Many semiconductor chips used in a wide range of applications require protection against physical attacks or tamper resistance. These attacks assume that a direct access to the chip is possible with either establishing electrical connections to signal wires or at least doing some measurements. The importance of protection against physical attacks is dictated by the amount of valuable and sensitive information stored on the chip. This could be secret data or company secrets and intellectual property (IP), electronic money for service access, or banking smartcards. The security in chips serves to deter prospective attackers from performing unauthorized access and benefiting from it. There are many areas that rely on tamper resistance of silicon chips. One of the first was car industry with theft protection and car alarms. Then in the early 1990s service providers such as PayTV, satellite TV, and utility companies realized that their service can be stolen if the access and payment cards are not properly protected. From the late 1990s home entertainment companies realized that their game consoles became the target of dishonest users who wanted to run illegal copies of the games. These days many device manufacturers from computer peripherals and mobile phones to printers and computers are worried about possible IP theft by third parties – either competitors or subcontractors. All the above challenges force hardware engineers to find secure solutions – either better protected off-the-shelf chips or their own custom chips. As in most cases it is impractical to block direct access to the device and its components, protection against physical attacks became the essential part of the system design. These days we have a continuous battle between the manufacturers who invent new security solutions learning their lessons from previous mistakes and the hacker community which is constantly trying to break the protection in various devices. Both sides are also constantly improving their knowledge and experience. In this endless war, the front line shifts forward and backward regularly. Deep down, the
problem concerns both economics and law. On the one hand, when dishonest people try to steal property, there will be a demand to increase security. On the other, reverse engineering was always part of technological progress, helping to design compatible products and improve existing ones. The dividing line between legal (reverse engineering) and illegal (piracy) is difficult.

7.1 Attack Scenarios or Why Devices are Attacked

Attacks can be used for different purposes depending on the goal. Sometimes copying a profitable on-the-market product can give easy money. Larger manufacturers could consider stealing IP from the device and mixing it with their own IP to disguise the theft. Others could try to steal secrets from the device either to produce a competitive product or to steal service. Product designers should first think about the possible motives for attacking their devices and then concentrate on the protection mechanisms. The following attack scenarios should be considered during the system design.

1. **Theft of Service** could happen when electronic devices are used to provide access to some information or service. For example, cable and satellite TV companies control the channels a viewer can see. If a pirate can bypass security or simulate the device, the service provider will lose. As the pirates normally work in a large community, any success is distributed among all members of the group, so it incurs huge losses to the service provider.

2. **Cloning and Overbuilding** are one of the most widely used attack scenarios. Cloning is used by a large variety of attackers from individuals, who want cheaper electronic gadgets, to large companies interested in increasing their sales without large investment in design. For example, dishonest competitors may try to clone existing products to reduce development costs. Of course they will have to spend some effort to disguise the fact of piracy, but compared to honest development cost this is negligible. Normally, cloning requires reverse engineering of the device to some extent. Overbuilding takes place when a contract manufacturer builds more than the requested quantity of electronic devices. The extra devices can be then sold on the market. The design could also be sold to third parties.

3. **IP Piracy** is always a big concern for many developers from private individuals to large corporations. This involves extraction of information, passwords, and cryptographic keys. This can later be used to design better product with lower investment or to read encrypted sensitive information and trade secrets.

4. **Denial of Service** can be used by a competitor to damage a vendor’s product. This could happen when the device firmware is updated over a network. If the competitor manages to reverse engineer the device and work out the update protocol, he could launch a malicious update code and then switch off all the devices or even damage them by uploading bad code. For example, it is possible to permanently damage a field-programmable gate array (FPGA)
device by uploading a bad configuration file. Also, modern microcontrollers and smartcards have Flash memory for the program code. If an erase command is issued for all memory blocks then the device will stop operating for good. The developer should design firmware update features very carefully to make sure they cannot be used without proper authentication.

7.2 Levels of Tamper Resistance

It is not an easy task to estimate the protection level of a semiconductor chip as so many factors should be taken into consideration from the chip’s package and die layout to memory structure, memory type, programming and access interfaces, security fuses or secret key location, protection mechanisms, and other security features such as glitch detection, power supply voltage monitors, protection meshes, tamper resistance, etc. There is no straightforward way to evaluate the hardware security of a semiconductor device; what normally has to be done is to apply different attack methods and observe the result. The more attacks tested, the more confidence in the result. In order to estimate the level of security protection several tamper protection levels were introduced by IBM [1]. Their classification suggests six security levels starting from a zero level corresponding to the system without any security protection to a high level for the virtually unbreakable system. There might, of course, be all sorts of intermediate levels which can be used to compare the devices with each other.

1. **Level ZERO.** No special security features are used in the system. All parts have free access and can be easily investigated. Example: microcontroller or FPGA with external memory.

2. **Level LOW.** Some security features are used but they can be relatively easy defeated with minimum tools required such as soldering iron and low cost analog oscilloscope. Attack takes time but does not involve more than $1,000 of equipment. Example: microcontroller with unprotected internal memory but proprietary programming algorithm.

3. **Level MODL.** Security used protects against most low cost attacks. More expensive tools are required as well as some special knowledge. Total equipment cost does not exceed $10,000. Examples: microcontrollers sensitive to power analysis and power glitches.

4. **Level MOD.** Special tools and equipments are required for successful attack as well as some special skills and knowledge. Total equipment cost is up to $100,000. Examples: microcontrollers with protection against ultraviolet (UV) light attacks and old smartcard chips.

5. **Level MODH.** Special attention is paid to design of the security protection. Equipment is available but is expensive to buy and operate. Total equipment cost is up to $1,000,000. Special skills and knowledge are required to utilise the equipment for an attack. A group of skilled attackers may be required
with complementary skills to work on the attack sequence. Examples: modern smartcard chips with advanced security protection, complex application-specific integrated circuits (ASICs), and secure FPGAs.

6. **Level HIGH.** All known attacks are defeated and some research by a team of specialists is necessary to find a new attack. Highly specialised equipment is necessary, some of which might have to be designed and built. Total cost of the attack is over a million dollars. The success of the attack is uncertain. Only large organizations like semiconductor manufacturers or government funded laboratories could afford carrying such attacks. Examples: secure cryptographic modules in certification authority applications.

For applications or devices that include cryptography, U.S. and Canadian federal government agencies are required to use a cryptographic products that has been FIPS 140 (Federal Information Processing Standards) validated [2] or Common Criteria validated [3]. Most Common Criteria protection profiles rely on FIPS validation for cryptographic security. Within the FIPS 140–2 (or 140–1) validations, there are four possible security levels for which a product may receive validation.

- **Security Level 1** provides the lowest level of security. It specifies basic security requirements for a cryptographic module.
- **Security Level 2** improves the physical security of a Level 1 cryptographic module by adding the requirement for tamper evident coatings or seals, or for pick-resistant locks.
- **Security Level 3** requires enhanced physical security, attempting to prevent the intruder from gaining access to critical security parameters held within the module.
- **Security Level 4** provides the highest level of security. The physical security provides an envelope of protection around the cryptographic module to detect a penetration into the device from any direction.

The security level of a particular device does not last forever. It is possible that a low cost attack will be found in the future when the attack tools become cheaper or available as second-hand. In addition, technological progress opens doors to less expensive attacks reducing the protection level of some products.

### 7.3 Attack Categories

There are several ways how semiconductor chips could be physically attacked.

- **Side-channel attacks** allow the attacker to monitor the analog characteristics of supply and interface connections and any electromagnetic radiation by the device during normal operation.

- **Software attacks** use the normal communication interface of the device and exploit security vulnerabilities found in the protocols, cryptographic algorithms, or their implementation.
**Fault generation** uses abnormal environmental conditions to generate malfunctions in the device that provide additional access.

**Microprobing** can be used to access the chip surface directly, so we can observe, manipulate, and interfere with the device.

**Reverse engineering** is used to understand the inner structure of the device and learn or emulate its functionality. It requires the use of the same technology available to semiconductor manufacturers and gives similar capabilities to the attacker.

All microprobing and reverse engineering techniques are invasive attacks. They require hours or weeks in specialized laboratory and in the process they destroy the packaging. The other three are noninvasive attacks. The attacked device is not physically harmed during these attacks. The fault attack could also be semi-invasive. It means that the access to the chip’s die is required but the attack is not penetrative and the fault is generated with intensive light pulse, radiation, local heating, or other means.

Noninvasive attacks are particularly dangerous in some applications for two reasons. Firstly, the owner of the device might not notice that the secret keys or data have been stolen, therefore it is unlikely that the validity of the compromised keys will be revoked before they are abused. Secondly, noninvasive attacks often scale well, as the necessary equipment can usually be reproduced and updated at low cost.

The design of most noninvasive attacks requires detailed knowledge of both the chip and software. On the other hand, invasive microprobing attacks require very little initial knowledge and usually work with a similar set of techniques on a wide range of products. Attacks therefore often start with invasive reverse engineering, the results of which then help to develop cheaper and faster noninvasive attacks. Semi-invasive attacks can be used to learn the device functionality and test its security circuits. As these attacks do not require establishing any physical contact to the internal chip layers, expensive equipment such as laser cutters and focused-ion beam machines are not required. The attacker could succeed using a simple off-the-shelf microscope with a photoflash or laser pointer attached to it.

Attacks can be reversible when the device can be put back into the initial state, or irreversible with permanent changes done to the device. For example, power analysis and microprobing could give the attacker a result without harming the device itself. Certainly, microprobing will leave tamper evidence but usually that does not affect further device operation. On the contrary, fault injection and UV light attacks could very likely put the device into the state where the internal registers or memory contents are changed and cannot be restored. In addition, UV attacks leave tamper evidence as they require direct access to the chip surface.

### 7.3.1 Noninvasive Attacks

A noninvasive attack does not require any initial preparations of the device under test. The attacker can either tap the wires to the device or plug it into a test circuit for the analysis. Once found, these attacks could be easily scaled and their reproduction
does not involve very much cost. In addition, no tamper evidence is left after they are applied. Therefore they are considered to be the most serious threat to the hardware security of any device. At the same time it usually takes a lot of time and effort to find an attack on any particular device. This often involves reverse engineering the device in the sense of either disassembling its software or understanding its hardware layout. Quite common electrical engineering tools can be used for non-invasive attacks. These involve integrated circuit (IC) soldering/desoldering station, digital multimeter, universal chip programmer, prototyping boards, regulated power supply, oscilloscope, logic analyzer, and signal generator.

Noninvasive attacks can be either passive or active. Passive attacks, also called side-channel attacks, do not involve any interaction with the attacked device but, usually, observation of its signals and electromagnetic emissions. Examples of such attacks are power analysis and timing attacks. Active attacks, like brute force and glitch attacks, involve playing with the signals applied to the device including the power supply line.

The most widely used noninvasive attacks include playing around with the supply voltage and clock signal. Under-voltage and over-voltage attacks could be used to disable protection circuit or force a processor to do the wrong operation. For these reasons, some security chips have a voltage detection circuit, but this circuit cannot react to fast transients. Power and clock transients can also be used in some processors to affect the decoding and execution of individual instructions.

Another possible attack uses current analysis. We can measure with an oscilloscope the fluctuations in the current consumed by the device. Drivers on the address and data bus often consist of up to a dozen parallel inverters per bit, each driving a large capacitive load. They cause a significant power-supply short circuit during any transition.

Another possible threat to secure devices is data remanence. This is the capability of volatile memory to retain information for some time after power is disconnected. Static random-access memory (SRAM) storing the same key for a long period of time can reveal it on next power on. Another possibility is to “freeze” the memory by applying low temperature. In this case, SRAM can retain information for enough time to get access to the memory chip and read its contents. Data remanence can take place in nonvolatile memories as well; the residual charge left on a floating gate transistor may be detected. For example, it could affect a threshold level or time-switching characteristics.

The next possible way of attacking a device is playing around with its interface signals and access protocols. Also, if a security protocol is wrongly implemented, that leaves a hole for the attacker to exploit. Some microcontrollers and smartcards have a factory-test interface that provides access to on-chip memory and allows the manufacturer to test the device. If an attacker can find a way of exploiting this interface, he can easily extract the information stored inside the chip. Normally information on test circuits is kept secret by the manufacturer, but an attacker can try applying different voltages and logic levels to the pins in the hope that it will put it into test mode. This sometimes works for microcontrollers but in smartcards such test circuitry is usually destroyed after use. Also, embedded software developers
sometimes implement functions that allow downloading from internal memory for test and update purposes. That must be done in a way that prevents any access to the code without proper authentication, or so that the code can be sent out in encrypted form only.

### 7.3.2 Invasive Attacks

These attacks require direct access to the internal components of the device. If it is a security module or a USB dongle then it has to be opened to get access to the internal memory chips. In the case of a smartcard or a microcontroller, the packaging should be removed followed by focused ion beam (FIB) or laser depassivation to get access to the internal wires buried deep under the passivation layer of the chip. Such attacks normally require a well-equipped and knowledgeable attacker to succeed. Meanwhile, invasive attacks are becoming constantly more demanding and expensive, as feature sizes shrink and device complexity increases.

Some operations such as depackaging and chemical etching can still be performed by almost anyone with a small investment and minimal knowledge. There are also some attacks, for example optical reading of an old Mask ROM (read-only memory), or reverse engineering of a chip built with old technology and two metal layers, where gaining the access to the chip surface is enough to succeed. Despite the greater complexity of invasive attacks, some of them can be done without expensive laboratory equipment. Low-budget attackers are likely to get a cheap solution on the second-hand market for semiconductor test equipment. With patience and skill, it should not be too difficult to assemble all the required tools for under $10,000 by buying a second-hand microscope and using self-designed micropositioners.

Invasive attacks start with the removal of the chip package. Once the chip is opened it is possible to perform probing or modifying attacks. The most important tool for invasive attacks is a microprobing workstation. Its major component is a special optical microscope with a long working distance objective lens. Micropositioners are installed on a stable platform around the chip test socket and allow the movement of probe arms, with submicron precision, over a chip surface. A probing needle with an elastic hair at the end is installed on each arm and allows electrical contact to on-chip bus lines without damaging them.

On the depackaged chip, the top-layer aluminum interconnect lines are still covered by a passivation layer (usually silicon oxide or nitride), which protects the chip from the environment and ion migration. This passivation layer must be removed before the probes can establish contact. The most convenient depassivation technique is the use of a laser cutter. Carefully dosed laser flashes remove patches of the passivation layer. The resulting hole in the passivation layer can be made so small that only a single bus line is exposed. This prevents accidental contacts with neighboring lines and the hole also stabilizes the position of the probe, making it less sensitive to vibration and temperature changes. It is not usually practical to read the information stored on a security processor directly out of each single memory
cell, except for ROM. The stored data has to be accessed via the memory bus where all data is available at a single location. Microprobing is used to observe the entire bus and record the values in memory as they are accessed.

In order to read all memory cells without the help of the device software, we have to abuse a central processor unit (CPU) component such as an address counter to access memory for us. The program counter is already incremented automatically during every instruction cycle and used to read the next address, which makes it perfectly suited to serve us as an address sequence generator. We only have to prevent the processor from executing jump, call, or return instructions, which would disturb the program counter in its normal read sequence. Tiny modifications of the instruction decoder or program counter circuit, which can easily be performed by opening the right metal interconnect with a laser, often have the desired effect.

Another approach to understanding how a device works is to reverse engineer it. The first step is to create a map of the chip. It could be done by using an optical microscope with a digital camera to produce several meter large mosaics of high-resolution photographs of the chip surface. Basic architecture structures, such as data and address bus lines, can be identified quite quickly by studying connectivity patterns and by tracing metal lines that cross clearly visible module boundaries like ROM, SRAM, electrically erasable programmable ROM (EEPROM), arithmetic logic unit (ALU), and instruction decoder. All processing modules are usually connected to the main bus via easily recognizable latches and bus drivers. The attacker obviously has to be familiar with complementary metal-oxide-semiconductor (CMOS) IC design techniques and microcontroller architectures, but the necessary knowledge is easily available from numerous textbooks.

Most currently available microcontrollers and smartcard chips have feature sizes of $0.13–0.35$ μm and two to seven metal layers. Chips down to $0.25$ μm can be reverse-engineered and observed with manual and optical techniques, but require some specific deprocessing operations to remove one metal layer after another. For the latest generations of microcontrollers with more metal layers and features below the wavelength of visible light, it may be necessary to use more expensive tools such as scanning electron microscopes (SEM).

The most common tool used for failure analysis and to apply any modifications to the chip structure is a FIB machine. It consists of a vacuum chamber with a particle gun, comparable to a SEM. With a FIB machine the attacker can cut the metal and polysilicon interconnections and build new ones with a deep submicron precision. Using laser interferometer stages, a FIB operator can navigate blindly on a chip surface. Chips can also be polished from the rear side down to a thickness of just a few tens of micrometers. Using laser interferometer navigation or infrared (IR) imaging, it is then possible to locate individual transistors and contact them through the silicon substrate by FIB editing a suitable hole. This rear-access technique has probably not yet been used by pirates so far, but the technique is about to become much more commonly available and, therefore, has to be taken into account by designers of new secure chips. FIBs are primarily used by attackers today to simplify manual probing of deep metal and polysilicon lines. A hole is drilled to the signal line of interest and then filled with platinum to bring the signal to the surface,
where a several micrometer large probing pad is created to allow easy access. Modern FIB workstations cost less than a million dollars and are available in over a hundred organizations including universities. Some old FIB models are available on a second-hand market at a price of less than $100,000.

7.3.3 Semi-Invasive Attacks

There is a large gap between previously discussed noninvasive and invasive types of attack and many attacks fall into this gap, being not so expensive as classical penetrative invasive attacks but as easily repeatable as noninvasive attacks. Therefore, a new class of attack called semi-invasive was recently defined and introduced [4]. Like invasive attacks, they require depackaging the chip in order to get access to its surface. However, the passivation layer of the chip remains intact, as semi-invasive methods do not require creating contacts to the internal lines. This is because microprobing is not used for this attack technology and thus such expensive tools as FIBs are not required. Instead, less expensive laser microscopes can be used which can be made from a standard optical microscope by attaching a laser pointer to it.

Semi-invasive attacks are not entirely new. UV light has been used to disable security fuses in EPROM and one-time programmable (OTP) microcontrollers for many years. Modern microcontrollers are less susceptible to this attack as they were designed to withstand it. Advanced imaging techniques can be considered as semi-invasive as well. This includes various kinds of microscopy such as IR, laser scanning, and thermal imaging. Some of them can be applied from the rear side of the chip which is very useful for modern chips with multiple-metal-layer design. Some of these techniques allow observation of the state of each individual transistor inside the chip.

One of the main contributions to semi-invasive attacks is optical fault injection which can be used to modify the contents of SRAM and change the state of any individual transistor inside the chip. That gives almost unlimited capabilities to the attacker in getting control over the chip operation and abusing the protection mechanism.

Compared to noninvasive attacks, semi-invasive attacks are harder to implement as they require decapsulation of the chip. However, very much less expensive equipment is needed than for invasive attacks. These attacks can be performed in a reasonably short period of time. Also they are scalable to a certain extent, and the skills and knowledge required to perform them can be easily and quickly acquired. Some of these attacks, such as an exhaustive search for a security fuse, can be automated. If compared to invasive attacks, the semi-invasive kind do not normally require precise positioning for success because they are normally applied to a whole transistor or even a group of transistors rather than to a single wire inside the chip.
7.4 Breaking the Security with Noninvasive Attacks

7.4.1 Side-Channel Attacks

Some security-related operations a semiconductor chip performs can take a different time to complete depending on the values of the input data and the secret key. Careful timing measurement and analysis may allow recovery of the system’s secret key. This idea was first published in the scientific literature in 1996 [5]. Then later these attacks were successfully performed on an actual smartcard implementation of the RSA signature [6].

To conduct the attack one needs to collect a set of messages, together with their processing time, e.g., question-answer delay. Many cryptographic algorithms were found to be vulnerable to timing attacks. The main reason why this happens is in the software implementation of each algorithm. That includes performance optimization to bypass unnecessary branching and conditional operations, cache memory usage, nonfixed time processor instructions such as multiplication and division, and a wide variety of other causes. As a result, performance characteristics typically depend on both the encryption key and the input data.

To prevent such attacks the techniques used for blinding signatures can be used [7]. The general idea is to prevent the attacker knowing the input to the modular exponentiation operation by mixing the input with a chosen random value.

Timing attacks can be applied to chips whose security protection is based on passwords, or to access control systems that use cards or keys with fixed serial numbers, for example, Dallas iButton products. The common mistake in such systems is the way the serial number of the entered key is verified against the database. Very often the system checks each byte of the key against one entry in the database and stops as soon as an incorrect byte is found. Then it switches to the next entry in the database until it reaches the end. So the attacker can easily measure the time between the input of the last key and the request for another key and figure out how many coincidences were found. With a relatively small number of attempts, he will be able to find one of the matching keys.

To prevent these attacks, the designer should carefully calculate the number of CPU cycles that take place when the password is compared and make sure they are the same for correct and incorrect passwords. For example, in the Motorola 68HC08 microcontrollers family the internal ROM bootloader allows access to the internal memory only if the correct eight-byte password was entered first. To achieve that, extra NOP commands were added to the program making the processing time equal for both correct and incorrect bytes of the password. That gives good protection against timing attacks. In the early versions of the Texas Instruments MSP430 microcontrollers such compensation was not made. As a result it was possible to guess the access password to the user Flash memory [8].

Some microcontrollers have an internal resistor-capacitor (RC) generator mode of operation in which the CPU running frequency depends upon the power supply voltage and the die temperature. This makes timing analysis more difficult as the
attacker has to stabilize the device temperature and reduce any fluctuations and noise on the power supply line. Some smartcards have an internally randomized clock signal to make measurements of the time delays useless for the attack.

A computing device’s power consumption depends on its current activity. The consumption depends on changes of state of its components, rather than on the states themselves, because of the nature of CMOS transistors. When an input voltage is applied to a CMOS inverter, a transient short-circuit is induced. The rise of the current during this transient is much higher than the static dissipation caused by parasitic leakage current. Using a resistor in the power supply line, these current fluctuations can be measured. In electromagnetic analysis (EMA) a small coil placed close to the chip is used to acquire electromagnetic emission. As both the power analysis and the EMA were covered in previous chapters they will not be discussed in details here.

### 7.4.1.1 Brute Force Attacks

‘Brute force’ has different meanings for cryptography and semiconductor hardware. In cryptography, a brute force attack would be defined as the methodical application of a large set of trials for a key to the system. This is usually done with a computer or an array of FPGAs delivering patterns at high speed and looking for success.

One example could be the password protection scheme used in microcontrollers, such as the Texas Instruments MSP430 family. The password itself is 32 bytes (256 bits) long which is more than enough to withstand direct brute force attack. But the password is allocated at the same memory addresses as the CPU interrupt vectors. That, firstly, reduces the area of search as the vectors always point to even addresses within memory. Secondly, when the software gets updated, only a small part of the password is changed because most of the interrupt subroutines pointed to by the vectors are very likely to stay at the same addresses. As a result, if the attacker knows one of the previous passwords he could easily do a systematic search and find the correct password in a reasonable amount of time.

Brute force can be also applied to a hardware design implemented into an ASIC or a FPGA. In this case the attacker tries to apply all possible logic combinations to the input of the device while observing all its outputs. This kind of attack could be also called black-box analysis because the attacker does not have to know anything about the design of the device under test. He only tries to understand the function of the device by trying all possible combinations of signals. This approach works well only for relatively small logic devices. Another problem the attacker will face is that designs implemented in FPGAs or ASICs have flip-flops, so the output will probably be function of both the previous state and the input. But the search space can be significantly reduced if the signals are observed and analyzed beforehand. For example, clock inputs, data buses, and some control signals could be easily identified, significantly reducing the area of search.

Another possible brute force attack, applicable to many semiconductor chips, is applying an external high voltage signal (normally twice the power supply) to the
chip’s pins to find out whether one of them has any transaction like entering into a factory test or programming mode. In fact, such pins can be easily found with a digital multimeter because they do not have a protection diode to the power supply line. Once sensitivity to a high voltage is found for any pin, the attacker can try a systematic search on possible combinations of logic signals applied to other pins to figure out which of them are used for the test/programming mode and exploit this opportunity.

The attack could be also applied to the device communication protocol in order to find any hidden functions embedded by the software developer for testing and upgrade purposes. Chip manufacturers very often embed hardware test interfaces for postproduction testing of their semiconductor devices. If the security protection for these interfaces is not properly designed, the attacker can exploit it to get access to the on-chip memory. In smartcards such test interfaces are normally located outside the chip circuit and physically removed after the test operation, eliminating any possibility of use by outsiders.

7.4.1.2 Fault Injection Attacks

Glitch attacks are fast changes in the signals supplied to the device and designed to affect its normal operation. Usually glitches are inserted in power supply and clock signals. Every transistor and its connection paths act like an RC element with a characteristic time delay. The maximum usable clock frequency of a processor is determined by the maximum delay among its elements. Similarly, every flip-flop has a characteristic time window (of a few picoseconds) during which it samples its input voltage and changes its output accordingly. This window can be anywhere inside the specified setup cycle of the flip-flop, but is quite fixed for an individual device at a given voltage and temperature. So if we apply a clock glitch (a clock pulse much shorter than normal) or a power glitch (a rapid transient in supply voltage) this will affect only some transistors in the chip and cause one or more flip-flops to adopt the wrong state. By varying the parameters, the CPU can be made to execute a number of completely different wrong instructions, sometimes including instructions that are not even supported by the microcode. Although we do not know in advance which glitch will cause which wrong instruction in which chip, it can be fairly simple to conduct a systematic search. For example, the bootloader in the Motorola MC68HC05B microcontroller checks the bit0 of first EEPROM address and grants external access to the chip’s memory only if it is “1”, otherwise it goes into endless loop. If the chip supply voltage is temporarily reduced, the CPU fetches an FFh value from the EEPROM rather than the actual value and this corresponds to the unsecured state of the fuse. Similarly, if the clock source is disturbed by short circuiting the external crystal resonator, multiple clock glitches are produced. That way there is a good chance of escaping from the endless loop into the memory access code.

Applying clock glitches to some microcontrollers could be difficult. For example, the Texas Instruments MSP430 microcontroller family operates from an internal RC
generator in bootloader mode and it is difficult to synchronize to the internal clock and estimate the exact time of the attack. Some smartcards benefit from having randomly inserted delays in the CPU instruction flow, which makes applying the attacks even more difficult. Using power analysis could help, but requires very sophisticated and expensive equipment to extract the reference signal in real time.

Power supply voltage fluctuations can shift the threshold level of the transistors. As a result some flip-flops will sample their input at different time or the state of the security fuse will be read incorrectly. This is usually achieved by either increasing the power supply voltage or dropping it for a short period of time, normally from one to ten clock cycles. Power glitches can be applied to a microcontroller with any programming interface as they could affect both the CPU operation and the hardware security circuit. In general, they are harder to find and exploit than clock glitches because in addition to the timing parameters, the amplitude and rising/falling times are variables.

A glitch could be an external electric field transient or an electro-magnetic pulse. One approach is using a miniature inductor consisting of several hundred turns of fine wire around the tip of a microprobe needle. A current injected into this coil will create a magnetic field, and the needle will concentrate the field lines [9].

7.4.1.3 Data Remanence

Security processors typically store secret key material in SRAM, from which power is removed if the device is tampered with. It is widely known that, at temperatures below –20°C, the contents of SRAM can be “frozen”; therefore, many devices treat temperatures below this threshold as tampering events. However, it was showed that the conventional wisdom no longer holds and that data remanence can be a problem even at higher temperatures [10].

Security engineers are interested in the period of time for which an SRAM device will retain data once the power has been removed. The reason for this is as follows. Many products perform cryptographic and other security-related computations using secret keys or other variables that the equipment’s operator must not be able to read out or alter. The usual solution is for the secret data to be kept in volatile memory inside a tamper-sensing enclosure. On detection of a tampering event, the volatile memory chips are powered down or even shorted to ground. If the data retention time exceeds the time required by an opponent to open the device and power up the memory, then the protection mechanisms can be defeated.

Data remanence affects not only SRAM but other memory types as well, like DRAM, UV EPROM, EEPROM and Flash [11]. As a result, some information still can be extracted from memory that has been erased. This could create many problems for secure devices which assume that all the sensitive information is gone once the memory is erased.

Unlike SRAM which has only two stable logic states, EPROM, EEPROM and Flash cells actually store analog values in the form of a charge on the floating gate of a metal-oxide-semiconductor (MOS) transistor. The floating-gate charge shifts
the threshold voltage of the cell transistor and this is detected with a sense amplifier when the cell is read. The maximum charge the floating gate can accumulate varies from one technology to another and normally is between $10^3$ and $10^5$ electrons. The amount of trapped charge can be detected by measuring the gate-induced drain leakage current of the cell or its effect can be observed indirectly by measuring the threshold voltage of the cell. In older devices, which had the reference voltage for the sense amplifier tied to the device supply voltage, it was often possible to do this by varying the device supply voltage. For example, all information can be successfully extracted from the Microchip PIC16F84A microcontroller after it has been erased several times (Fig. 7.1). In newer devices, it is necessary to change the parameters of the reference cell used in the read process, either by re-wiring portions of the cell circuitry or by using undocumented test modes built into the device by chip manufacturers.

The changes in the cell threshold voltage caused by write/erase cycles are particularly apparent in virgin and freshly-programmed cells. It is possible to differentiate between programmed-and-erased and never-programmed cells, especially if the cells have only been programmed and erased once, since virgin cell characteristics will differ from the erased cell characteristics. The changes become less noticeable after ten program/erase cycles.

7.5 Breaking the Security with Invasive Attacks

Invasive attacks start with partial or full removal of the chip package in order to expose the silicon die. There are several methods, depending upon the package type and the requirements for further analysis. For microcontrollers and smartcards, partial decapsulation is normally used, so that the device can be placed in a standard programmer unit and tested. Some devices cannot be decapsulated and still maintain their electrical integrity. In this case the chip die has to be bonded to a chip carrier.
using a bonding machine which connects to the bonding pads on the die with thin aluminum or gold wire. Such bonding machines are available from different manufacturers and can be bought second-hand for less than $10,000. The contacts to the die can be also established using microprobing needles on a probing station.

It is a common opinion that decapsulation is a complicated process which requires a lot of experience. In fact it is not and anyone capable of carrying out chemical or biological work in the context of a standard high-school program can do this. All the necessary experience could be obtained by decapsulating a dozen different samples. Some precautions should be taken as the acids used in this process are very corrosive and dangerous; ideally, the work should be performed in a fume cupboard to prevent inhalation of the fumes from acids and solvents. Eyes should be protected with safety goggles and appropriate acid-resistant gloves should be worn as the acid will cause severe burns if it accidentally comes into contact with the skin. Protective clothing should be worn as well.

The process of manual decapsulation usually starts with milling a hole in the package so that the acid will affect only the desired area above the chip die (Fig. 7.2). The tools necessary for this operation are available from any do it yourself (DIY) shop for less than $20. The commonly used etching agent for plastic packages is fuming nitric acid (>95%), which is a solution of nitrogen dioxide in concentrated nitric acid. It is very strong nitrifying and oxidizing agent; it causes plastic to carbonize, and it also affects copper and silver in the chip carrier island and pins. Sometime a mixture of fuming nitric acid and concentrated sulphuric acid is used. This speeds up the reaction with some types of packages and also prevents the silver used in bonding pads and chip carrier from reacting. The acid is normally applied in small portions with a pipette into a premilled hole in a chip preheated to 50–70°C (Fig. 7.2). After 10–30 seconds the chip is sprayed with dry acetone from a washing bottle to remove the reaction products. This process has to be repeated several times until the die is sufficiently exposed. To speed up the process, the chip can be placed in a sand bath and the acid can be preheated in a glass beaker.
The acid residues can be removed from the etched plastic and from the chip surface by ultrasonic treatment. For that the chip is placed into a beaker with acetone and then put in an ultrasonic bath for 1–3 minutes. After washing the chip with acetone and drying it in an air jet, we have a clean and fully operational chip (Fig. 7.3).

A very similar approach can be used for decapsulating chips from the rear side. The only obstacle is the copper plate under the chip die which reacts slowly with the fuming nitric acid. That could create problems if the automatic decapsulator is used because the surrounding plastic will be etched away before this copper plate and the chip leads are very likely to be damaged. However, access to the rear side of the die can be established without using chemical etching. The chip package can be milled down to the copper plate which is then removed mechanically. The residues of the glue used to attach the die to the plate can be removed with solvents or by scraping it off with a wooden toothpick stick.

The same partial decapsulation technique can be used for smartcards as well (Fig. 7.3), although not all of them would maintain electrical integrity. Very often the chip has to be decapsulated completely and then bonded onto a chip carrier.

### 7.5.1 Deprocessing

The opposite process to chip fabrication is called deprocessing. A standard CMOS chip has many layers. The deepest doping layers inside the substrate form the transistors. The gate oxide layer isolates the gate from the active area of the transistors. The polysilicon layer on top of it forms the gates and interconnections. The interlayer oxide isolates conducting layers. Metal layers, usually made of aluminum, form the signal wires, and they are connected with other layers through “via” plugs. Finally, a passivation layer made out of silicon oxide or nitride protects the whole structure from moisture and air which could harm the die. In plastic packages the passivation layer is covered with a polymer layer, usually polyimide, to protect against sharp grains in the compound during the package formation.
There are two main applications of deprocessing. One is to remove the passivation layer, exposing the top metal layer for microprobing attacks. Another is to gain access to the deep layers and observe the internal structure of the chip.

Three basic deprocessing methods are used: wet chemical etching, plasma etching, also known as dry etching, and mechanical polishing [12]. In chemical etching each layer is removed by specific chemicals. Its downside is its isotropic nature, i.e., uniformity in all directions. That produces unwanted undercutting. As a result, narrow metal lines will have a tendency to lift off the surface. Isotropic etching also leads to etching through holes such as vias, resulting in unwanted etching of underlaying metallization (Fig. 7.4). Plasma etching uses radicals created from gas inside a special chamber. They react with the material on the sample surface to form volatile products which are pumped out of the chamber. As the ions are accelerated in an electric field they usually hit the surface of the sample perpendicularly. The removal of material is strongly anisotropic (directional). Only the surfaces hit by the ions are removed, sides perpendicular to their paths are not touched. Mechanical polishing is performed with the use of abrasive materials. The process is time-consuming and requires special machines to maintain the planarity of the surface. From the inspection perspective, the advantages of using polishing over wet and dry etching techniques is the ability to remove layer by layer and view features in the area of interest within the same plane (Fig. 7.4). It is especially useful on multilayer interconnect processes fabricated with advanced planarization techniques.

7.5.2 Reverse Engineering

This is a technique aimed at understanding the structure of a semiconductor device and its functions. In case of an ASIC or a custom IC, that means extracting information about the location of all the transistors and interconnections. In order
to succeed, a general knowledge of IC and VLSI design is required. All the layers formed during chip fabrication are removed one-by-one in reverse order and photographed to determine the internal structure of the chip. In the end, by processing all the acquired information, a standard netlist file can be created and used to simulate the device. This is a tedious and time-consuming process, but there are some companies which do such work as a standard service [13].

When it comes to reverse engineering smartcards and microcontrollers, both structural and program-code reverse engineering are required to understand how the device works. First, the security protection needs to be understood by partial reverse engineering of the chip area associated with it. Thus if memory bus encryption was used, the hardware responsible for this should be reverse engineered. Then, finally, the internal memory contents have to be extracted and disassembled to understand device functions.

A slightly different approach is required for reverse engineering FPGAs. Even if the security protection is defeated and the attacker manages to extract the configuration bitstream file from the device, he will have to spend a substantial amount of time and effort to convert it into the logic equations and primitive blocks for further simulation and analysis.

The most important tool for reverse engineering silicon chips down to 0.18 μm feature size is an optical microscope with a digital camera to produce mosaics of high-resolution photographs of the chip surface. Not every microscope would do. As light cannot pass through the chip, the microscope should have reflected light illumination. The image should be sharp and without geometric distortion and color aberration, otherwise it will not be possible to stick all the images together. The most important parameters of the microscope are resolution and magnification. The resolution of a microscope mainly depends upon its objective lenses and is defined as the smallest distance between two points on a specimen that can still be distinguished as two separate entities. Resolution is a somewhat subjective value in microscopy because at high magnification an image may appear nonsharp but still be resolved to the maximum ability of the objective.

Layout reconstruction requires the images of all the layers inside the chip to be combined. The images are normally taken automatically using a motorized stage to move the sample and special software to combine all the images together [14]. Normally, for semiconductor chips fabricated with 0.13 μm or smaller technology, images are created using a SEM which has a resolution better than 10 nm.

### 7.5.3 Microprobing

The most important tool for invasive attacks is a microprobing station (Fig. 7.5). It consists of five elements: a microscope, stage, device test socket, micromanipulators and probe tips. The microscope must have long working distance objectives – sufficient enough to accommodate six to eight probe tips (Fig. 7.5) between the
sample and the objective lens. It should also have enough depth of focus to follow the probe tip movement. Usually the microscope has several objectives to accommodate different magnification and depths of focus. Lower magnification with greater focus depth is used for coarse location of the probe tip and higher magnification for placing the tip on a conductor wire or a test point. The chip is normally placed in a test socket that provides all the necessary signals and is controlled by a computer.

On a stable platform around the test socket, several micropositioners are installed. They allow us to move a probe tip with submicron precision. The probe tip can be either passive or active. The passive tip, for example Picoprobe T-4, can be used for both eavesdropping and injecting signals, but as it is normally connected directly to an oscilloscope, it has low impedance and high capacitance. As a result it cannot be used for probing internal signals on the chip, except for the bus lines which are usually buffered. Another application for the passive tips is making connections to the bonding pads on a fully decapsulated chip. The active tip, for example Picoprobe 12 C, has a FET amplifier close to the end of the tip. Active tips offer high bandwidth (1 GHz for Picoprobe 28) with low loading capacitance (0.02 pF for Picoprobe 18 B) and high input resistance (>100 GΩ for Picoprobe 18 B). The probe tips are made out of a tungsten wire which is sharpened to <0.1 μm at the end for probing small features.

Modern probing stations benefit from full automatic control over the microscope, stage and micropositioners. For simple applications, a manually controlled probing station is enough and can be bought second-hand for less than $10,000. Passive probe tips are very cheap (less than $5 each) but active probes are quite expensive – over $60 for the tip plus over $1,000 for the tip holder with amplifier. However, they can easily be built from a $2 operational amplifier and a passive tip (Fig. 7.5).
7.6 Breaking the Security with Semi-Invasive Attacks

7.6.1 UV Attacks

These are among the oldest attacks used on microcontrollers since their release in the mid 70’s. UV attacks were often considered invasive attacks before. But as most of them require only decapsulation of the chip, they certainly belong to the class of semi-invasive attacks. These attacks can be applied to many OTP and UV EPROM microcontrollers as their protection is designed to withstand low cost noninvasive attacks only.

The attack can be divided into two stages – finding the fuse and resetting it to the unprotected state with a UV light. As the security fuse is normally designed such that it cannot be erased earlier than the program memory, the UV light cannot be applied to the whole chip. Either the memory must be protected with opaque material, or the UV light can be applied to the fuse selectively by using a microscope or a UV laser.

As well as EPROM memory, many floating-gate memory devices are also susceptible to UV attacks [15]. Meantime, chip designers have more freedom in choosing different protections against such attacks. As the EEPROM and Flash cells can change their state in both directions, the obvious thing to do is to use an erased state of the cell to indicate the alarm state and a programmed state to correspond to disabled security. Minimal changes to the control logic will do the job. This is widely used in Flash microcontrollers from many manufacturers.

7.6.2 Advanced Imaging Techniques

Visual observation under a microscope is the first step in semiconductor analysis. As feature sizes of transistors shrink each year, structures on the chip surface become more and more difficult to observe. Down to 0.8 μm technology, it was possible to identify all the major elements of microcontrollers – ROM, EEPROM, SRAM, CPU and even instruction decoder and registers within the CPU. On chips built using 0.5 μm or 0.35 μm processes, one can hardly distinguish ROM, Flash and SRAM, whereas on chips with 0.25 μm or smaller transistors, almost nothing can be seen. This is caused not only by the small feature sizes, but most of all by multiple metal layers covering the chip surface (up to eight on modern 0.13 μm chips). In addition, planarization technology often involves filling blank spaces on metal layers with metal pads which block the optical path as well.

One approach is to use IR light, either reflected or transmitted, and observe the chip from its rear side. Silicon is almost transparent to photons with wavelengths >1100 nm. However, highly doped silicon wafers used in some modern chips, are less transparent to IR light and more intensive light source or an IR camera with higher sensitivity is required. Backside imaging is widely used in failure analysis
tasks, from locating the failures in transistors or interconnections to navigation during a FIB work. There are special microscopes designed for such applications, for example the BEAMS V-2000 from Hypervision. Needless to say, such systems cost a tremendous amount of money and can only be afforded by relatively large companies. Yet, low budget laboratories could use NIR extended microscopes with IR-sensitive video cameras.

Laser radiation can ionize an IC’s semiconductor regions if its photon energy exceeds the semiconductor band gap ($\geq 1.1$ eV or $<1100$ nm). Laser radiation with 1.06 μm wavelength (1.17 eV photon energy) has a penetration depth of about 700 μm and provides good spatial ionization uniformity for silicon devices. In active photon probing, a scanned photon beam interacts with an IC. Photons with energies greater than the band gap of silicon generate electron-hole pairs in the semiconductor. Photons with lower energies can still interact with p-n junctions, but with only a heating effect taking place, which is significantly weaker than the photovoltaic effect.

There are different scanning techniques used for photon probing in failure analysis [12]. As a photon source they normally use a laser scanning microscope. Although such microscopes have a big advantage of fast scanning – about one frame per second, their price is too high for small research labs. Usually, less expensive but much slower approach of a stationary laser source and a sample moved on an X-Y motorized stage can be used.

There are two major laser scanning techniques which can be used for hardware security analysis. One is called optical beam induced current (OBIC) and is applied to an unbiased chip to find the active doped areas on its surface [16]. Another, called light-induced voltage alteration (LIVA), applied to a chip under operation [17]. In OBIC, photocurrents are used directly to produce the image. For that, the analyzed chip is arranged with its power supply pin connected to a current amplifier and the values are registered on the computer via an acquisition board. The active areas can be seen as they produce higher current, but most of the chip is covered with metal layers which the laser cannot penetrate, so these areas do not produce any current (Fig. 7.6). In LIVA, images are produced by monitoring the voltage changes of the constant current power supply as the optical beam is scanned across the IC surface. It can be seen that memory cells have different states: where the cell holds a “1” the top is brighter, and where it is a “0” the bottom is (Fig. 7.6).

### 7.6.3 Optical Fault Injection

It is a new class of attacks on secure microcontrollers and smartcards [18]. 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 and can be carried them out using a $5$ laser pointer. For example, this can be used to set or reset any individual bit of SRAM in a microcontroller. Unless suitable countermeasures are taken, optical fault injection 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.

A standard SRAM cell consists of six transistors. 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 Fig. 7.7. The transistors VT1 and VT2 create the CMOS inverter; together with the other similar pair, they create the flip-flop which is controlled by the transistors VT3 and VT6.

If the transistor VT1 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 VT4, the state of the cell would be changed to the opposite value. The main difficulties we might anticipate are focusing the ionizing radiation down to several micrometers spot and choosing the proper intensity. In the original experiments the Microchip
PIC16F84 microcontroller with 68 bytes of on-chip SRAM memory was used [18]. The light from a photoflash lamp was focused using the microscope optics. By shielding the light from the flash with an aperture made from aluminum foil the state of only one cell can be changed. The array, under maximum magnification, is shown in Fig. 7.7. 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 the black circle, the cell changed its state from “0” to “1” or remained in state “1.”

EPROM, EEPROM and Flash memory cells are even more sensitive to fault injection attacks. This happens because the currents flowing inside the floating gate cell are an order of magnitude smaller that inside the SRAM cell. There are some new optical fault injection attack techniques introduced recently. One is local heating attacks [19] which use lasers 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. Another is bumping attacks [20] aimed at data extraction from secure embedded memory, which usually stores critical parts of algorithms, sensitive data and cryptographic keys. As a security measure, read-back access to the memory is not implemented leaving only authentication and verification options for integrity check. Verification is usually performed on relatively large blocks of data, making brute force searching infeasible. By attacking the security in three stoffs, the search space can be reduced from infeasible $2^{100}$ to affordable $2^{15}$ guesses per block of data. This progress was achieved by finding a way to preset certain bits in the data path to a known state using optical fault injection.

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 IR light or X-rays, while bus encryption can be defeated by attacking registers directly.

### 7.6.4 Optical Side-Channel Analysis

Transistors emit photons when they switch. This has been well known for decades and is actively used in failure analysis. So far, observation of such emissions was associated with sophisticated and expensive equipment, because only a very limited number of photons emitted per every switch – usually $10^{-2}$ to $10^{-4}$. The peak of emission is in the near-infrared spectrum (900 to 1200 nm) and this poses restrictions on sensors selection. The emission comes from an area close to the drain and primarily from the n-channel MOS transistor. Optical emission significantly increases at higher power supply voltages [21].
Optical emission has good correlation with power analysis and can be used for characterization of leaking areas for later improvement of protection against power analysis attacks. The results, presented in Fig. 7.8, reveal that optical emission has higher bandwidth and thus data appearing at different times can be separated for further analysis.

Modern low-cost charge-coupled device (CCD) cameras are adequate for detecting photons emitted by modern CMOS circuits. Photomultiplier tubes (PMT) are very fast, but they have limited sensitivity in the IR region. Monochrome CCD cameras have good IR sensitivity and low dark current, which is important with long exposure times.

A standard microscope setup with a CCD camera mounted on top and a sample in a test socket can be used for analysis. Hobbyist astronomical cameras with active cooling appeared to be best suited for low-cost optical emission analysis by having good IR sensitivity and extremely low dark current. The emission acquired from a microcontroller using a $2\times$ objective lens is presented in Fig. 7.9. A closer look with a $10\times$ objective revealed that the value of data read from the internal memory is clearly visible (Fig. 7.10).
Modern deep-submicron chips emit photons as well. However, the front-side approach no longer works due to multiple metal layers which completely block the emission. For chips built with 0.35 μm and smaller technology, a backside approach is required. In order to achieve reasonable emission, the power supply voltage must be increased by 30–50%. The example of optical emission acquired from the backside of 130 nm FPGA chip above the SRAM area is presented in Fig. 7.11.

Optical emission analysis can lead to possible data extraction from semiconductor chips. That way, the security can be compromised in various chips from microcontrollers and smartcards to FPGAs and ASICs. Possible countermeasures include asynchronous designs and employing data encryption.

### 7.6.5 Optically Enhanced Position-Locked Power Analysis

This is a great example of combining noninvasive and semi-invasive attack methods for better result [22]. Optical enhancement of power analysis is a new and innovative technique that allows the current through an individual transistor to become visible
in the IC power trace. In conventional power analysis, power consumption is measured for a whole chip rather than on 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. Also, the power fluctuations are affected by the number of bits being set or reset (Hamming weight of data), rather than the actual value of the manipulated data.

By focusing a laser on a specific area on the chip surface, it is possible to monitor the logic state of an individual transistor, as well as the activity of a particular memory cell. This is highly useful for security analysis, allowing faster and less expensive solutions.

When a laser is focused on the VT1 transistor of the PIC16F84 SRAM memory cell (Fig. 7.7) and writing into this cell happens, an about 0.4 mA change in the power trace can be observed (Fig. 7.12). For comparison, the conventional power analysis results showed a similar change in the power consumption for a single-bit difference in the memory contents (Fig. 7.12). However, the same technique applied to the memory cells being read, did not produce any noticeable results, unless many power traces were averaged. This is because writing into an SRAM cell causes a significantly larger current response than a read operation with low laser-injected current.

By focusing the laser on the region between the VT1 and VT4 transistors, significantly higher changes in the power trace can be observed (Fig. 7.13), and what is more important, for both write and read operations (Fig. 7.13), thus allowing access detection. This happens because the timing characteristic of the cell changes when both inverters are influenced with a laser.

Interesting results can be achieved with the laser focused on the area between the cell-select transistors (VT3 and VT6). In this case, access to any of the memory cells inside the memory array column will produce a very noticeable difference in the power consumption. The same response can be obtained when a laser with higher power is focused between VT1 and VT4. These approaches can be used for access-event triggering.
Modern chips normally have multiple metal layers over their active areas, preventing direct access with a laser beam. Accessing the chip from its rear side can circumvent this obstacle and the power-trace difference is very similar to the one from the front side.

7.7 Countermeasures Against Physical Attacks

Since the late 1990s, smartcard vendors have made invasive attacks much more difficult. Smartcards typically have a sensor mesh implemented in the top metal layer, consisting of a serpentine pattern of sensor, ground and power lines. All paths in this mesh are continuously monitored for interruptions and short circuits, and cause reset or zeroing of the EEPROM memory if alarmed. In some recent smartcards further protection against microprobing attacks is used such as EEPROM data memory bus encryption. Even if the attacker manages to pick up the signals from the data bus he will not be able to recover passwords, secret keys or other sensitive information from it. This protection was aimed at preventing invasive and semi-invasive attacks. At the same time noninvasive attacks could still be successful as the CPU normally has full access control to unencrypted information.

Another improvement worth mentioning is moving from the standard building-block structures like CPU instruction decoder, register file, ALU and I/O circuits, to a complete ASIC-like logic design. This design approach is called “glue logic” and it is widely used in smartcards. Glue logic makes it virtually impossible to tap into the card’s information by manually finding signals or nodes to attack physically (Fig. 7.14). The glue logic design could be done automatically with using special design tools.

Technological progress on its own is increasing the costs to the attackers. Ten years ago it was possible to use a laser cutter and a simple probing station to get access to any point on the chip surface, but for modern deep submicron semiconductor chips very sophisticated and expensive technologies must be used.
That excludes most potential attackers. For example, the structure of the Microchip PIC16F877 microcontroller can be easily observed and reverse engineered under a microscope (Fig. 7.15). The second metal layer and polysilicon layer can still be seen even if buried under the top metal layer. This is possible because each subsequent layer in the fabrication process follows the shape of the previous layer. Under a microscope the observer sees not only the highest layer but also edges that reveal the structure of the deeper layers. In 0.5 μm and smaller technologies, for example in the Microchip PIC16F877A microcontroller, each predecessor layer is planarized using chemical-mechanical planarization (CMP) process before applying the next layer. As a result the top metal layer does not show the impact of the deeper layers (Fig. 7.15). The only way to reveal the structure of the deeper layers is by removing the top metal layers either mechanically or chemically.

Modern smartcard protection features now typically include internal voltage sensors to protect against under- and over-voltages used in power glitch attacks; clock frequency sensors to prevent attackers slowing down the clock frequency for
static analysis and also from raising it for clock glitching attacks; top-metal sensor meshes (Fig. 7.16); internal bus hardware encryption to make data analysis more difficult; and light sensors to prevent an opened chip from functioning. Software access to internal memory is often restricted by passwords, so that simple hacks to read out all the memory on the bus are no longer available.

7.8 Conclusion

There is no such thing as absolute security. A determined hacker can break any protection given enough time and resources. The question is how practical it would be. If it takes ten years to break a device which in three years is replaced by a successor with even better security, then the defense has won. On the other hand, the vulnerability could be buried within the design blocks itself. What if your secure system was designed from insecure components? In the end, the overall security of your system is determined by the least secure element. Even if you implement a provably secure protocol, your system could be broken if the key can be easily extracted from the hardware by mechanical or optical probing. Therefore, whenever you design a secure system, proper security evaluation of all the components must be performed. Of course it is impossible to avoid all problems; a reasonable goal is to make the process of breaking your design more expensive and time-consuming. With luck, potential attackers will switch to other products rather than spending money and effort on breaking yours.

The first step in designing adequate protection is to understand motivations of prospective attackers and possible attack scenarios. Then depending on the most likely class of attackers – outsiders, insiders or funded organizations, more realistic estimation of their capabilities can be made. From that a decision on what to protect and the level of protection can be made.

References