3 create the CMOS inverter; together with the other similar pair, they create the flip-flop which is controlled by the transistors T_5 and T_6.

If the transistor T_3 could be opened for a very short time by an external stimulus, then it could cause the flip-flop to change state. By exposing the transistor T_4, the state of the cell would be changed to the opposite. The main difficulties we might anticipate are focusing the ionizing radiation down to several μm^2 and choosing the proper intensity.

3 Experimental Method

For our experiments we chose a common microcontroller (Microchip PIC16F84), which has 68 bytes of SRAM memory on chip (Fig. 2). A standard depackaging procedure was applied to the chip and the result of this operation is shown as well in Fig. 2.

The SRAM memory array is located in the centre of the bottom section of the chip. This area is shown with 80× magnification on Fig. 4.

Because we had a very limited equipment budget, and the laser we had appeared unsuitable, we decided to use a cheap photoflash lamp (a Vivitar 550FD, bought secondhand from a camera shop for $30). Although the luminosity of a flashlamp is much less than that of a pulsed laser, with appropriate magnification the necessary level of ionization might be achieved. We used duct tape to fix

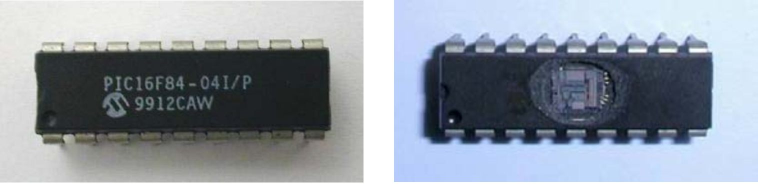

Fig. 2. Microcontroller PIC16F84 original and depackaged

the photoflash lamp on the camera port of a Wentworth Labs MP-901 manual probing station (Fig. 3). Magnification was set to the maximum – 1500×.

The microcontroller was programmed, such that its memory could be uploaded and downloaded via a serial interface connection. By filling the whole memory with constant values, exposing it to the flash light, and downloading the result, we could observe which cells changed their state. We used the TTL-level control input of the flash to remote control it from a connected PC and changing the capacitor recharge time allowed us to control the energy output. The output power of the lamp was set to the maximum possible in this experiment.

By shielding the light from the flash with an aperture made from aluminum foil, we succeeded in changing the state of only one single SRAM cell. The final state of the cell depended on the area exposed to the flash. This confirmed our intuition that it would be possible to change the contents of SRAM using a low cost semi-invasive attack.

4 Results

We found we could change any individual bit of an SRAM array. The array, under maximum magnification, is shown in Fig. 5. Focusing the light spot from the lamp on the area shown by the white circle caused the cell to change its state from 1 to 0, with no change if the state was already 0. By focusing the spot on the area shown by black circle, the cell changed its state from 0 to 1 or remained in state 1.

It can be seen from Fig. 4 that the SRAM array is divided into eight equal blocks. By exposing cells in different blocks, we found that each block corresponds to one bit plane of information. The result of this operation is shown in Fig. 6.

We built a map of the addresses corresponding to the physical location of each memory cell by exposing each cell in turn to the photoflash light. The result is presented in Fig. 7, with the left edge corresponding to the bottom side of the block. It can be seen that the addresses are not sequential, but divided into three groups.

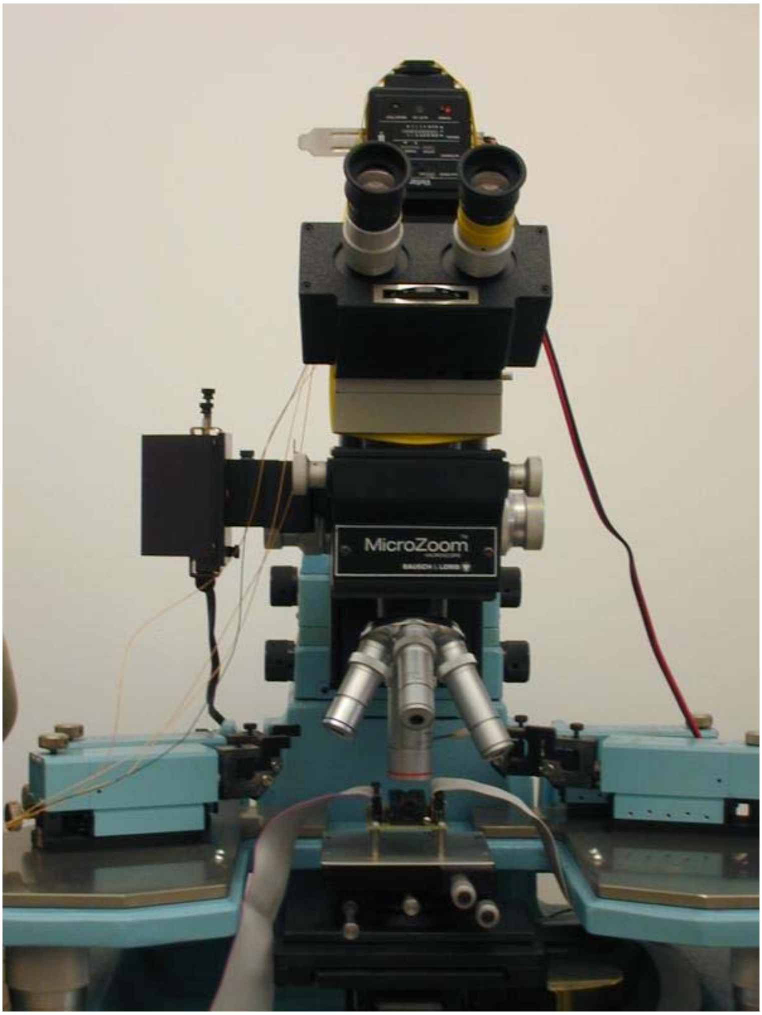

Fig. 3. Wentworth Labs MP-901 manual prober with Vivitar 550FD photoflash lamp mounted on top

This shows how simple semi-invasive attack methods can be used for reverse engineering a memory address map. The only limitation is that the flash does not produce even and monochromatic light, so it is very difficult to control the area where the spot of the light will be applied. This problem can be solved by replacing the flash with a suitable laser.

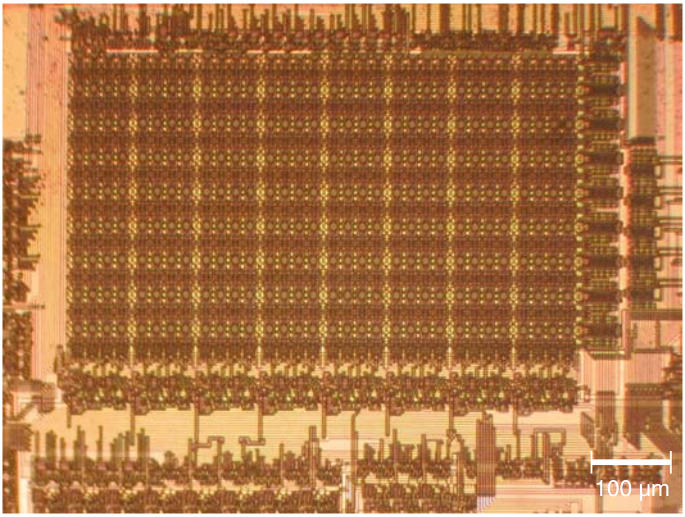

Fig. 4. SRAM memory array with magnification

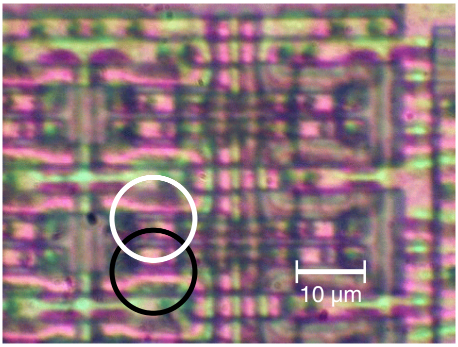

Fig. 5. SRAM memory array with maximum magnification

5 Implications and Further Work

This work shows that optical probing attacks are possible using low-cost equipment. We have repeated the experiments using a laser pointer (Fig. 8), which we bought on a local market for $8, and a motorized stage. The same results were achieved, but there were several practical differences. On the one hand that we could probe the chip surface automatically, and at a rate which we estimate could be driven as high as 100 flashes per second. On the other hand we had to be more careful with alignment because of the narrower aperture and lower power. The pointer was designed as a Class II laser device (< 1 mW), but we operated it with a supply current that should result in up to 10 mW light out-

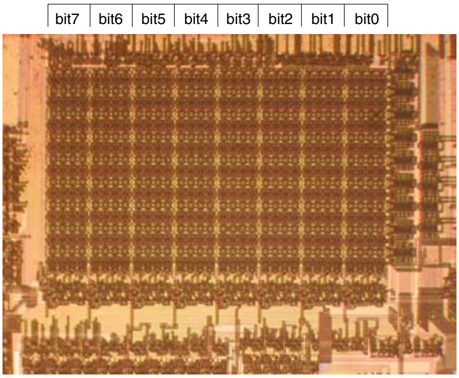

Fig. 6. Allocation of data bits in the SRAM memory array

30h	34h	38h	3Ch	40h	44h	48h	4Ch	10h	14h	18h	1Ch	20h	24h	28h	2Ch	0Ch
31h	35h	39h	3Dh	41h	45h	49h	4Dh	11h	15h	19h	1Dh	21h	25h	29h	2Dh	0Dh
32h	36h	3Ah	3Eh	42h	46h	4Ah	4Eh	12h	16h	1Ah	1Eh	22h	26h	2Ah	2Eh	0Eh
33h	37h	3Bh	3Fh	43h	47h	4Bh	4Fh	13h	17h	1Bh	1Fh	23h	27h	2Bh	2Fh	0Fh

Fig. 7. Allocation of addresses in each bit block of SRAM memory array

put. We can focus it to around 1 μm on the chip surface and its wavelength is around 650 nm.

We used our automated probing equipment to implement attacks on a number of semiconductor devices. The best designed of the modern secure microcontrollers are not vulnerable to attacks using single laser flashes, as their protection state depends on a number of bits of physical storage. However, a number of designs can be unprotected by changing the state of the flip-flop that latches the read-protect state. We strongly recommend that designers of ICs should study their designs carefully to ensure that there are no single-transistor failures that can subvert the chip's security policy.

Attack experiments have been conducted on smartcards too. It may be helpful at this point to recall some of the earlier literature on fault analysis. In [3], Boneh, Demillo and Lipton pointed out that the faulty computation of an RSA digital signature leaks the signing key. For example, when doing an RSA signature the secret computation $S = h(m)^d \pmod{pq}$ is carried out mod p, then mod q, and the results are then combined, as this is significantly faster. However, if the card returns a defective signature S_p which is correct modulo p but incorrect modulo q, then we will have $p = \gcd(pq, S_p^e - h(m))$.

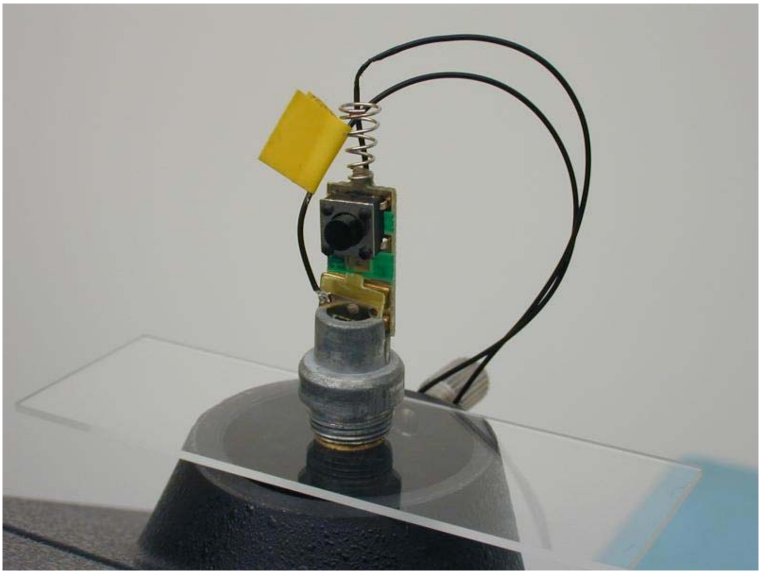

Fig. 8. Disassembled laser pointer mounted to the microscope camera port

In [1], Anderson and Kuhn pointed out that interference with jump instructions is an even more powerful and general attack: an attacker who can cause conditional branches in the smartcard code to be taken wrongly may, for example, reduce the number of rounds in a block cipher to one or two, making key recovery straightforward. The first of these two types of attack has been implemented successfully using our technique, but an NDA prevents us from giving further information.

Further scientific work in our plan includes a fuller investigation of the potential for attacks by an opponent with a moderately resourced laboratory, by which we mean a modern probing station with a multiple wavelength laser. We are commissioning such equipment and plan to use it to explore the potential for fault induction through the rear of the chip using infrared light. We have also obtained access to a suitable X-ray source and will investigate whether it can be used to induce useful faults. The significance of this is that X-rays can penetrate top-layer metal, as well as most types of protective packaging likely to be encountered in practice.

6 Countermeasures

The optical probing attack described above is a new and devastating technique for attacking smartcards and other security processors. We anticipate that, like the power analysis attacks reported by Kocher in [6], it could have a significant commercial effect on the industry, in that it will force a thorough reappraisal of security claims and the introduction of new defensive technology.

Following Kocher, we decided to delay the announcement of our attack until proper defenses were available. Existing high-end chip-defense techniques, such as top-layer metal shielding and bus encryption, may make an attack using these techniques more complicated, but are not enough. A sophisticated attacker can defeat metal shielding by using infrared light or X-rays, while bus encryption can be defeated by attacking registers directly.

The defensive technology that we have developed uses self-timed dual-rail logic. Conventional digital logic uses a clock to synchronize activities; but the cost of clocking rises as devices become more complex, and this has led to a surge of interest in design techniques for self-timed, or asynchronous, circuits – circuits that do not use clocks. Such circuits need some mechanism whereby functional components in a circuit can signal that they are ready to receive data, or are done. One way of doing this is to introduce redundancy into the data path.

In dual-rail logic, a 0 or 1 is signaled not by a low or high voltage on a single wire, but by a combination of signals on a pair of wires. For example, 0 may be 'LH' and 1 may be 'HL'. When used in self-timed circuits, 'LL' signals quiescence. The principal drawback of this simple arrangement is fragility: bugs tend to cause the emergence of the unwanted 'HH' state, which propagates rapidly throughout the circuit and locks it.

Our innovation was to turn this fragility to advantage, by making 'HH' into an error signal. This signal can be raised deliberately by tamper sensors, causing the device to lock [12]. Of more interest here is the fact that matters can be so arranged that single device failures cause are unlikely to cause the output of sensitive information [11]. We believe that such robustness will be a requirement for many high-security devices in future.

The engineering details are non-trivial. For example, an obvious concern is that almost any undetected malfunction could be exploited by the attack of Boneh et al. on RSA signatures. Colleagues have therefore developed a modular multiplication unit using our technology. Similarly, although bus encryption can remove the need to protect on-chip memory arrays, there remains the risk of attacks on the program counter and other registers. Other colleagues have therefore developed registers, and a memory management unit, that use our technology [11].

7 Conclusion

Standard CMOS circuitry is extremely vulnerable to attack using optical probing. By exposing a transistor to a laser beam, or even the focused light from a flash, it can be made to conduct. This gives rise to many effects that can be used by an attacker. We have described here how the illumination of a certain area of an SRAM cell can be used to set it to either 0 or 1. Other memory technologies, such as EPROM, EEPROM and Flash, can also be manipulated in various ways.

However, this is only the beginning. Given only moderately expensive equipment, an attacker may be able to induce a fault in a CMOS integrated circuit, in any targeted transistor, and at precisely the clock cycle of her choice. This is quite devastating. Hardware countermeasures will be necessary.

References

1. R.J. Anderson, M.G. Kuhn, "Low Cost Attacks on Tamper Resistant Devices", in M. Lomas et al. (ed.), Security Protocols, 5th International Workshop, Paris, France, April 7–9, 1997
2. R.J. Anderson, "Security Engineering – A Guide to Building Dependable Distributed Systems", Wiley 2001
3. D. Boneh, R.A. DeMillo, R.J. Lipton, "On the Importance of Checking Cryptographic Protocols for Faults, Advances in Cryptology – Eurocrypt 97", Springer LNCS vol 1233 pp 37–51
4. D.H. Habing, "Use of Laser to Simulate Radiation-induced Transients In Semiconductors and Circuits", IEEE Trans. Nuc. Sci., Vol NS-12, No 6, pp 91–100, Dec. 1965
5. A.H. Johnston, "Charge Generation and Collection in p-n Junctions Excited with Pulsed Infrared Lasers", IEEE Trans. Nuc. Sci., Vol NS-40, No 6, pp 1694–1702, 1993
6. P. Kocher, "Differential Power Analysis", Advances in Cryptology – Crypto 99, Springer LNCS vol 1666 pp 388–397
7. "Handbook of Optical Constants of Solids", edited by Edward D. Palik, Orlando: Academic Press, 1985, pp 547–569
8. J.M. Rabaey, "Digital Integrated Circuits: A Design Perspective", Prentice-Hall, 1995
9. K. Yun, "Memory", UC San Diego, Adapted from EE271 notes, Stanford University, `http://paradise.ucsd.edu/class/ece165/notes/lecC.pdf`
10. J.J. Quisquater, D. Samyde, "ElectroMagnetic Analysis (EMA): Measures and Countermeasures for Smart Cards", International Conference on Research in Smart Cards, E-smart 2001, Cannes, France, pp 200–210, Sept. 2001
11. S.W. Moore, R.J. Anderson, P. Cunningham, R. Mullins, G. Taylor, "Improving Smartcard Security using Self-Timed Circuits", Asynch 2002, proceedings published by IEEE Computer Society Press
12. S.W. Moore, R.J. Anderson, M.G. Kuhn, "Improving Smartcard Security using Self-Timed Circuit Technology", Fourth AciD-WG Workshop, Grenoble, ISBN 2-913329-44-6, 2000