BRNO UNIVERSITY OF TECHNOLOGY
VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FACULTY OF INFORMATION TECHNOLOGY
FAKULTA INFORMAČNÍCH TECHNOLOGIÍ

DEPARTMENT OF INTELLIGENT SYSTEMS
ÚSTAV INTELIGENTNÍCH SYSTÉMŮ

ANALYSIS OF ATTACKS ON (MICRO)CHIPS AND DEVELOPMENT OF ENHANCEMENT OF THEIR ROBUSTNESS/SECURITY

ANALÝZA ÚTOKŮ NA (MIKRO)ČIPY A NÁVRH ZVÝŠENÍ JEJICH ODOLNOSTI/BEZPEČNOSTI

DOCTORAL THESIS
DISERTAČNÍ PRÁCE

AUTHOR **Ing. DOMINIK MALČÍK**
AUTOR PRÁCE

SUPERVISOR **prof. Ing., Dipl.-Ing. MARTIN DRAHANSKÝ, Ph.D.**
ŠKOLITEL

BRNO 2019

Abstract

Nowadays, microchips are used virtually everywhere, from simple home devices to confidential military equipment. In many scenarios, sensitive data is being processed by these devices. For example, in the case of electronic personal documents, fingerprints, facial images, and personal data are processed by the chip; and in some cases also iris images. Auditing proclaimed functions and a level of security of such microchips is becoming a valued service. In this doctoral thesis, we present an experimentally proven process for the microscopic analysis of chips, feasible in a low-cost setup. The described process was demonstrated on a chip acquired from the Czech biometric passport—from extracting the chip out of the plastic card up to analysis of the acquired microscopic images. We investigated and evaluated various potentially viable methods for logic element recognition; without the employment of machine-learning. Additionally, hardware-oriented attacks are discussed and followed by proposals for countermeasures leading to the hindering of microscopic analysis.

Abstrakt

S využitím mikročipů se dnes setkáváme prakticky na denní bázi, od jednoduchých zařízení pro domácí použití až po utajované vojenské vybavení. V mnoha případech navíc svěřujeme těmto zařízením velmi citlivá data, jako i v případě elektronických dokladů – otisky prstů, fotografie obličeje, osobní data; a v některých případech například i obraz oční duhovky. Ověření deklarované funkčnosti a míry zabezpečení takových mikročipů se tak stává žádanou službou. V rámci této disertační práce prezentujeme experimentálně ověřený proces mikroskopické analýzy mikročipů proveditelný v nízkonákladovém režimu. Popsaný proces jsme poté demonstrovali na čipu z českého biometrického pasu – od získání čipu z plastové karty až po jeho analýzu na základě získaných mikroskopických snímků. V rámci analýzy jsme prozkoumali a porovnali různé metody bez strojového učení potenciálně využitelné k rozpoznávání logických elementů. Dále jsme provedli zhodnocení aktuálních hardwarově orientovaných útoků na mikročipy. V návaznosti na toto zhodnocení jsme navrhli možná protiopatření zaměřená primárně na ztížení procesu mikroskopické analýzy.

Keywords

Microchip, chip, chip package, deprocessing, dry etching, wet etching, security analysis of chips, microscopic analysis, SmartMX, MIFARE Classic.

Klíčová slova

Mikročip, čip, pouzdro čipu, deprocesing, plasmatické leptání, chemické leptání, bezpečnostní analýza čipů, mikroskopická analýza, SmartMX, MIFARE Classic.

Reference

MALČÍK, Dominik. *Analysis of attacks on (micro)chips and development of enhancement of their robustness/security.* Brno, 2019. Doctoral thesis. Brno University of Technology, Faculty of Information Technology. Supervisor prof. Ing., Dipl.-Ing. Martin Drahanský, Ph.D.

Analysis of attacks on (micro)chips and development of enhancement of their robustness/security

Declaration

I hereby declare that this thesis is my original work and has been created under the supervision of prof. Ing., Dipl.-Ing. Martin Drahanský, Ph.D. Some information was provided by Ing. Vladislav Hrachovec (ON SEMICONDUCTOR), Karine Danilo (PRESTO ENGINEERING) and Andrey Denisyuk, Ph.D. (TESCAN Brno). I listed all of the literary sources and publications that I have used.

.....
Dominik Malčík
August 31, 2019

Acknowledgements

I would like to thank prof. Martin Drahanský, for his support, provided advice, and all the effort he had to expend regarding this doctoral thesis. Furthermore, I would like to express special thanks to Ing. Vladislav Hrachovec (ON SEMICONDUCTOR), Karine Danilo (PRESTO ENGINEERING) and Andrey Denisyuk, Ph.D. (TESCAN Brno) for valuable discussions. Last but not least, my family and my girlfriend Radka also deserve a big thank for the support and patience they maintained.

Contents

Chapter 1

Introduction

1.1 Motivation

Nowadays, many different types of chips are used virtually everywhere in the real world. There exist various incentives why one would like to see what is under the hood of a chip. Starting with QA departments of the chip producers, continuing over competition scanning, up to security auditing (i.e., auditing a device that was proclaimed secure). Such investigations are available in the market; however, these are mostly available in the commercial sector. The companies providing these services are keeping their know-how concealed. In the academic sphere, we are witnessing rather isolated attempts of such analyses scattered among various departments than a coherent work. We would like to contribute to the knowledge and capabilities maintained in academia with the mapping of the microscopic analysis in detail and, of course, with providing additional value to the status quo.

A few decades back, the chips were rather simple, and actually not that small. Observations, and even understanding of such devices, were possible with not much effort, i.e., we have decapsulated and deprocessed a single chip containing only four NAND cells. Analyses of these uncomplicated devices requires just an optical microscope and a few hours of work in a lab. On the contrary, dealing with contemporary chips used in biometric passports is a completely different challenge (and we have to admit that there are even more advanced chips, e.g., conventional CPUs, GPUs). The growing integration density and increasing die area result in the enormous complexity of the devices. Thus, it is impossible to apply manual only approaches. Moreover, the equipment capable of dealing with such advanced devices is very costly and thus not always available to low-cost attackers, as we are.

1.2 Thesis Contribution

Our focus laid mainly in the area of microscopic analysis of the microchips. We had to manage the whole process of chips decapsulation and deprocessing in order to get to the dies we wanted to investigate. Thus, the contributions of this thesis are broader than just a single topic. We truly believe that this thesis will help other institutions that want to tackle microscopic analysis of chips to manage the process based on the very detailed instructions and processes we present in this thesis. We believe that the main contributions are:

- A detailed description of methodologies for obtaining, decapsulation and deprocessing of the chips including improvement of the process of obtaining chips from thermoplastic compounds—plastic cards used for wrapping smartcard chips.
- Proposals for security improvements of the chips, focused primarily on hindering microscopic analysis. Techniques like 3D integration, MEMS, or battery-backed security fuses are presented in this thesis.
- Analysis of a chip belonging to the SmartMX family (used in biometric passports, ID cards, etc.) performed in a low-cost setup. Although not having regular access to needed equipment, we were able to completely analyze the chip construction—described in detail in this thesis, partially deprocess it and analyze the gained specimens.
- The comparison of methods that are not commonly used for processing of the chip silicon layer images without the employment of typically used machine-learning methods. Such an approach removes the need for complex training and verification data sets.

1.3 Thesis Organization

The text of this doctoral thesis consists of ten chapters. Chapter 2 provides basic insight into physics—semiconductor and transistors related knowledge, logic gates composition, and relevant logic families preview. This overview introduces the reader into fundamental principles employed in integrated circuits. Chapter 3 advances to a higher level of abstraction and presents typical chip components commonly used for building microchips. A general overview of possible attacks applicable on the chips with emphasis on hardware-oriented invasive attacks is presented in Chapter 4. The topic covering attacks overflows into Chapter 5, where the attacks are further discussed, and possible security enhancements are proposed. Chapter 6 introduces the reader into the biometric passports topic. Information provided in this chapter allows the reader to gain a broader view into the domain of e-documents that is important for the following chapters, where the analysis of microchips used in these e-documents is performed. The whole process from obtaining the chips to post-processing images acquired by a microscope is presented in detail in Chapter 7. Chapter 8 is devoted to the analysis of the Czech biometric passport chip. Comparison of various methods potentially usable for logic elements recognition without the employment of learning approaches is given in Chapter 9. Finally, Chapter 10 provides a final summary of the work with a conclusion and future work proposals.

Chapter 2

Chip Manufacturing

In order to be able to analyze the chips, one has to know how the chips are manufactured and what the fundamental rules of physics are that allow us to produce such powerful devices. Let us formulate a brief overview in this chapter of the knowledge required.

2.1 Materials, Electricity, and Conductivity

This thesis is not dealing primarily with physics nor with electricity itself. Therefore, let us assume a basic knowledge of physics regarding electricity and conductivity among readers. Nevertheless, a few facts should not be omitted.

One of the very basic electromagnetic properties is the ability to conduct electricity. The conventional current flows from the positive terminal of a source through the load to the negative terminal of that source. In metals, conductivity is performed by electrons, and electrons flow in the exact opposite direction than conventional current direction. This historical agreement is kept valid, although it may be confusing. According to the properties of materials concerning conductivity and resistance, we split these into three main groups—conductors, semiconductors, and insulators. All these three groups play their unique role in the production of microchips.

Conductivity (σ, sometimes also κ or γ) is reciprocal with electrical resistivity (ϱ). The SI unit of conductivity is siemens per meter ($\mathrm{S\,m^{-1}}$) and the SI unit of resistivity is ohm-meter ($\Omega\,\mathrm{m}$). The reciprocal relation is expressed as $\sigma = \frac{1}{\varrho}$ and thus $\varrho = \frac{1}{\sigma}$ can be easily transposed. Resistivity (and also conductivity) varies according to the type of material, its purity, and temperature—see Table 2.1. The difference in resistivity among conductors, semiconductors, and insulators is briefly shown in Table 2.1. [91], [35], [110], [158]

Table 2.1: Resistivity in conductors, semiconductors and insulators. (Source: [158]; Martin Drahanský, FIT BUT course Intelligent Sensors)

Conductors	
	Resistivity ($\Omega\,\mathrm{m}$)
Ag	1.6×10^{-8}
Al	2.8×10^{-8}
Cu	1.7×10^{-8}

Semiconductors	
	Resistivity ($\Omega\,\mathrm{m}$)
Ge	0.47
Si	3.0×10^{3}
InSb	2.0×10^{4}

Insulators	
	Resistivity ($\Omega\,\mathrm{m}$)
Mica	9.0×10^{14}
Silica	3.0×10^{14}
Diamond	1.0×10^{14}

It is worth noting the difference between electrical resistivity (ϱ) and resistance (R); while resistivity is a property of material, so-called material constant, the resistance is

a property of a particular object. The same relation can be seen between conductivity (σ) and conductance (G). Conductivity is the intrinsic property, while conductance is the extrinsic property. In other words, conductivity is the inherent property of material. Conductance is then the property of an object dependent on various aspects, e.g., amount, mass, size, or physical shape.

Conductance and resistance are also in a similar relationship, as conductivity and electrical resistivity—$G = \frac{1}{R}$ SI unit of conductance is siemens (S), that is according to the given equation Ω^{-1}. Conductance can be computed from conductivity (σ), cross-sectional area (A) and length (l) as follows—$G = \sigma\frac{A}{l}$. Based on all relationships among the mentioned physical magnitudes, resistance is calculated in a very similar way—$R = \varrho\frac{l}{A}$. [35], [158].

2.1.1 Conductors

Conductors, also sometimes called metals, conduct electricity quite well. Conductivity in metals is performed by electrons. Each atom of metal typically provides one free electron to electron gas. Unlike with semiconductors, the conductivity decreases with temperature. It is given by an increase of collisions among electrons in the material. Each collision then causes a slowdown of such a movement of the electron gas. Related to the scope of this thesis, conductors are used in microchips mainly for interconnections, but also for transistor manufacturing. [158], [76]

2.1.2 Semiconductors

The relatively poor conductivity of semiconductors is redeemed by a good possibility to control their electrical properties. This is exactly what is needed for the construction of "switches" that are essential for logic in electronic devices. Another positive fact about semiconductors is that semiconductors are very often built on a silicon basis. And silicon is one of the most affordable materials on Earth.

The electrical resistivity of semiconductors decreases exponentially with rising temperatures. In other words, the conductivity increases exponentially with temperature, which is the complete opposite behavior to metals.

Pure semiconductors (intrinsic semiconductors, undoped semiconductors) are not conducting current very well, since there is always the same number of holes as of free electrons in such a semiconductor. To upgrade a semiconductor to a more suitable version, there exists a technique called doping. That is, in fact, the intentional and controlled introduction of impurities into the crystal structure. This technique allows creation of p-type semiconductor (doped with, e.g., indium, boron) and n-type semiconductor (doped with, e.g., phosphorus, arsenic). So-called extrinsic semiconductors are created. Consequently, it is possible to build a PN junction between differently doped regions. [158],[76]

2.1.3 Insulators

The last group represents materials with very poor conductivity given by the very tight binding of electrons in their atoms. Insulators are used to separate conductors or semiconductors, ensuring no conductivity among them. Another important insulating feature among insulators, except the mentioned conductivity insulation, is thermal insulation. The insulators have their limits given primarily by high resistance (the ability to prevent current from passing through) and so-called breakdown voltage, or dielectric strength. The

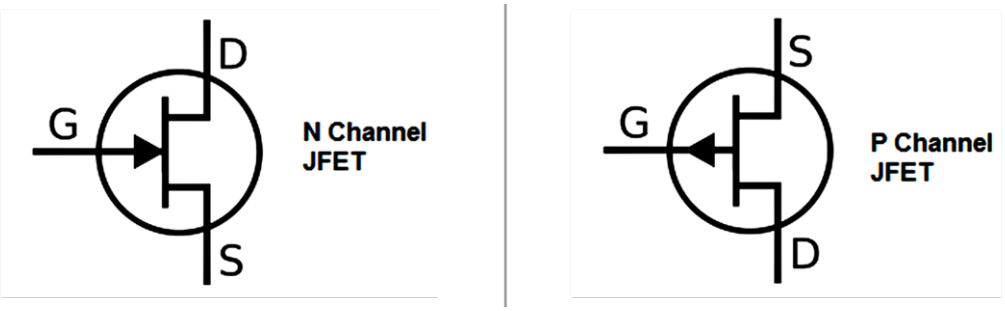


Figure 2.1: Unipolar JFET transistor schematic. Left: n-type; Right: p-type. (Source: [94])

breakdown voltage value is intrinsic for each material and tells us at what voltage level the particular insulator loses its insulating properties and allows electric current to flow. [158], [76]

2.2 Transistors

Transistors are three terminal-active electric devices. An active device is a device that can amplify the electric signal. Compared to that, passive devices are only transporting the signal. The transistors are made of different semiconductor materials—generally from silicon or germanium.

The main two functions of transistors are switching—used in digital electronics—or amplification, which is used in analogue electronics.

There exists different technologies for the production of transistors; nevertheless, the most important one for microchips is MOSFET (Metal Oxide Semiconductor Field Effect Transistor). These transistors are unipolar, either pMOS type or nMOS. The subsequent technique for chip production is called CMOS (Complementary Metal Oxide Semiconductor) which uses both nMOS and pMOS transistors equally and as a big advantage, these transistors are consuming power only during the switching part. [158], [76]

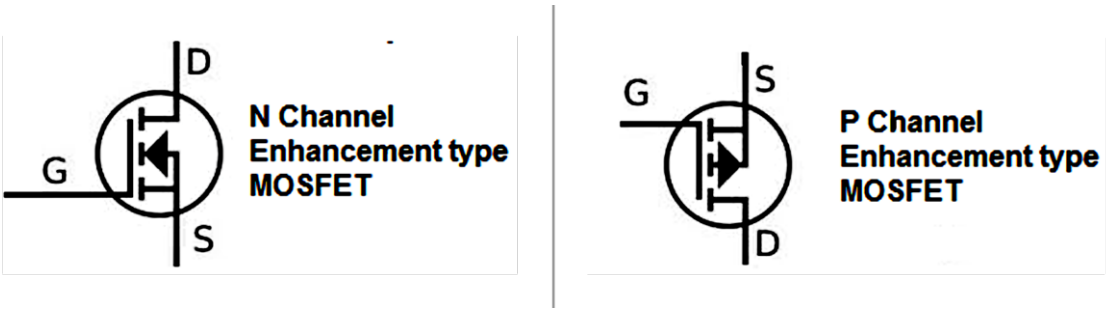


Figure 2.2: Unipolar MOSEFT enhancement-mode transistor schematic. Left: n-type; Right: p-type. (Source: [94])

2.2.1 Unipolar Transistors

Unipolar transistors are, as the name says, of either p-type or n-type. The current is conducted by positive holes, or respectively by negative electrons. This type of transistor is

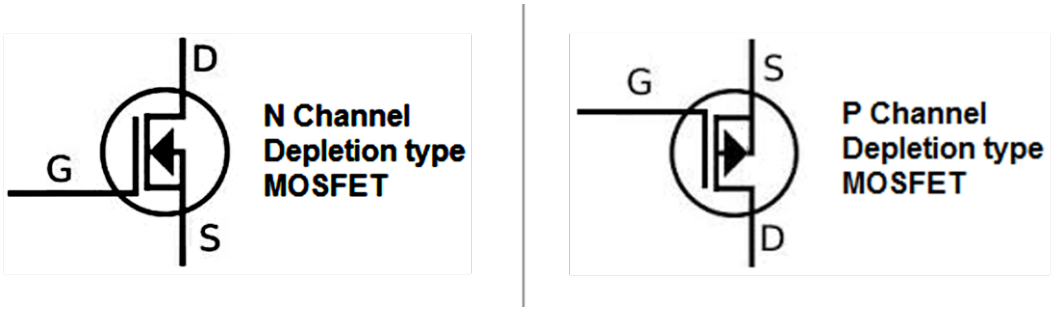


Figure 2.3: Unipolar MOSEFT depletion-mode transistor schematic. Left: n-type; Right: p-type. (Source: [94])

the most common transistor type that we encounter in many electronic devices. Unipolar transistors are so-called voltage controlling. There is no need for a biasing current to control the current flow between source and drain, because the channel between the source and drain terminals is controlled by an electric field induced by the voltage put on the gate terminal.

The FET (Field Effect Transistor) construction does not have a PN junction in its main current carrying path, but the PN junction is controlling the size, or rather the conductivity, of the channel between source and drain electrodes. This conductivity, as stated before, is controlled by the voltage applied to the gate terminal.

Generally, there are more types of FET transistors than just MOSFET, e.g., JFET (see Fig. 2.1) or JUGFET (Junction Gate Field Effect Transistor). Nonetheless, as was said before, MOSFET transistors are the most common ones. They can be produced in two main ways—enhancement mode (Fig. 2.2) or depletion mode (Fig. 2.3)—according to the expected purpose of use. Enhancement mode transistors are off at zero gate-source voltages. Depletion mode is the opposite of that; the transistor is on at zero gate-source voltage.

A positive property of FET-type transistors is that they consume only a very little amount of energy. Due to this fact, their usage in electronic devices using batteries is very convenient. [158], [76], [65], [135]

2.2.2 Bipolar Transistors

BJT (Bipolar Junction Transistor) transistors are called bipolar, because they operate with both types of charge carriers, holes, and electrons. There are two basic arrangements of bipolar transistors—PNP and NPN (see Fig. 2.4). These two arrangements are expressing the physical arrangement of the PN junctions present in the particular transistors. As it can be deducted from the acronyms, there are two PN junctions in bipolar transistors. The three terminals are labeled as the Emitter (E), the Base (B), and the Collector (C) to distinguish among them. Unless they are unipolar transistors, the bipolar ones are current-regulating devices. This means that we can control the amount of current flowing through these transistors related to the biasing voltage applied to the base terminal.

There are three possible connection types of the transistors in an electronic circuit—Common Base configuration, Common Emitter configuration and Common Collector configuration. [158], [76], [65], [135]

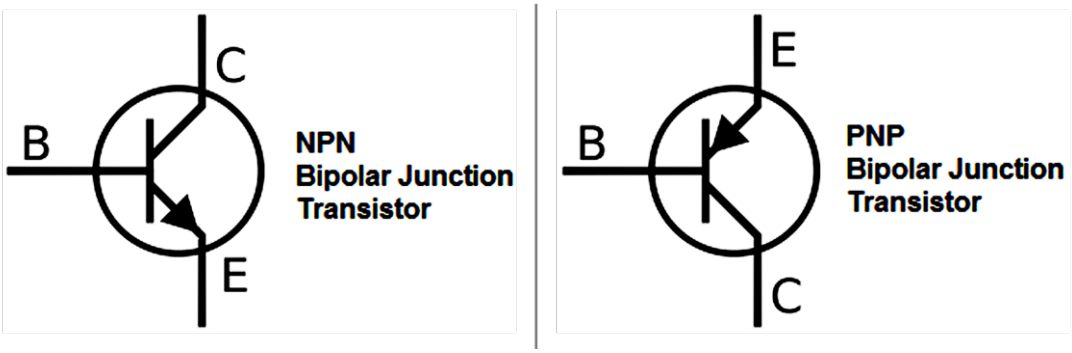


Figure 2.4: Bipolar transistor schematic. Left: NPN transistor type; Right: PNP transistor type. (Source: [94])

2.2.3 Physical Limits

The era of Moore's law [108] is slowly coming to an end, or at least to a slowdown, given by the complexity and costs of each next node in integration. When looking at current Intel roadmaps, it can be seen that we will witness hitting 10 nm technology for massive production in 2019. Another step, the 7 nm process, is planned for 2020–2021. Furthermore, it was already proven that a single atom transistor for subsequent quantum architecture can turn into a reality [56]. Of course, a lot of research is still needed.

The radius of a single silicon atom is around 0.2 nm [151]. For building up a conventional transistor, it needs to put together 5–6 atoms. That leads us to a thought of a 1 nm theoretical limit. On the other hand, below 5 nm, it is already known that the technology will have to deal with quantum tunneling and other similar scientific issues still considered mysterious. Thus, a limit below 5 nm is a bit experimental and not supported by the hard facts.

This unavoidable technical progress is also bringing about significant changes in the analysis methods that have to accommodate these new approaches. As the transistors are becoming three-dimensional, the classic decomposition and consequent observation of planar layers with the use of any type of capable microscope might be getting insufficient.

2.3 Logic Gates

Logic gates is in our case are small devices that are able to accept Boolean values (zeros and ones representing true and false) and to process the inputs into a resulting Boolean value or possibly into more Boolean values. Boolean logic is used because of its simplicity and possibility to represent true and false values by physical means—e.g., high (1) and low (0) voltage levels. In history, there were also attempts to use more states than only two—high and low—in order to have wider alphabet available for computing. Nevertheless, it was proven by time that two state values are the best fit given noise and the physical construction of transistors. Therefore, Boolean logic prevails. [158], [76]

Table 2.2: Truth table.

x	y	$\neg$ x	x $\wedge$ y	x $\vee$ y	x $\Rightarrow$ y	x $\Leftrightarrow$ y	x $\uparrow$ y	x $\downarrow$ y	x $\oplus$ y
0	0	1	0	0	1	1	1	1	0
0	1	1	0	1	1	0	1	0	1
1	0	0	0	1	0	0	1	0	1
1	1	0	1	1	1	1	0	0	0

2.3.1 Boolean Logic

In Boolean logic, generally everything is computed with values from a set $\{0, 1\}$. This set represents two possible values—True and False. All following operations are depicted in Table 2.2.

There exists three basic operations in Boolean logic—conjunction ($\wedge$, AND), disjunction ($\vee$, OR), and negation ($\neg$, NOT). Conjunction and disjunction are binary operations, negation is a unary operation.

From the abovementioned basic operations, we can build other operations, i.e., secondary operations—material conditional ($\Rightarrow$, there is no explicit logic gate representing this operation), biconditional ($\Leftrightarrow$, XNOR) and exclusive or ($\oplus$, XOR).

Finally, we are getting to other important derived operations, especially interesting in the sphere of the physical design of chips—inverted conjunction ($\uparrow$, NAND) and inverted disjunction ($\downarrow$, NOR). Why particularly these two operations are the most interesting ones? Simply put, each of them is functionally complete on its own. In other words, it is possible to build all other operations just with binary NAND logic gates, or respectively with binary NOR gates. If we look at the functional completeness from another perspective, it also says that with each of these operations, we can build all possible truth tables. There is no other operation that is functionally complete as such. It always has to be paired with at least one other operation to create a set of functionally complete operations, e.g., $\{\neg, \wedge\}$, $\{\neg, \vee\}$. This is exactly the reason why the most commonly used logic gates are just these two. When we dig even deeper, NAND gates usually prevail. We will get to the root of NAND's dominance later in this chapter. [76]

2.3.2 Basic Logic Gates

In this section we will describe basic logic gates with respect to NAND logic and NOR logic. These terms come from the fact that these two operations are functionally complete and thus can build all other logic gates purely out of each of them. What is worth noticing are the slight differences in composition of the particular logic gates in NAND and NOR logic. Common schematic symbols according norms IEC—IEC 60617-12, US—ANSI/IEEE Std 91-1984 and ANSI/IEEE Std 91a-1991, DE—DIN 40700 are shown in Fig. 2.5. [158], [76]

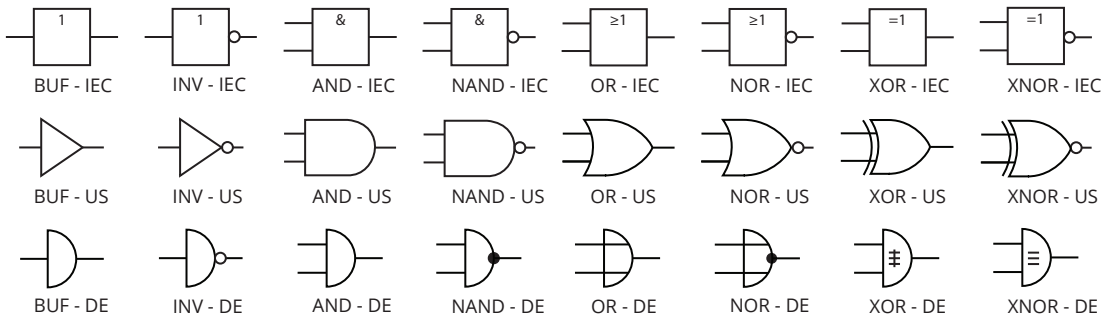


Figure 2.5: Comparison of schematic symbols. (Source: [171])

NOT

This is a logic gate with one input and one output, providing an inverted value of the input. NAND and NOR construction is very similar due to the fact that there is the same value always on both inputs of NAND or NOR gate (see in Fig. 2.6). In Truth Table 2.2, it can be seen that both operations act similarly for both zeros and ones on input. [158], [76], [135]

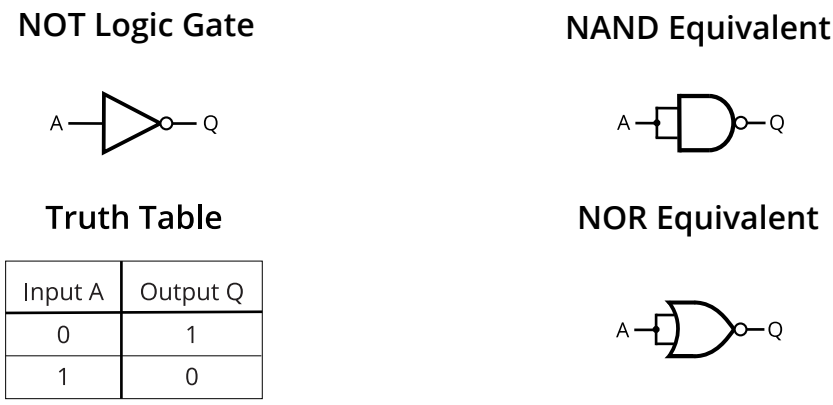


Figure 2.6: NOT gate and its NAND and NOR construction. (Source: [135])

AND

Binary operation conjunction is represented by the AND logic gate (see Fig. 2.7). This gate is in fact Boolean multiplication and returns a value of 1 only in case both input values equal 1. Construction in NAND logic is simple, because NAND is in fact NOT AND. To get AND out of NAND we place one inversion (NOT) after a NAND cell. In NOR logic we have to realize when the NOR gate outputs 1 (see Table 2.2)—this is only in case both input values are 0. Compared to that, AND outputs 1 only in case both inputs are 1. Then it is as easy as inverting both of the input values before a NOR gate. [158], [76], [135]

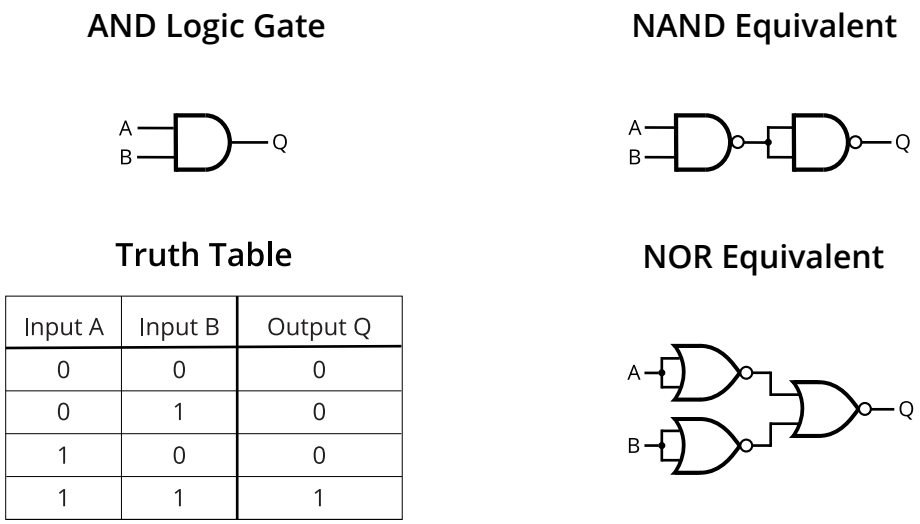


Figure 2.7: AND gate and its NAND and NOR construction. (Source: [135])

OR

OR gate is very similar to AND (see Fig. 2.8). For NAND logic we see in the Truth table the only difference—when both inputs are of the same value, we need to get exactly the opposite result as we have with the NAND gate. This naturally leads to performing inversions on both input values. For NOR logic, we use the same approach as with NAND and AND—NOT is added after NOR gate to form OR results. [158], [76], [135]

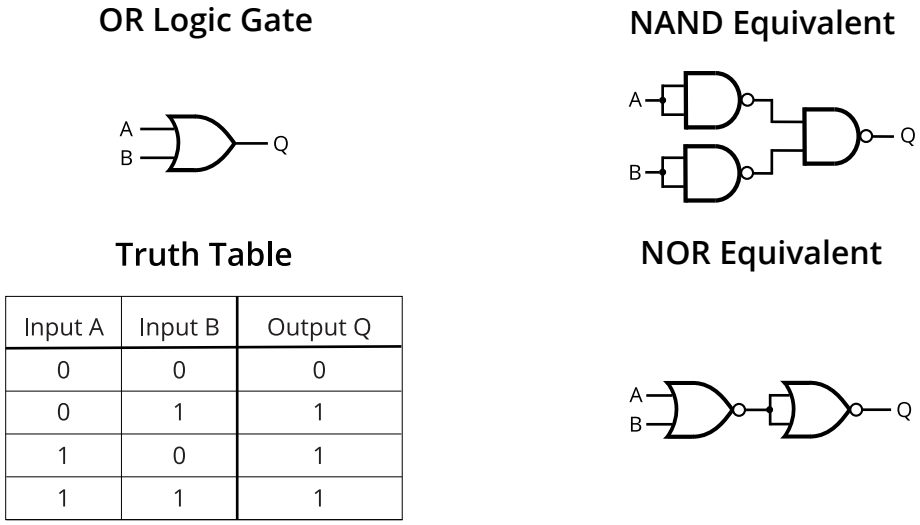
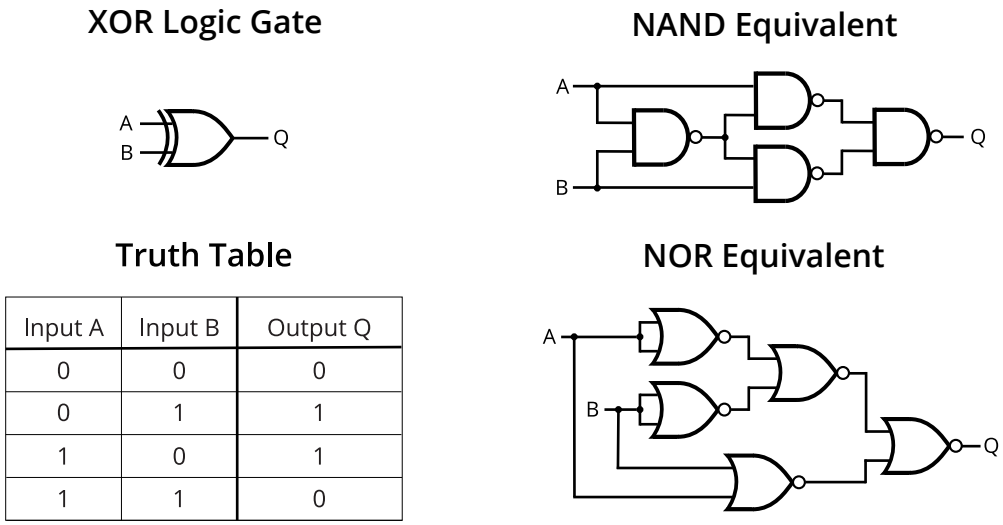


Figure 2.8: OR gate and its NAND and NOR construction. (Source: [135])

XOR

Exclusive OR (see Fig. 2.9), also called Boolean addition (therefore sign $\oplus$), outputs 1 if exactly one of the inputs is 1. When inputs are 0, there is simply nothing to add. And in the case both inputs are 1, the resulting 1 overflows, because we cannot express number 2 just with 1 bit. This leads to a thought that this gate is probably good for adding numbers in a proper combination with other gates that ensure the correct handling of the overflown (carried) 1. Construction with NAND elements is similar to OR construction with one extra NAND inserted in front of the modified OR part. This insertion guarantees the correct processing of input A equals 1 and B equals 1. To prepare the setup with NOR logic, we need to see when we can get value 1 on a NOR gate output first—only in case both inputs of such gate are 0. That means we have to combine AND and NOR blocks in parallel in order to prepare the proper input for our last NOR logic gate, which produces the final XOR output. [158], [76], [135]

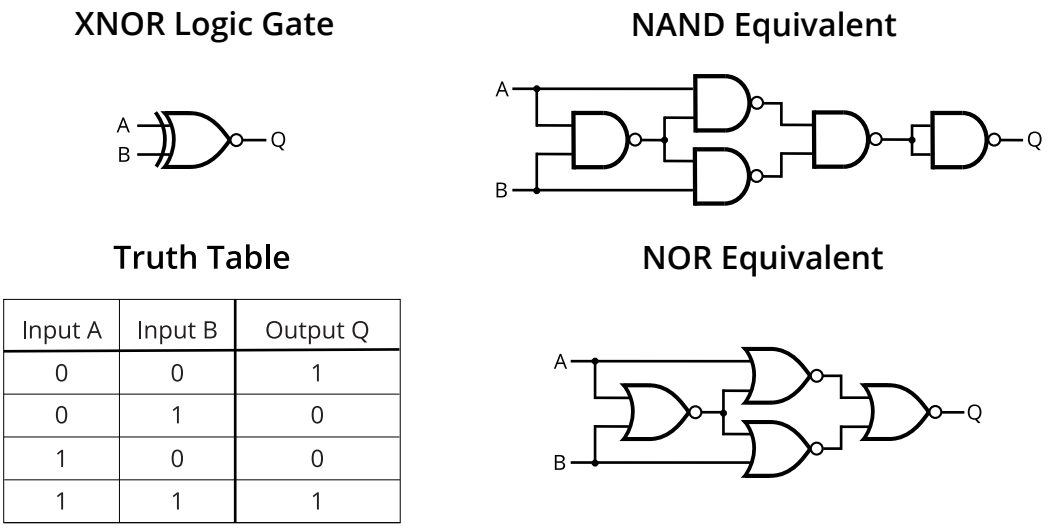


Input A	Input B	Output Q
0	0	0
0	1	1
1	0	1
1	1	0

Figure 2.9: XOR gate and its NAND and NOR construction. (Source: [135])

XNOR

This gate is an inversion of the XOR gate (see Fig. 2.10). Nevertheless, we can see as well that it is presenting logical biconditional operation. For NAND structure, the XOR block is used as a base and the result is inverted at the end. With NOR gates we might do the same, however this would be too complex from the number-of-gates perspective. It is possible to take an AND block made of NOR gates and use the same trick as was done in XOR construction with NAND gates. The AND block is almost what is needed, except the situation when there is 0 on both inputs. This specific situation is solved with inserting a NOR block before the AND part. [158], [76], [135]

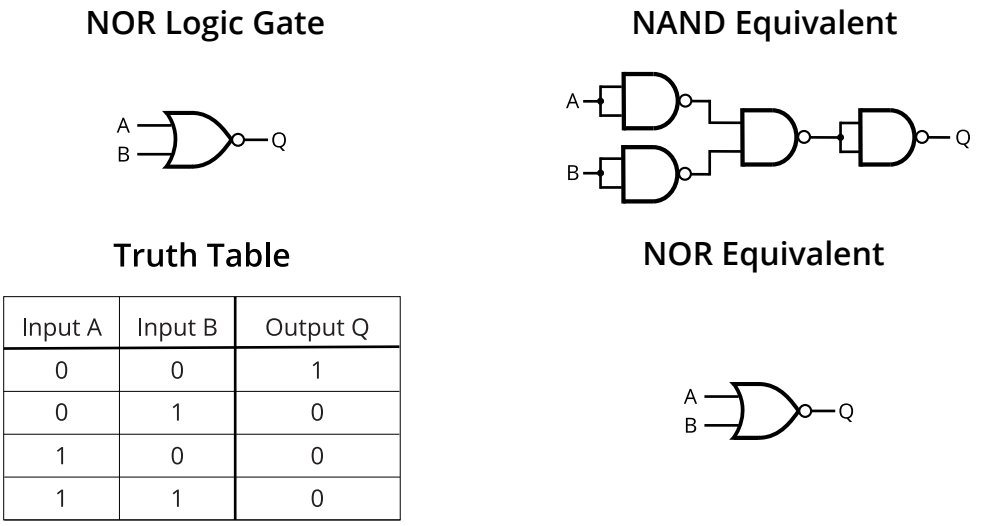


Input A	Input B	Output Q
0	0	1
0	1	0
1	0	0
1	1	1

Figure 2.10: XNOR gate and its NAND and NOR construction. (Source: [135])

NOR

The most straightforward approach is to take the OR gate and invert its results (see Fig. 2.11). This is how the NOR gate is built with use of NAND gates. [158], [76], [135]



Input A	Input B	Output Q
0	0	1
0	1	0
1	0	0
1	1	0

Figure 2.11: NOR gate and its NAND and NOR construction. (Source: [135])

NAND

Analogically to building NOR gates out of NANDs (see Fig. 2.12), the NAND gate made of NORs is assembled in the same manner—inverting AND gate results. [158], [76], [135]

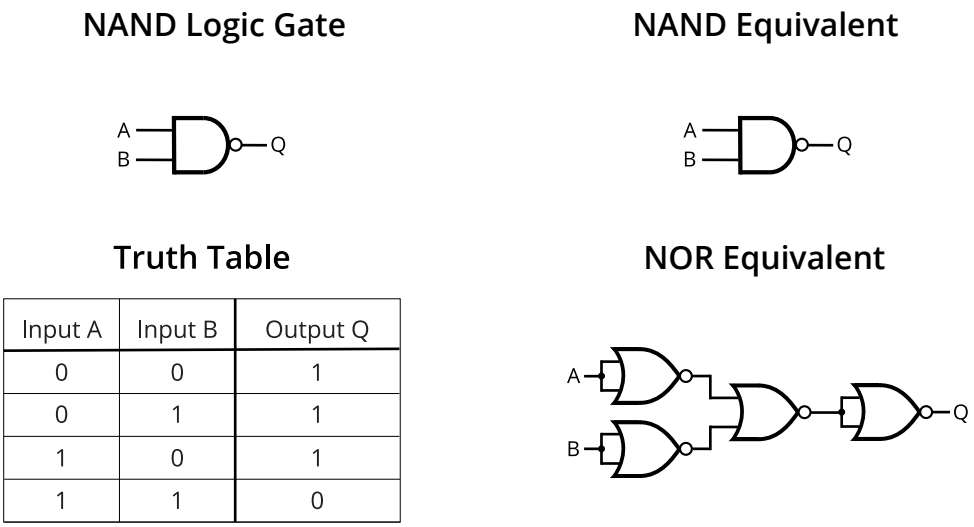


Figure 2.12: NAND gate and its NAND and NOR construction. (Source: [135])

2.4 Logic Families

As the technology evolved from the very beginning, there were different eras along the path used for mainstream mass production. There existed many variations of each of the families mentioned in the following lines. Nonetheless, the given overview is perfectly sufficient to serve as a summary for the scope of this thesis.

The most important logic families using bipolar transistors:

RTL

Resistor-transistor logic was the early stage technology for building integrated circuits. In the times of RTL, transistors were the most expensive parts of electronic devices. This logic could use only one bipolar transistor supplemented with resistors to compose a NOR gate. Low noise resistance and high power dissipation were the biggest disadvantages of this technology that practically prevented its wider use. [111], [76]

DTL

Diode-transistor logic was a successor of RTL, removing its disadvantages that were mentioned in the previous paragraph. The logic was performed by diodes, and diodes belong to passive components that don't consume power. Transistors were used then "only" for amplification of the signal. The speed of such circuits was still poor. High and low logic levels are 5.0 V and 0.2 V respectively. [65]

ECL

Emitter-coupled logic, sometimes also called Current-mode logic, is one of the fastest logic families. Its extreme high speed is achieved by keeping all BJT transistors saturated, nevertheless not deeply saturated. Due to this fact, times required to charge and discharge the transistors are rapidly reduced. A basic element for ECL is a BJT differential pair. As readers may object, while keeping transistors saturated, it has to be supplied with the appropriate power. The trade-off of ECL compared

to RTL or DTL is clear—the power consumption has to be sacrificed to gain speed. Because the voltage levels for high (-0.9 V) and low (-1.7 V) are very close to each other, noise immunity is low, which is considered a strong weakness of ECL. [76], [135]

TTL

Transistor-transistor logic is another family based on bipolar transistors and resistors. Transistors perform both functions in TTL—logic and amplification. The logic level 0 range is 0.0 V-0.8 V. The voltage range for logic level 1 is 2.0 V-Vcc, where Vcc is usually 5 V or in low voltage version 3.3 V. Compared to ECL, the TTL is significantly slower. However, design complexity and power consumption are lower. Therefore, we might encounter combinations of TTL and ECL elements in order to optimize the circuit according to performance and power requirements. In comparison with CMOS technology, TTL consumes more power in its rest state. On the other hand, the power consumption does not rise so quickly with the rising of the operating frequency. Radiation sensitivity or electric discharge sensitivity features are significantly better with the use of TTL. [158], [76], [135]

IIL, sometimes also I2L

Integrated injection logic uses only bipolar junction transistors. Due to the absence of resistors, the spatial and power requirements are very low compared to RTL, DTL, and ECL. Hence this technology is suitable for LSI (Large Scale Integration) and even for VLSI (Very Large Scale Integration). Performance of such devices is also on a much better level than with RTL or DTL. Although the logic levels are even closer than in ECL—high 0.7 V, low 0.2 V—IIL still keeps good noise immunity. It is possible, because IIL is operated by current instead of voltage. [135]

Let us now proceed from bipolar to unipolar transistors. Here is again a brief description of technologies based on this type of transistor:

PMOS

The P-type metal-oxide-semiconductor was massively used earlier than NMOS. The reason for this initial prevalence was simple—it was easier to produce this type of transistor. Conductivity in p-type is provided by positive holes, nevertheless the holes are slower than electrons used in "sister" NMOS technology. That is why devices based on PMOS tended to be slower, especially in performance of high to low transition. Moreover, power dissipation whenever the output is high, even in a static state, draws a huge disadvantage compared to the currently predominantly used CMOS. [135]

NMOS

The N-type metal-oxide-semiconductor uses n-type FET transistors where conductivity is given by electrons. Analogous to PMOS, n-type transistors are slow in transitions from low to high. The same analogy can be found in power consumption—power is used even in a steady state; in the case of NMOS, it is logic 0, or in other words in the low state. The positive side of NMOS is the speed, as it is faster than PMOS and CMOS in many cases, which uses both p-types and n-types. Unfortunately, one more negative fact regarding NMOS and PMOS has to be taken into account—both families are more susceptible to noise than CMOS. This is caused by asymmetric input logic levels. [135]

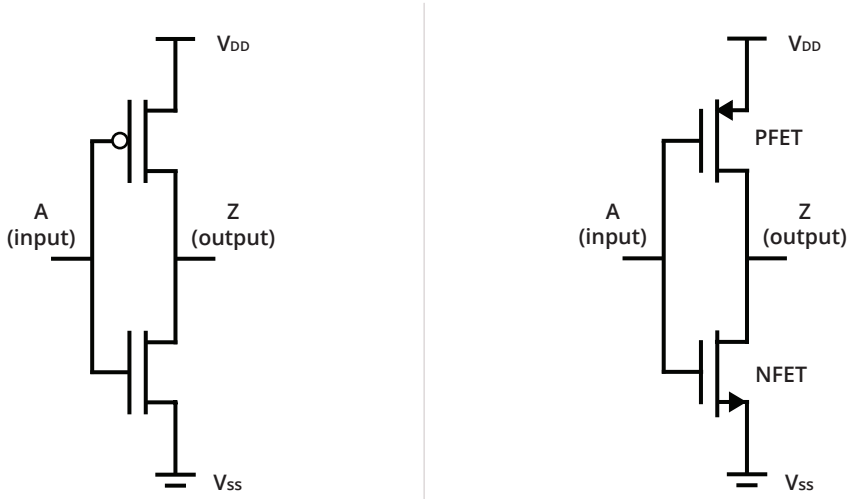


Figure 2.13: NOT schematic displayed with use of 2 different symbols. (Source: [170])

CMOS

Complementary metal-oxide-semiconductor uses n-type and p-type Field-Effect transistors (FET). It is undoubtedly most common and the most widely used technology for building integrated circuits. Therefore, we devote a few more lines in Chapter 2.4.1 to CMOS. [158], [76]

2.4.1 CMOS

Complementary metal-oxide-semiconductor was a natural step in the technology development following PMOS and NMOS. Big advantages of CMOS are the low power consumption in a static state as well as speed. Circuits made with this technology consume power mainly only for switching between states. This technology is currently still the most used one for the mainstream production of various integrated circuits.

According to previous statements in this chapter, the most used logic gates are NOT, NAND, and NOR. We will look closer at each of them up to CMOS transistor level detail. Nevertheless, before we get to this depth, it has to be clear how each of the p-type and n-type transistors work. PMOS uses positive holes, thus when we apply positive voltage to a gate, we close the positive channel. That means no conductivity between drain and source. Simply put, this type of transistor is "switched on" when we apply logic zero to the gate. That is why in some types of schematic symbols (see Fig. 2.13, 2.14, 2.15) of PMOS transistors there is a circle before gate to depict this behavior—the transistor is switched on when there is a zero on the gate. It is a kind of inversion, hence the symbol. NMOS behaves in the opposite manner. The channel of a transistor is opened in case there is logic 1 (positive voltage) on the gate.

These two types of transistors are then combined into CMOS technology with a ratio close to 1:1, therefore the word complementary. All particular advantages and disadvantages of each type of transistor are described in Chapter 2.4. Readers probably noticed that all the features of the p-type are complementary to n-type features, and this is the whole magic behind CMOS; getting the best of both transistor types. By designing the circuit wisely, static power dissipation can be nearly zero and the speed of the circuit can be

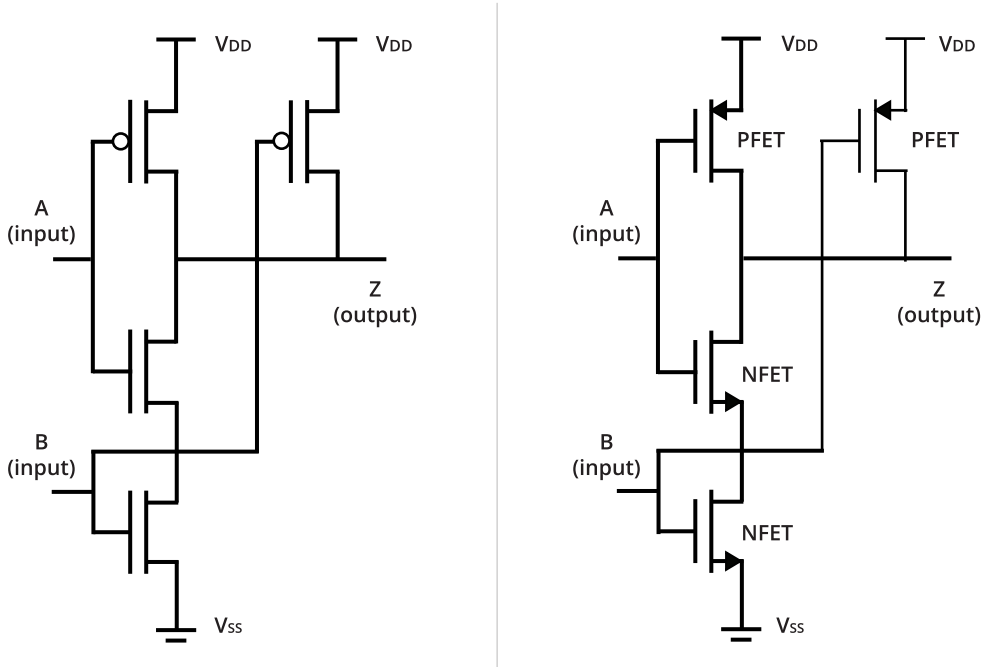


Figure 2.14: NAND schematic displayed with use of 2 different symbols. (Source: [170])

highly optimized. CMOS normally consumes power only when switching—i.e., dynamic power dissipation. [158], [76]

NOT Physical Structure

The demonstration of CMOS starts with an inverter. In Fig. 2.13 a NOT gate is depicted with two types of schematic symbols. As we know from the previous text, the PMOS type is switched on when there is low value on gate input (A). In this particular case, the transistor lets V_{DD} propagate to output (Z). At the same time, the NMOS transistor is switched off. If we change the input value (A) to high, the NMOS transistor is switched on and as it can be correctly expected, the PMOS transistor goes off. Thus, the output (Z) value will become low.

This inverter demonstrates all the essential principles of CMOS. First of all, only transistors are used, no resistors or diodes are needed—this ensures low spatial requirements of the CMOS based designs. The PMOS transistor is connected to V_{DD}, the NMOS to V_{SS}. In this parallel setup, there is always only one of the transistors switched on, no direct connection between V_{DD} and V_{SS} is possible and so no static power dissipation is observed. [170], [65], [76], [135]

NAND Physical Structure

In NAND realization, an inventer can be seen on the left top part of each NAND representation in Fig. 2.14. Nevertheless, the NMOS transistor of the inverter is not connected directly to V_{SS}, but serialized with another NMOS transistor from the input (B) pair of transistors. This serialization is for ensuring V_{SS} propagation only in case when both inputs are high. [170], [65], [76], [135]

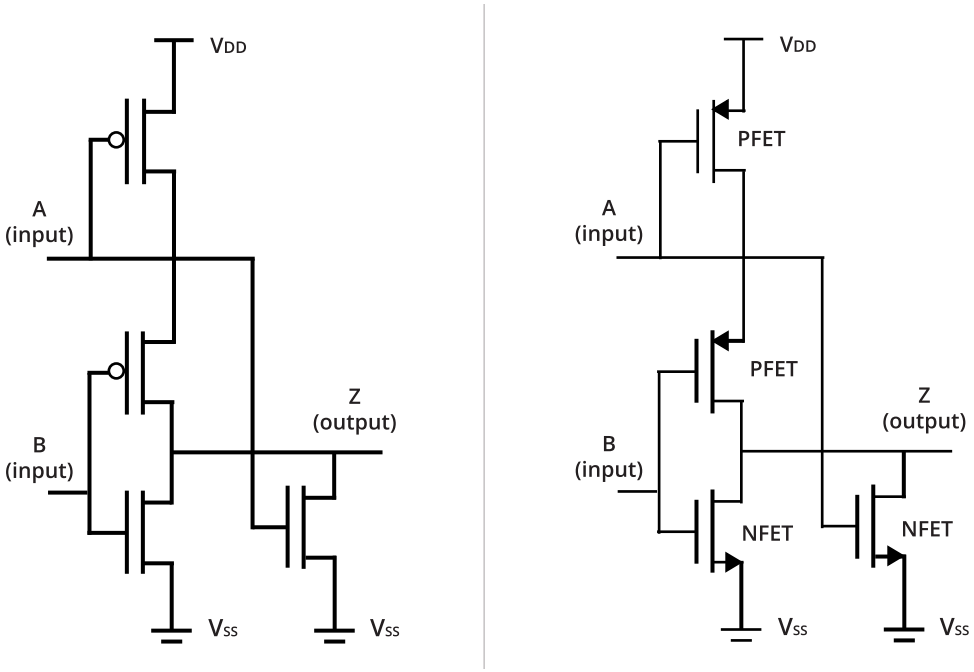


Figure 2.15: NOR schematic displayed with use of 2 different symbols. (Source: [170])

NOR Physical Structure

Of course, the NOR gate (see Fig. 2.15) uses the same principles and similar approach as NAND—only changed accordingly. Here, the V_{SS} prevails on output (Z) for all cases, except when both input values (A and B) are low. In such a case V_{DD} is transported to the output through two serialized PMOS transistors. [170], [65], [76], [135]

Chapter 3

Chip Components

When looking at chips from a distance, it can be seen that there are three main types of basic components—memory elements, logic elements, and interconnections. Chips also consist of other parts like power distribution, mechanical elements, heat management, obfuscation elements, insulation, etc. These parts are, of course, very important for ensuring the final properties of a chip. Nevertheless, from reverse engineering angle, we basically want to safely get rid of these to clearly see the relevant parts listed further.

3.1 Memory

The fundamental purpose of memory elements is to hold some information, either temporarily or permanently, and provide it for further processing when required. Depending on where is a certain memory needed in the architecture, the memory components are divided into groups determined mainly by speed and price parameters. Price aspect is, in fact, resulting in more other price-related parameters, e.g., capacity, the complexity of the manufacturing process, spatial requirements of memory cells on a chip. Very important features are power consumption and volatility. Hierarchy of common memory types is shown in Fig. 3.1. Each of the displayed memory types is described in the following sections. It ought to be mentioned that the boundaries between the memory classes displayed in Fig. 3.1 are not absolutely sharp, overlapping among some of them can be seen. This is

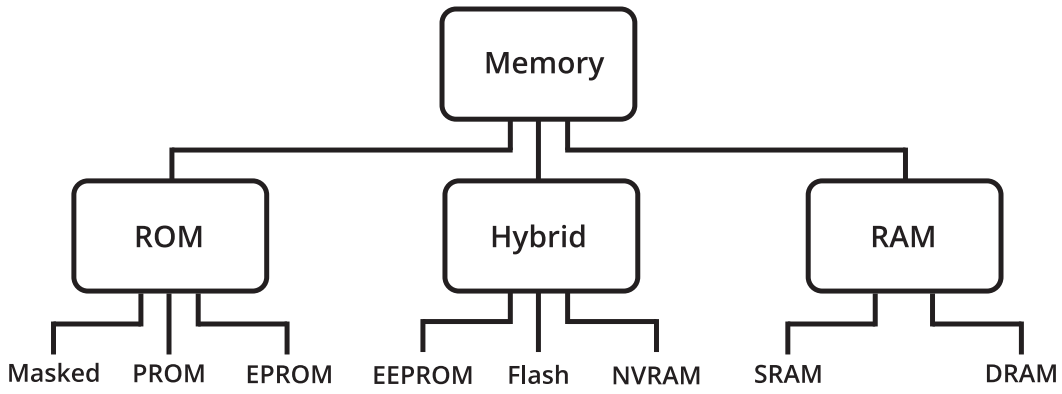


Figure 3.1: Memory types hierarchy. (Source: [11])

Figure 3.2: Various types of Mask ROM acquired by an optical microscope. (Sources: [119], [156])

caused by the evolution and endeavor to adopt the best features across all the types. [11], [170], [135], [65]

3.1.1 Read-Only Memory

ROM (Read Only Memory) is a non-volatile memory that holds its content even out of power. Typical ROM memory was originally meant to be non-writable. The information was stored inside during the manufacturing process only. Therefore, the usage was limited to holding functionality that was considered stable for the particular system (typically firmware, BIOS). Discovering a bug in ROM, especially in large series, may lead to serious economic harms due to the impossibility to fix the discovered problem in pure ROM memory. It does not apply to bugs only, but also to later discovered security issues or needed new features. This missing ability to update ROM leads naturally to evolution depicted in the following list—it has started with PROM (Programmable ROM) and continued over various enhanced versions of ROM that try to ease all the negative aspects of this type of memory. Despite we can overcome the biggest disadvantage of the original ROM memory and write new contents, writing into ROM is usually not an easy task. It may require special equipment, it can be very slow or introduces a limited amount of erase cycles. [11], [170], [135], [65]

Mask ROM

Mask ROM is the simplest and purest version of ROM (various examples of mask ROM can be seen in Fig. 3.2). The requested look-up table is implemented as hard-wired through combinational logic—combinational logic is implemented by Boolean circuits; output is a pure function of the present input only, not reflecting any previous state. The result is then represented directly by its physical implementation via semiconductor electronic components and their interconnections. Mask ROM holds its name from a part of the manufacturing process—a mask has to be prepared for transposition of the circuit to the silicon by photolithography. After the surface is masked, etched, and interconnected, the data is present on the chip in a hard form, without a possibility to be changed later. On the other hand, because the data is stored in its purest form without any overhead or logic behind it, mask ROM is very efficient in cost per bit ratio. [11], [170], [135], [65]

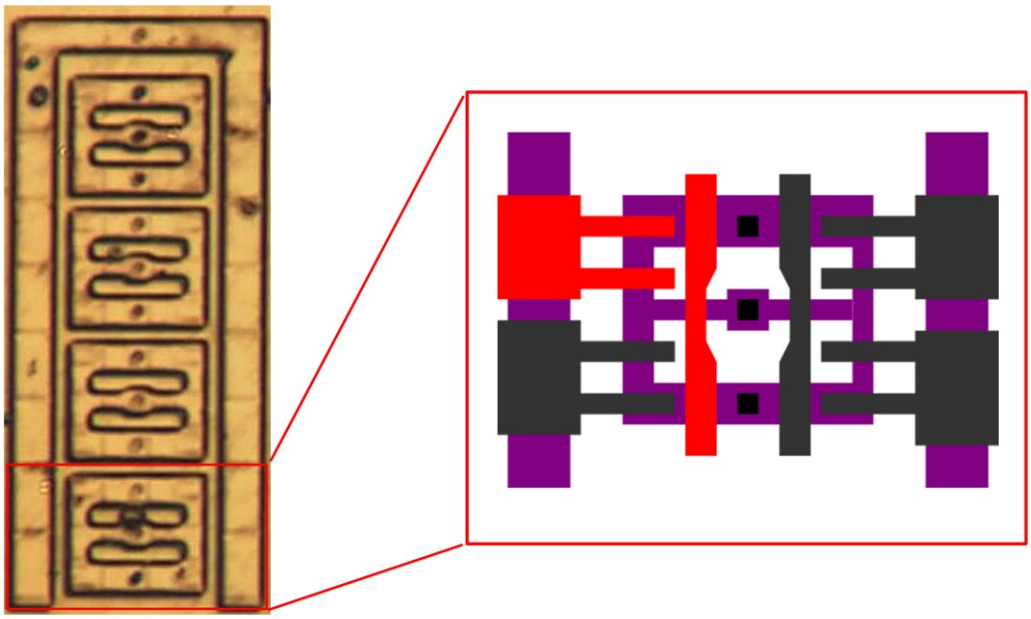


Figure 3.3: Cut-out of 2048 kB EEPROM 24C026 produced by STMicroelectronics. Left: 16-bit block. Right: 4-bit block scheme. (Source: [106])

Programmable ROM

Programmable ROM (PROM) allowed manufacturers to program ROM. Unfortunately, only once and mostly only via PROM programmer. The programming is performed by applying high-voltage pulses and thus destroying some parts of the connections permanently, usually diode fuses. This drastic step results in the final representation of data on the chip. Compared to masked ROM, the manufacturers were not undertaking such economic risk, because the programming of PROM could be made before sending goods to the market, not exactly at the time of chip production. It is possible to program only a few of the general-purpose ROM chips for testing and later update the code for the production version of the device according to the final needs. [11]

Erasable Programmable ROM

Erasable Programmable ROM (EPROM) is an expression of another step towards the more convenient type of ROM memory. The need for repeatedly re-writable ROM resulted in EPROM, sometimes also called UV-EPROM, because of using UV light for data erasure. A common sign of chips containing EPROM is a fused quartz window over the chip allowing the entrance of strong UV light. EPROM is the first type of ROM based on the floating-gate transistor (principles of the floating-gate transistor will be described later in Chapter 3.1.2) that allows storing electric charge without being powered on. As exposure to UV light affects the whole chip, it is clear that there is no possibility of focused partial erasure of the EPROM. This might be seen as a disadvantage, together with the time needed to reliably erase all parts of the memory. Moreover, the demand for special UV lamps and necessity to detach the ROM chip from a product to expose it to the UV light makes it even more complicated. [11], [170], [135], [65]

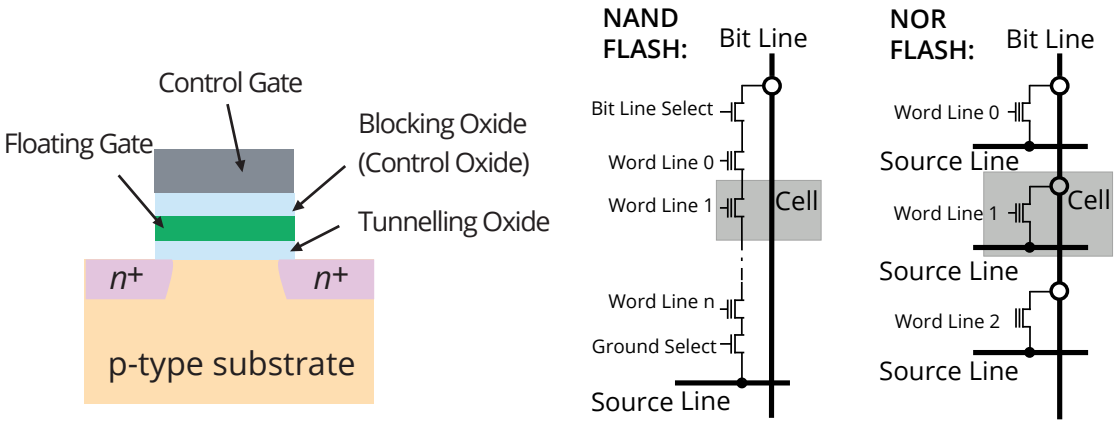


Figure 3.4: Flash memory. Left: Floating gate transistor illustration. Right: NAND Flash and NOR Flash memory cells arrangement. (Sources: [164], [116])

3.1.2 Hybrid Memory

This group emerged from the evolution of the original ROM and RAM memories. As evolution progressed, the clear boundaries between the original groups became fuzzy, and a new group was formed. The hybrid memory is combining features of ROM and RAM classes. It is non-volatile, like ROM, and data can be written/erased as needed, like RAM (although not as fast as RAM). [11]

EEPROM

Electrically Erasable Programmable ROM (see Fig. 3.3) symbolizes the top-level in evolution pyramid of ROM memory types. The same as the predecessor—EPROM—it is based on floating-gate transistors, cells are made of NOR gates. Unlike EPROM, this memory type allows in-circuit rewriting of data without any extraordinary effort. Data retention time is typically in tenths of years. EEPROM basically removes all main negative aspects of the original read-only memory. The last improvement above EPROM enabled data updates at customers' side just with appropriate software without recalling products back to the manufacturer. Erasing and writing new data into EEPROM is slow though and normally requires higher voltages. Therefore, the aim is to use this memory in smaller sizes at places where it is not necessary to rewrite the data very frequently.

Flash Erasable Programmable ROM might be considered a subset of EEPROM. Nevertheless, there are significant differences between these two memory types. Because original EEPROM is in general byte-wise oriented, built on NOR gates, its architecture has to be naturally more complex to allow this fine access to transistors. Price rises, of course, with cell complexity and so flash memory is cheaper and less spatially demanding—flash is a block-wise oriented, mostly built of NAND gates. On the other hand, flash EPROM offers very often less erase cycles endurance—typically from tenths of thousands up to hundreds of thousands. The recent development of flash memories and wide expansion into many exposed spheres like, e.g., hard drives, removable memory devices, supports its heavy evolution. After all, there are also recently emerging technologies that will most probably outperform flash memory in the near future. It only has to be waited out for the price to

drop on these innovations. For example, intel's 3D XPoint is much faster and enduring in terms of erase/write cycles. [11], [170], [135], [65]

Flash Memory

Physical design and key features of flash memory are established on very basic elements, NAND or NOR cells. Information is stored in the floating-gate transistor (see Fig. 3.4). This special type of transistor forming the NAND or NOR cells can hold a charge in the floating-gate layer for a long time without being powered on. The floating-gate layer is insulated from both sides by oxide layers. Such a design prevents spontaneous charge dissipation from this layer. The question is then how to transport the desired charge to this layer and how to remove it from there on purpose? There are two commonly used techniques—quantum tunneling and hot electron injection. With each type of cells, there has to be a different approach applied for reading, erasing, and writing data. Moreover, each charge break through the barrier damages it a bit. Therefore, flash memory suffers from a limited amount of erase/write cycles.

Cells can be of the following three types—single-level cell (SLC, 1 bit of information per cell), multi-level cell (MLC, 2 bits of information per cell) or triple-level cell (TLC, 3 bits of information per cell). The less information per cell, the lower predisposition for error occurrences, higher speed, and also higher cost. Each logic level is most often represented by different charge level in the cells. That means for SLC is needed to have two voltage states—high and low, for MLC, we already need four states, etc. This puts more demands on the logic with more levels of charges. On the other hand, the integration density allows us to produce high capacity media with very small dimensions.

The early stages of flash memory were using mainly NOR-type, because of the original purpose of use—replacement of ROM memories, or rather EPROM memories. NOR type based flash cells provide the possibility to access each cell at any time, also called Random Access. In contrast, erase and write speeds are low. Its suitability is therefore predominantly for rarely erased or programmed media.

NAND flash memory appeared on the scene in later stages. It brought a higher amount of erase cycles and better storage density over the NOR type. On the other hand, non-existence of Random Access to each bit and a necessity to read/write data in blocks predetermines NAND flash for different use cases, i.e., as a secondary data storage. [11], [164], [116], [170], [135], [65]

NVRAM

Non-Volatile Random Access Memory is a memory that does not require constant power to retain data. With advantages of speed and endurance belonging typically to Random Access Memory, the RAM cells can be easily supported with a battery to remove its volatility. Such a non-volatile RAM is in general called NVRAM (Non-Volatile RAM). A widely used contemporary representative of NVRAM is Flash memory. Flash memory can be considered a successor of EEPROM or battery-powered SRAM. Flash memory is basically a subset of EEPROM, precisely EEPROM with small blocks R/W ability that makes it much faster and usable in a variety of use cases. Unlike the mentioned predecessors, flash memory is not as fast as the other two stated types. Nevertheless, the speed decrease is not so drastic, especially with the use of NOR cells. Also, the endurance of flash memory is limited—flash cells are able to bear only a certain amount of write cycles. The endurance can be as little as 100 erase cycles, however, with recent flash chips it may be lifted up to 1 million cycles.

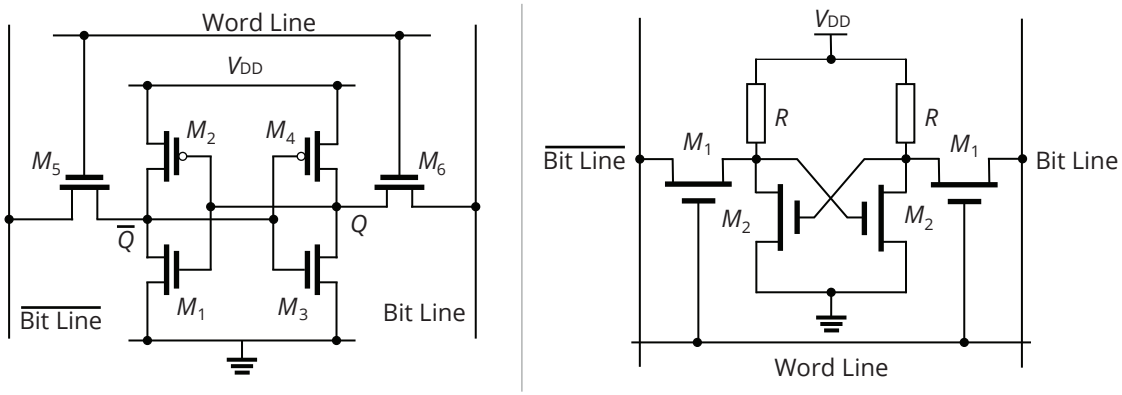


Figure 3.5: Static RAM cells schematics. Left: Standard SRAM cell made of 6 transistors. Right: SRAM cell made of 4 transistors and 2 resistors. (Sources: [65], [173])

Other features of flash memory outweigh all its disadvantages. Namely the possibility to read and write data in small blocks, shock resistance, high temperature, and pressure tolerance, etc. [11], [170], [135], [65]

3.1.3 Random Access Memory

Classic Random Access Memory (RAM) is a volatile type of memory. In general, RAM is much faster in reading out data than hybrid memories, ROM and its derivatives. There are two main sub-groups of RAM—static and dynamic RAM. The main difference between the groups is in their construction and in the way they store data. More details are revealed in the following sections.

Static RAM

Static Random Access Memory is capable of holding data without refreshing it as long as a power supply is attached, therefore the name static. SRAM memory cells are very fast, but on the contrary, the conventional SRAM cell consists of 6 transistors (see Fig. 3.5). That signifies higher manufacturing price and spatial demands. There are possibilities to build SRAM memory cell with either four transistors (depicted in Fig. 3.5) or even with two transistors though. A tradeoff of the two-transistor version of SRAM is the fact that these cells consume more power. [113], [89], [11], [170], [135], [65]

In general, SRAM is used in smaller capacities, closer to computing units just for the speed reasons. Power consumption depends on the frequency of read/write operations—higher frequencies result in higher consumption. In the idle state, the consumption is relatively low. Cells can be either synchronous when synced with system or bus clock, or asynchronous when working at independent frequency.

SRAM cells can be found at various levels of caching (demonstration of SRAM cells is depicted in Fig. 3.6) or in registers of CPUs and computational units. Common property of the registers is that they are very fast and located as close as possible to the place where the stored information is needed. On the other hand, capacity is very small to allow placement directly on a chip. Cache, depending on the level, is a compromise between size and distance from the logic of a chip.

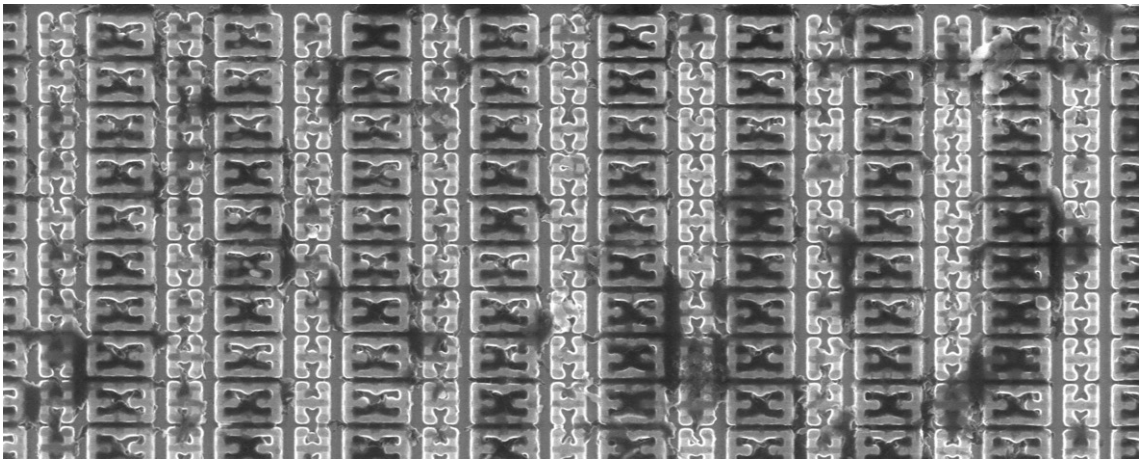


Figure 3.6: SRAM cells on the die of STM32F103VGT6 acquired by a scanning electron microscope. (Source: [173])

Dynamic RAM

Dynamic Random Access Memory requires periodic refreshing of stored values. The value is stored as a charge of a capacitor (see Fig. 3.7). Before the capacitor spontaneously discharges, the value has to be renewed, otherwise would be lost. During the refreshment phase, DRAM is inaccessible. The speed disadvantage compared to SRAM rises out of this fact, because reading or writing is possible only in specific time periods. Moreover, when reading the value stored in the capacitor, the charge is destroyed and has to be renewed as well.

On the other hand, construction of such cells is much simpler than SRAM cells. With the same integration level, DRAM capacity is much higher than SRAM. Only 1 to 3 transistors and a capacitor are needed for the manufacturing of one cell to store 1 bit. Spatial requirements are, therefore, minimal compared to SRAM. The same as with SRAM, the memory can be either synchronous or asynchronous.

Due to stated properties of DRAM memory, these cells are usually seen as the main memory in contemporary computers and graphic cards. [11], [170], [135], [65]

3.2 Logic in Hardware

Logic in a chip design can be of various types and extent. From a simple two-input logic gate, controller of memory cells, individual algorithms, up to comprehensive computational systems on a single chip (SoC). From the perspective of this thesis regarding microscopic analysis, the worst about the layout of the logic implemented directly in hardware on chips is the fact that with each technology node, there are significant differences in the appearance of the transistors. Moreover, there are differences also across producers (see Fig. 3.8).

The overall appearance of logic in a chip is mainly influenced by the used design approach. The main differences among these approaches are primarily driven by time (time-to-market) and cost limitations. In general, the design approaches can be divided into fully-custom (usually named as custom) and semi-custom. The semi-custom designs are either cell-based or array-based. Each of the approaches has its advantages and disadvantages, mostly characterized by two opposed features—costs and performance. The custom

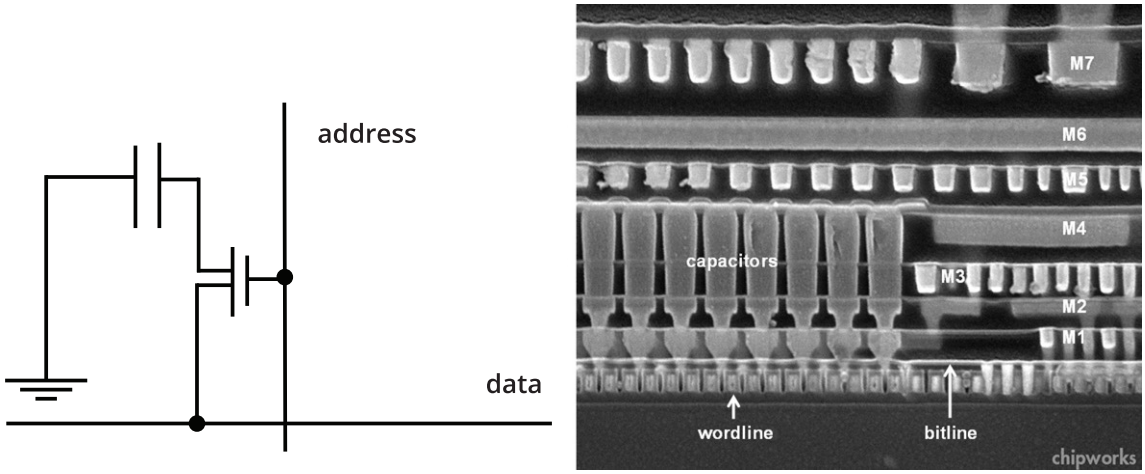


Figure 3.7: Left: DRAM cell schematic. Right: Cross-section of an Intel DRAM memory. (Sources: [65], [36])

design is the most costly; however, it can offer the best features of the final product and also a better price per single piece. On the opposite side, there is gate array-based design (or even CPLD/FPGA design) that is much better in time-to-marker (thus also better from the initial costs perspective); nevertheless, the performance and cost per chip are significantly worse than with the custom design. Finally, the standard cell-based design is positioned in the middle of the two approaches. For many applications, the gate array design is not suitable because of insufficient performance of the final products. [170], [78]

Custom design

The custom design approach is more demanding especially in early stages of the production process because the design is done transistor by the transistor in order to optimize the outcome (spatial optimization, power consumption optimization, routing optimization, etc.). The higher initial cost is paid back in high volumes of chips by lower per chip costs and also by better features of the final circuit (e.g., higher speed, better heat management).

Standard cell design

Standard cell design is a price and time oriented compromise that offers sufficient performance and utilization of the chip area for most of the use cases. Chip producers offer their customers a proprietary library that contains "standard" cells (NAND, NOR, D-latches, Flip-Flops, etc.). Out of these cells, customers can assemble their functionality and interconnections with the use of software tools available on the market. Risks for the customers using this approach are relatively low because time-to-market is short, one-time cost for the design is low as well, and the standard cells are already optimized by the producer. On the contrary, routing of the interconnects cannot be optimal as with the custom design.

Gate array design

The vendors are able to prepare a regular structure of gates organized usually in a matrix or in lines. These cells are not interconnected from the production phase. The customer then creates the interconnections in order to achieve the desired functionality. The indisputable advantage of this approach is the speed of creating a working

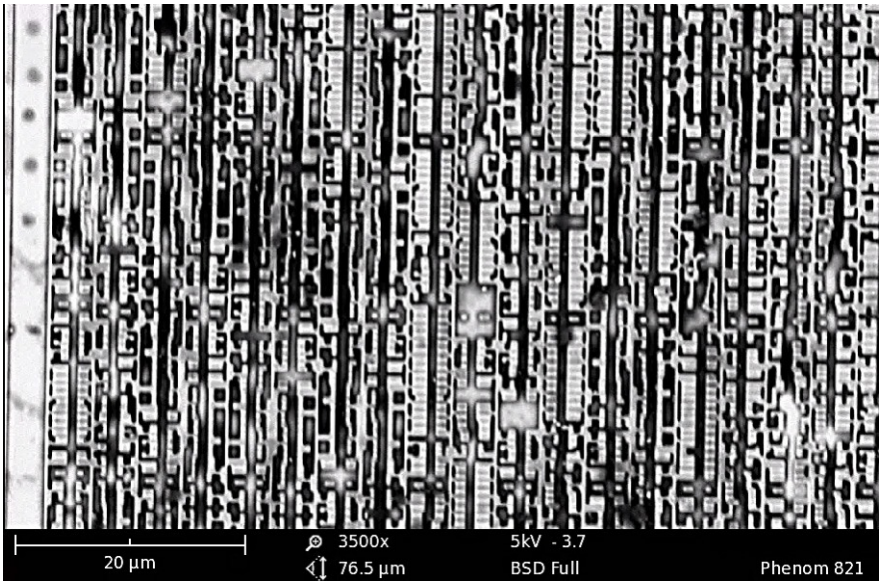


Figure 3.8: Logic cells of MIFARE Classic 1 kB acquired by scanning electron microscope Phenom Pure G2. (Source: author's work.)

product accompanied with low price—the producers sell these standard unconnected gate arrays to many customers, and thus the initial costs are spread among all the customers. The disadvantage is then non-optimal utilization of the chip surface, low speed (almost similar to CPLD/FPGA implementation) and in higher volume also higher price per piece.

On top of the mentioned approaches, there exist many components (blocks) that are designed and optimized as a whole. These parts can be used and combined with inter-block connections within a single device. With this approach, it is possible to ensure requested performance in the critical blocks and save resources by applying prefabricated parts. When we get back to the presented design approaches, it is also possible to combine the approaches in a single device to meet all criteria. [170], [78]

3.3 Interconnection

Interconnection is absolutely the striking part of each microchip. If a chip was just a silicon layer with transistors and no metal layers connecting the transistors together, it would be much easier for designers and analysts; but also, for frauds. Because of an astronomical number of transistors placed on a single chip, the interconnection is a very complex standalone discipline that is mostly done by computers, or rather by fully-automatic or semi-automatic software (see Fig. 3.9). With the very recent chips, it is virtually impossible to draw connections by hand, otherwise would take years to produce a single design. Length of interconnects influences many critical aspects of the chips, namely parasitic resistance, power consumption, thermal properties, frequency, etc. The most important task is to find the ideal connection setup for the whole chip design in order to minimize the length of all conductive lines and thus to reduce the negative impacts of its length. Making the interconnections shorter influences positively also the price of the design.

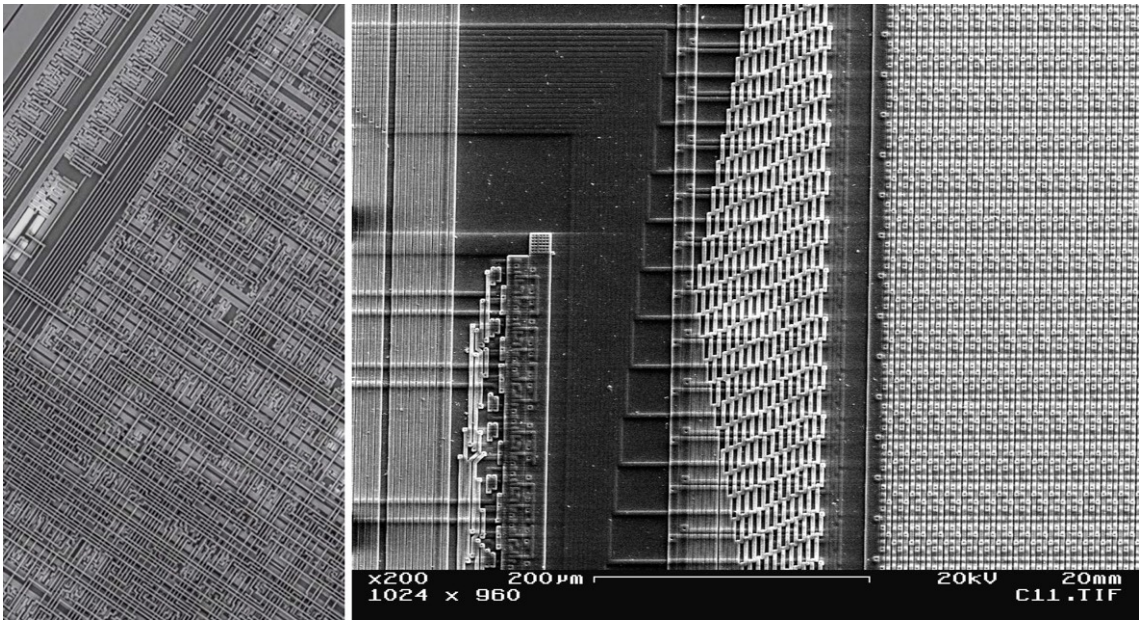


Figure 3.9: Left: Interconnection net in a chip layout. Right: Intel 486DX4 connection pyramid leading to ROM. (Sources: [5], [124])

The very purpose of interconnections is to connect power, ground, clock, and other signals to the chip and correctly route these signals within the chip to obtain requested results. Layers leading such signals are commonly called metals or metal layers. These layers are in schemes and cross-sections marked as M0–Mx according to the number of such layers in a chip, for example, in Intel's 10 nm metal stack there are metal layers from M0 to M10 (Fig. 3.10). Metal layers are split by insulation layers—generally named oxides or oxide layers. Connections between metal layers are provided by so-called vias. The most common materials for the production of interconnects are copper, aluminum, tungsten, silicide, cobalt, and a variety of compounds based on the mentioned elements. [170], [139], [78], [134]

With respecting the Moore's law [108], there is a constant trend in miniaturization of elements integrated on a chip. Followed by a tendency of increasing the average area of the chips. These two aspects combined together lead to consequent more or less linear growth of a number of transistors present on a single chip. This rapid growth requires more interconnections and also increases the average length of a single connection.

3.3.1 Types of Interconnects

Because of the substantial extent of interconnects in a single chip, the connections can be divided according to spatial aspect to:

- Local.
- Intermediate (or semi-global).
- Global.

For better illustration of this division, see Fig. 3.10. It can be observed that the further from the silicon substrate, the longer distances of interconnects are expected. A number

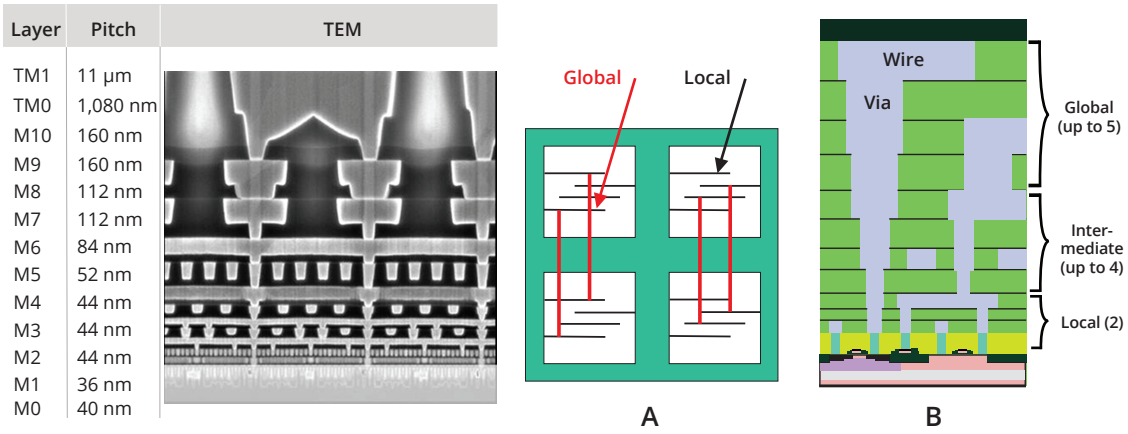


Figure 3.10: Illustration of interconnection complexity. Left: Intel 10 nm Metallization Stack—TEM of 10 nm intel chip cross-section. Middle and right: Local, intermediate, and global interconnections. (Sources: [138], [134])

of metal layers ordinarily grows with complexity and area of a chip. Nevertheless, it also depends on the chip purpose. For example, memory chips are less demanding in terms of the number of interconnects. Most of the memory cells are very simple; moreover, the memory chip structure is very regular. On the contrary, logic implemented in hardware is very irregular by nature and usually also very complex. Therefore, the design of interconnects suffers primarily with designing of the logic parts. Most of the performance requirements usually lie in these parts, where it is highly convenient to make the interconnects as short as possible to eliminate long delays. [170], [139], [78], [134]

As was mentioned before, there is also a function-wise division of the interconnects into the following basic groups:

- Signaling.
- Clocking.
- Power and ground distribution.

Width and height of interconnects depend on the technology used for chip production. The technologies are evolving very rapidly; billions of dollars are spent every year within big companies[1] on the modernization of the technologies. Formerly, each node of the semiconductor fabrication expressed the size of elements on the chip. Nowadays, nodes are rather marketing tools that show to the public an achieved next step in technology, although not describing the actual size of the placed elements. Currently, we are talking about nodes marked as 10 nm or 7 nm and in the near future 5 nm followed by 3 nm. As it can be seen, the technologies are slowly approaching the size of an atom—the size of a single atom, depending on the element, is in range 0.1–0.5 nm [151], that means 0.5–5 Å. Contemporary chips can contain hundreds of kilometers of wiring inside a single few millimeters rectangle shaped device. Moreover, because of lengths alike, one can imagine the delays of signals.

The situation deteriorates with the scaling of the technology, length of interconnects (L in Fig. 3.11 and in the following formulas) increases with complexity and spatial demands

[1] Intel Corporation, Advanced Micro Devices, Texas Instruments, NXP, etc.

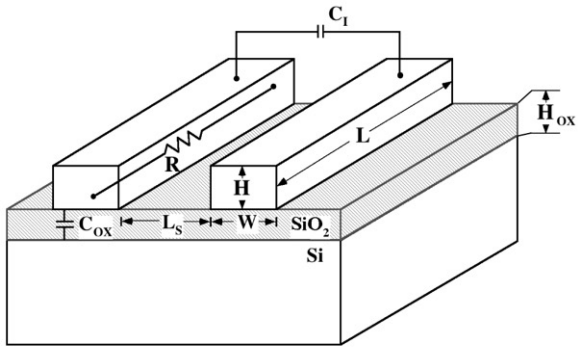


Figure 3.11: Illustration of interconnection formulas. (Source: [134])

of the chips. Other aspects such as the height of oxide layer (H_{ox}), the length of space between lines (L_s), the width of line (W), the height of line (H) decrease with higher integration. This results in a higher resistance, capacitance, and crosstalk ($X_{crosstalk}$). Deriving from Ohm's law, resistance influences current and voltage, that means power consumption and thermal properties. Capacitance influences, besides other things, the crosstalk, which means error occurrence. To eliminate crosstalk, special techniques can be used. For example, interleaving of signaling wires with V_{cc} and Gnd wires, or well-known twisting of a pair with use of differential wires (this technique requires 2n wires, which is not always possible) or so-called active noise cancellation with use of drivers adequately distributed. Crosstalk is directly proportional to the ratio of capacitance between wires to total capacitance, see Equation 3.8. [134], [170]

The rising delay can also be reduced with the use of drivers (inverters, buffers, etc.) along the path. Without the use of the drivers, the delay is increased quadratically with the length of the wire (Equation 3.1). With the use of it, only linear increase can be observed. With splitting the line into n smaller pieces, the delay is reduced as stated in Equation 3.2. On the contrary, each driver requires space, power, and correct routing of the affected wires. Moreover, each driver has also a delay, expressed in Equation 3.3 as τ_G; this equation describes the whole wire delay. [134], [170]

$$\tau_L = \frac{3.56\kappa_{ox}\varepsilon_o\rho}{\lambda^2}L^2 \tag{3.1}$$

$$\frac{\tau_L}{n} = \frac{3.56\kappa_{ox}\varepsilon_o\rho}{\lambda^2}(\frac{L}{n})^2 \tag{3.2}$$

$$(\frac{\tau_L}{n} + \tau_G)n = \frac{3.56\kappa_{ox}\varepsilon_o\rho}{\lambda^2}(\frac{L^2}{n}) + n\tau_G \tag{3.3}$$

The material used for building oxid layers is expressed by relative permittivity (κ_{ox} and ε_r) in Equations 3.6 and 3.7. Material for wiring is represented by electrical resistivity (ρ) in Equation 3.4. Total capacitance (Equation 3.5) is composed of the following addends—capacitance between wires within a single layer (C_I) and capacitance between layers (C_{ox}). For illustration, see Fig. 3.11. [134], [170]

$$R = \rho\frac{L}{S}; S = WH \tag{3.4}$$

$$C_{total} = C_I + C_{ox}; C_{ox} = C_{lower-ox} + C_{upper-ox} \tag{3.5}$$

$$C_{ox} = \kappa_{ox}\varepsilon_o \frac{WL}{H_{ox}}; \kappa_{ox} = \varepsilon_r = \frac{\varepsilon}{\varepsilon_o} \tag{3.6}$$

$$C_I = \kappa_{ox}\varepsilon_o \frac{HL}{L_s} \tag{3.7}$$

$$X_{crosstalk} \propto \frac{C_I}{C_{total}} \tag{3.8}$$

Let us consider physical quantities occurring in the equations above. As stated in the previous paragraph, the only chance to reduce the negative influence of interconnects delay is to upgrade the materials that the metal and oxide layers are made of. There are possibilities for how to tune conventional material to have better properties with regard to the desired state. Alternatively, completely new technologies will take place, i.e., optical wiring and using light for signaling. However, it is also important to mention that the local wires at lower metal levels are getting shorter (see Fig. 3.10), with lower delays. The speed at local block connections is the most important for the overall chip performance. With the thoughtful design of higher levels of interconnection among the basic blocks, good final results in performance are achieved. [134], [170], [139], [78]

Chapter 4

Attacks

Despite the fact that this chapter is named Attacks, it does not mean that all types of examinations of chips are with a malicious purpose. However, let us call all such actions "attacks" in the following text, regardless of its purpose. A chip can be attacked, e.g., during the manufacturing process, then it is rather called quality assurance, verification or similarly. Nevertheless, the used techniques are not far from the ones used when performing the real attack. The difference between a real attack and some kind of quality assurance is in very limited knowledge about the device in the case of the real attack. During the real attack, the chip can be considered a black box, and therefore, the real attack is the more demanding task.

After the past decade, where there were few successful attacks publicly performed, i.e., MIFARE Classic reverse engineering and its PRNG crack [115], the chip producers pay much more attention to employing various types of sophisticated countermeasures against a wide variety of attacks. Consequently, performing some of the attacks is becoming extremely difficult, especially with the recent chips that are used for security-oriented applications. "Normal" chips can be still produced with the use of not so recent fabrication technology, and thus observations are much easier than in the case of the cutting-edge nodes. Protection employed in such chips is sometimes virtually none. Nevertheless, our focus is aimed at chips that are used in biometric passports, id cards, and similar personal documents (i.e., Czech e-Passports, German ID cards), where the need for protection is essential.

Attacks in general or the attack scenarios can be a very broad topic; thus we will build this chapter primarily with respect to the goal of this doctoral thesis—the microscopic analysis and the physical security of microchips, especially concerning chips typical for smartcards.

4.1 Motivation

There are various types of attacks supported by disparate motives. First of all, it has to be understood why there exist any type of attack on chips. Let us look at the incentives that lead frauds to attack the chips. [126], [61], [129], [26], [142], [166]

Intellectual property protection

As there are more companies producing chips and competing with each other, the companies want to make sure that the competitors have not infringed their patents. For this reason, it is not enough to blindly believe statements of rivals that they do not misuse intellectual property (IP) of others. Analysis of the competitors' products

is made to prove that there is no reason to sue the other party, or to collect evidence for trials regarding IP violation.

Creation of unlicensed duplicates—overbuilding
The overbuilding is an effort to quickly and cheaply gain precise knowledge about a certain product to produce the same or similar unlicensed copies. This action is performed by some suppliers of the microchips. The process of designing a chip is very costly, and so copied chips can be sold for a significantly lower price or with much higher commercial margins.

Verification of own production
Due to the fact that there are many IP cores used as building blocks of larger aggregated units, and because of production being moved to economically convenient locations, each producer has to verify that the products meet the given requirements. Especially in the production of security-related chips. The customer has to verify whether there is, e.g., no backdoor, no hardware trojan, etc. This task may be very complex and expensive as it means to verify the device thoroughly.

Competition scanning
Each company wants to see how far the competitors in the implementation of new technologies are. As each company analyses its own products because of needed quality verification during the manufacturing process, the third party products are also analyzed in the same way. This kind of mutual examination contributes to the general evolution of technologies. It is used not only among commercial subjects but also in the military sector. Typical military technological espionage helps to keep balance in the progress of high-tech technologies and inventions. This kind of analysis is usually done undercover.

Security auditing
When introducing a security-related chip to the market, performing an independent security audit might be a way to build trust in the product. Some of the use cases may need a proven level of security that cannot be guaranteed just by a statement of the producer.

Misuse of victims' hardware
Attackers may need to use a set of hardware devices that they either cannot obtain or price for the devices is enormously high. In such cases, finding a way to circumvent security mechanisms and seize desired hardware without alerting the victim might be an acceptable solution.

Theft of secrets; theft of services
The attacked hardware, is in many cases, an access key to some service. It can be proving identity, entitlement to use some application or ownership of some secret. If the attacker finds a way how to gain such an asset, the targeted service may be used by an unauthorized user. We can even talk about identity theft in specific cases of stealing secrets. Let us state one more example—multimedia providers are protecting streaming of the provided media by some form of encoding, i.e., TV providers. When this type of protection is cracked and keys spread within a community (commonly a vast community), the losses can get to serious numbers.

Denial of service
Knowledge of potentially available vulnerabilities or weaknesses in hardware implementation can lead to a denial of service attack that prevents users from using services that they are authorized for. There are many use cases where this attack might be convenient for the attackers—delaying access to monetary resources, denial of access to critical information, etc.

Academic purposes
The same as the purpose of this thesis, various types of academic explorations can lead to attacking or reverse engineering a chip or its part.

Missing documentation
This motivation is not very likely to happen with contemporary chips, because these are manufactured in the corporate sector, where all parts of the process (including design and documentation) are scrupulously protected and backed up. Nevertheless, some historic straightforward designs or designs of enthusiasts may be subject to missing documentation.

Other personal interests
There may be no other background for performing an attack than gaining recognition in the community or only attacking devices as a hobby because of one's curiosity.

4.2 Attacks Classification

Attacks on chips can be divided into three main groups according to the degree of tampering with the chip—invasive, semi-invasive and non-invasive. The differences among the attack classes can be intelligibly represented by the following fundamental aspects—cost, time, expertise. Generally, invasive attacks are the most demanding in all aspects, whereas non-invasive attacks are the most undemanding ones. In between these two classes, there is the recently defined class of semi-invasive attacks [148]. This class naturally emerged to fill the gap between invasive and non-invasive attacks. Let us introduce each class separately with typical examples of possible attacks belonging to them. In this chapter, we will focus primarily on the invasive and semi-invasive attacks because of the orientation of this thesis. [148], [127], [126]

In real attack scenarios, it is very common to combine several attack approaches across the classes. Performing strictly one type of attack is often very difficult, especially without complementary information and hints that can be gained from the other attack types.

4.3 Invasive Attacks

Invasive attacks are undoubtedly the most demanding ones. These attacks require several disciplines to be combined with a significant level of expertise—especially chemistry, physics, and computing. Therefore, carrying out such complex experiments requires either various experts to be available or a single general-purpose scientist capable of all these. The second mentioned has become almost impossible with recent technologies. Equipment needed for proceeding with standard steps is also very expensive and requires regular upgrades to keep up with the unflagging introduction of new technologies among chip producers. There exist many experiments performed in the past decades that claim that also second-hand equipment obtainable for about ten thousand dollars might be sufficient for invasive attacks. It is

correct, partially—some attacks can be performed with a second-hand optical microscope and some self-made positioning system with microprobing stations. However, considering state-of-the-art security chips, such invasive attacks are hardly possible in low-cost. Producers have already implemented appropriate countermeasures to make the attacks extremely difficult to happen. Moreover, die size has grown several times in general, and the integration level has been increased as well. That results in a completely different level of complexity that requires recent equipment (in other words, expensive equipment). [127], [148], [126]

For the successful performance of an invasive attack, it is recommended to have several samples of the same chip available. First of all, because of risks connected to carrying out the attacks—a specimen can be accidentally destroyed. Secondly, some steps are destructive and non-reversible, that means a few samples have to be sacrificed in order to read the next steps of the attack.

Before starting any invasive attack, the chips have to be decapsulated (see in detail in Chapter 7.2). Different approaches to decapsulation can lead to different types of attacks. Sometimes we want to etch out only a part of the packaging that is directly above the chip surface with preserving all wiring in the package. This allows us to connect the chip into a circuit and see it in operation with the advantage of having access to the surface with, for example, microneedles. For more destructive attacks, like microscopic analysis, where the aim is to acquire images of all layers, the complete decapsulation is performed.

4.3.1 Microscopic Analysis

Microscopic analysis belongs to the invasive attacks—it requires chip decapsulation and decomposition. As standalone, we could describe it as observing the layers' surface with the use of any type of microscope with the aim to understand the visually ingested representation of the structures. In general, microscopic analysis is a part of all the attacks mentioned below. In other words, observation of a chip with the use of a microscope is an inseparable part of the attacks. The closest connection can be seen, especially with reverse engineering, where the microscopic analysis plays a key role—acquisition of images of all layers with a consequent understanding of the structures. The way of understanding the structures is nowadays heavily aided by software—for more details, see Chapter 7.5 and Chapter 9. We describe the whole process of microscopic analysis in detail in Chapter 7.

The microscopic analysis itself can be divided into a few branches, each differing in the purpose of doing such analysis. Nevertheless, it is very often only a part of a wider analytical process. Microscopic analysis is being used on a wide range of devices, from very simple integrated circuits representing a basic functionality, e.g., logic gates, adders, multiplexors, to very complex chips containing more separate units on a single chip (smart cards, cryptographic devices, CPUs, GPUs, etc.). General-purpose chips (CPUs, GPUs, DSPs, etc.) cannot be analyzed in the same manner as specific purpose chips, because the general-purpose ones do not express the behavior (functionality) of an application directly in hardware, but in the form of general instructions used to form different programs. Therefore, general-purpose chips are not among the primary aims of this thesis. [127], [148], [126], [30]

Branches of Microscopic Analysis

Microscopic analysis can be used for different purposes—fault analysis, system-level inspection, or process analysis. Nature is always the same. First, the samples have to be

prepared for observation (x-ray, removing a package, deprocessing[1]); then, scanned with the use of an appropriate type of microscope, and finally, the acquired data has to be processed and analyzed—layer images, cross-section images, EDX (Energy-Dispersive X-ray spectroscopy), etc.[160], [159].

Fault analysis
Fault analysis is mainly performed by manufacturers and designers of chips during the manufacturing process to assure the quality or to discover faults emerging unintentionally somewhere inside the process.

System-level inspection
System-level inspection is carried out because of, e.g., patent protection, competition analysis, (partial) circuit extraction, security assurance, etc. Put simply; it is a process that is looking at a chip as a working system and tries to reveal the physical structures with an understanding of the functionality. This branch of microscopic analysis is the main area of interest in this doctoral thesis.

Process analysis
Process analysis is a kind of analysis that provides information about the structure of the specimen, used materials, and manufacturing techniques. This approach can be useful for system-level inspection as well, especially for the phase of preparing specimens.

4.3.2 Microprobing

One of the most important invasive attacks is microprobing. It requires a lab spot with a microscope with long working distance objectives, stage, device test socket, micromanipulators, and microprobe tips, all also called microprobing station as a whole. Such devices vary in provided features and naturally in price that can ordinarily oscillate from five thousand dollars to up to several million. A brief overview of a microprobing station parts follows. [86], [146], [127], [148], [126]

Microprobes
There are two types of microprobes, active and passive. Each type is suitable for different microprobing tasks. For simpler use cases it is needed to accommodate 2–8 microprobes at the same time.

Passive tips are very often connected directly to an oscilloscope, thus having low impedance and high capacitance. These tips can be used for signal injections or for eavesdropping of buffered internal buses or outgoing signals at bonding pads of a decapsulated chip. [86]

For internal signals (the unbuffered ones), the active probes are needed. The active probes are equipped with a FET amplifier nearly at the end of the tip. Characteristics of the active probe are exactly the opposite of the passive one—high impedance and low capacitance.

Stage and micromanipulators
Stage and micromanipulators can be either manual or automatic. Purpose of these moving devices is to precisely navigate the stage, the specimen or the tips of probes to

[1] Sometimes also called delayering.

the desired position. With very standard devices, we usually talk about sub-micron precision of the movements. For smaller scale occasional tasks, the cheaper manual version might be sufficient. Nevertheless, for regular work with recent technologies, fully controlled device with better precision is highly recommended.

Objectives
Long working distance of objectives is necessary in order to be able to place probes under the objectives. Some stations have only one objective; some have a very typical revolver stand with objectives providing different magnitude and focus depth.

Device test socket
The device test socket is a socket connected to a computer or other controlling mechanism capable of providing basic signals (CLK, reset, voltage, GND, and input data). It is possible to interact with the chip in the desired way through the use of the socket—sending input values to the chip and observe its behavior.

For successful microprobing, it is necessary that the observed chip remains functional. Therefore, the appropriate procedure of specimen preparation has to be carried out to preserve the chip bonding and potentially also the chip packaging or its part. The second option is to bond the chip to a test package after removal of its original package. Use cases of the microprobing vary from verification of correct functionality over eavesdropping of few signals up to copying memory content. The microprobing itself is a process of connecting a chip to an external device capable of reading out signals or injecting signals with adequate timing. Subjects of this attack are in many cases buses that transfer information stored in memory or computed by a processing unit. Various techniques are used in order to make the microprobing harder—bus scrambling, data encryption, hiding data busses deeper in the structure of the chips, etc. [86], [146], [127], [148], [126]

A common procedure is to identify key parts of the chip and their connections—memory, processing unit, buses. It is usually necessary to partially or fully deprocess a few specimens (or to inspect it with a FIB) before we gain the knowledge needed for proceeding with the connection of the probes. With older chips, the needles can be attached to top levels; therefore, no special treatment is needed, except removing passivation layer (either completely or better only on contact points). Nevertheless, with contemporary chips containing many metal levels and protection mechanisms, it is needed to locate precisely the points of interest. Security by obscurity principle may sometimes be sufficient to hide buses and important contact points deep inside the chip structure. After the location of the targeted lines, the researcher has to drill a hole in the chip with laser or recently more often with the FIB (Focused Ion Beam) that allows aimed attachment of probes to the chosen place. If needed, the hole can be filled, for example, with platinum in order to prepare a convenient pad for comfortable probe landing. In the vast majority of cases, one probe is not sufficient to read the whole data set. It is usually possible to attach a subset of targeted contacts and with repetition of the same process read the whole data set step by step.

4.3.3 Circuit Manipulation

Circuit manipulation is one of the compelling techniques in the chip production industry, helping to speed up the time-to-market process. Circuit manipulation attacks were naturally derived from the possibility to edit circuits in the production process. The FIB technology provides significant power in performing accurate observations and modifica-

tions of microchips, reflecting the very recent manufacturing processes. [141], [127], [126], [136]

Focused Ion Beam

The FIB has been coming to the main stage of chip production as a very powerful tool. Unfortunately, the same power is also offered to the other side, to hackers. In production, FIB is used for rapid cost reduction in the prototyping process, in QA departments, etc. Formerly, each proposed change during the production cycle led to the creation of a new mask and carrying out the whole process from scratch again, which is not only time consuming, but also very costly. Nowadays, when there is a bug, optimization, or any different change needed, producers are able to rewire the prototype chip virtually immediately for a fraction of formerly expended resources.

The focused ion beam is a technique similar to Scanning Electron Microscopy (SEM). SEM is used for non-invasive observations, whereas FIB allows observing of the specimens (at low ion beam currents) and also editing of them (at high ion beam currents). As the main difference, SEM uses an electron beam, whereas FIB uses ion beam—ions are much more massive than electrons. Exposing a specimen to an electron beam causes reflections of these tiny particles. With the use of ions at high currents, the top layers of atoms are sputtered out of the specimen. Moreover, the FIB offers deposition of both conductive and also dielectric materials. This all is provided at nano-scale. The presented features of FIB allow rewiring of ICs with very little effort, of course, to some reasonable extent of the rewiring. Contemporary FIB workstations are capable of dealing with 10 nm nodes, i.e., OptiFIB Taipan Focused Ion Beam for Semiconductors from FEI[2]. [141], [127], [126], [136]

The biggest issue for hackers is the potential unavailability of FIB workstations, because of their price. Nevertheless, clever attackers can rent a FIB device for hourly rates in lower hundreds of dollars[3]. That can make the attacks feasible also for lower costs.

FIB Aided Attacks

Academic research has proven a few hypotheses and laid out only a few potential scenarios so far. The reason behind this lack of research attempts is that carrying out practically some of these FIB aided attacks requires knowledge of the attacked device and primarily regular access to a FIB device. After the successful investigation of the specimen, the FIB can be used precisely according to the results of the analysis. [86]

The target for FIB aided attacks is often to disable chip security mechanisms that are ordinarily integrated into modern security chips. This technique can be called FIB microsurgery. Despite it sounds easy, disabling a security mechanism requires deep knowledge of the attacked chip—attacker should reveal its layout with the use of reverse engineering. The countermeasures employed in the examined device should be mapped during the reverse engineering process. It is sometimes not possible in a simple way, because the chips can contain safeguards that are able to destroy parts of the IC after digging deeper. The safeguards can protect security features as well as the defensive mechanisms themselves.

Let us assume that we have enough information about a chip. Then the FIB workstation comes to light to disable intrusion detection, to remove protection technologies, to reroute

[2] https://www.fei.com/products/fib/optifib-taipan-for-semiconductors/

[3] Hourly cost of using FIB is £240 on this website: https://store.lboro.ac.uk/product-catalogue/loughborough-materials-characterisation-centre/electron-microscopy/focused-ion-beam-dualbeam-per-hour-charge [May 2018]

signals, to mill holes through chips' layers, to prepare pads for microprobing, etc. If the attacker is successful, the next step of the analysis begins—gaining and understanding the data. FIB aided attacks have the potential to serve in top-secret military analyses and thus can have a substantial impact on overall global security.

There arise attempts for new use cases of FIB technology in attacking chips, presented, for example, in [153], where researchers from TU Berlin accessed an older CPLD (Complex Programmable Logic Device) chip from its backside, milled holes with the use of FIB close to transistors and tried to observe infrared spectrum of the chips in use. The installed IR camera recorded the so-called heat signature of the encryption algorithms. This technique could give attackers insight into the programmed structure of programmable logic devices, where reverse engineering does not provide much relevant information.

As a summary, it can be stated that FIB aided attacks are very powerful, opening possibilities to hack virtually every chip. [86], [141], [127], [126], [136]

4.3.4 Reverse Engineering

Reverse engineering (RE) is a complex process of understanding the structure and functionality of an IC. Complexity varies from one chip to another depending on many factors, especially the purpose of the chip, the extent of its functionality, and used manufacturing technology. With chips using standard components, e.g., standard memory cells, IP cores, the tasks may lay in identifying these blocks and interconnections. On the contrary, with custom designs like ASIC (Application Specific Integrated Circuit), the whole field of transistors and their interconnections have to be understood. The target of RE dictates demands—reading masked ROM contents belongs to easier tasks, whilst complete reconstruction of a contemporary security chip takes weeks or rather months of work of a team of researchers. At the end of the complete RE, the outcome can be, for example, a netlist of the device. [160], [127], [88], [30], [57], [21], [166], [26], [32], [40], [128], [168], [126]

Common Procedure

In general, the process for performing RE can be shrunk to a few abstract steps that are then specifically performed in each particular case. First of all, a decent number of chips has to be obtained (step 1—obtaining chips). We intentionally talk about decent number, because as was stated before, the complexity of RE varies from chip to chip and also from task to task. For some tasks, only a few samples might be sufficient; for others, a dozen samples might not be enough. The chips are in the vast majority of cases encapsulated in a package, mostly plastic package. Therefore, the decapsulation process has to take place (step 2—decapsulation), which is nothing else than removing the whole package or its part. After that, bare chips are ready to be delayered (step 3—delayering) one-by-one in reverse order compared to the fabrication process. Each layer has to be documented with the use of an appropriate microscope (step 4—acquiring images). The last phase is the main analysis of the images containing, for example, recognition of logic blocks, reconstruction of the original blueprint, tracing interconnections (step 5—postprocessing and analysis). The steps needed for RE are described in detail in Chapter 7 and in Chapter 8.

Typical Targets

Although the incentives for performing RE may sometimes be non-standard, let us mention at least the common ones with simple examples providing better understanding. [160], [127], [88], [30], [57], [21], [166], [26], [32], [40], [128], [168], [126]

Read ROM content

In former implementations, secrets (chip ID, password, access code, etc.) could be stored directly in ROM memory (in the worst case in mask ROM). This happened because of the belief of manufacturers that whatever was hidden in hardware, especially inside something so tiny as a microchip, could be considered safe.

Understanding contents of ROM can also be a key to decryption of memory bus for reading out secrets stored in visually non-readable memories. Although using mask ROM memory in security chips is an abandoned idea (for the reasons stated above) there are still many chips holding various information just in the mask ROM memory.

Read non-ROM content

Nowadays, it is well-known that visually observable storage of information is a severe security risk. Therefore, information is stored in memories that have cells with a visually stable structure regardless of the value stored inside. There can be several ways on how to gain this information. These approaches always require a certain level of RE; in other words, a certain level of understanding the IC.

Some memories can be read just with simple microprobing of the memory bus and reading out unencrypted data; some can be accessed through the processing unit after decryption of the content stored in the memory. Nonetheless, the attacker has to understand the essential principles used in HW, e.g., addressing instructions and data, the common process of boot-loading, and so forth. There exist different attempts of accessing contents of memories, e.g., [133], [146], [105], [16], [145], that are more or less dependent on RE.

Discovering HW design flaws exploitable for cheaper attacks

Because invasive attacks are the most demanding ones, searching for flaws that might be potentially exploitable on a different level of interaction with the chips (on semi-invasive or non-invasive level) is a common target for reverse engineering. As with the case of MIFARE Classic, there are still tons of cards spread among users all over the globe. In order to prepare a quickly usable non-invasive attack against so popular MIFARE Classic, RE was performed with positive results [115]. Unfortunately, this is not the only case, also the successor card—MIFARE DESFire—was successfully attacked [120].

Reconstruction of an algorithm implemented in HW

Many chips carry on their surface hardware implementation of some special functionality. The best reason for implementing some function directly in hardware should be performance optimization. Unfortunately, in many cases, authors intentionally tried to hide the physical implementation as a security measure. Security through obscurity is violating essential Kerckhoffs's principle [84].

Reconstruction of the whole IC

Reverse engineering of the whole device is undoubtedly the final escalation of difficulty

within RE. All steps of RE have to be fully mastered and performed in such a way. It is easy to imagine a simple flip-flop or memory microchip reproduction. Compared to that, cloning of contemporary Microcontroller Units (MCU) or Systems on Chip (SoC) with specific design and many anti-RE measures is tedious even for the very professionals.

Gaining configuration bitstream file from a CPLD or an FPGA chip
When it comes to an understanding the functionality of these general-purpose chips that allow programming of their functionality, gaining access to the bitstream configuration file is essentially equivalent to hardware RE. Obtaining such data is a matter of understanding how the configuration data is stored in the programmable device or from where it is loaded into the device after losing power. Then, these memory elements or buses used for the data transfer are usually subject to the attack.

4.4 Semi-Invasive Attacks

Semi-invasive attacks category was defined in the previous decade in [149] and later further described in [148]. It can be noticed that this term is slowly being accepted by the community. Let us adopt this classification as well. By definition, semi-invasive attacks still require access to a chip's die (require full or partial decapsulation), nevertheless the attacks are not penetrative. In other words, there is no need for delayering of the chip or for milling holes to create connections to lower layers. Avoiding delayering means, in fact, avoiding one of the most tedious sub-tasks of invasive-attacks. Subsequently, cost demands are rather approaching the cost level of non-invasive attacks. As the invasive attacks become more resource demanding, the semi-invasive group turns to be interesting even more. However, the semi-invasive attacks still maintain a certain level of complexity, especially due to the need to precisely locate the spot for the attack deployment.

In general, semi-invasive attacks aim at keeping the chip surface intact. Thus the chip can stay fully operational. Moreover, the passivation layer still protects the chip against unwanted interaction. This gives us, for instance, the following possibilities to interact with the examined chips—IR/UV light, X-ray, laser, electromagnetic field, heat emission, fault injection, etc. This class can be further divided into passive semi-invasive attacks and active semi-invasive attacks. The target of passive attacks is mostly to read some information, commonly memory content. Whereas the active sub-group typically induces faults in the device, as mentioned, with X-ray, electromagnetic field, light, etc.

There are many techniques used primarily in fault analysis area that are based on the interaction of electromagnetic waves or particles with the chip surface or even with levels inside the chip body. These techniques can be used for building semi-invasive attack scenarios, e.g., LIVA (Light-Induced Voltage Alteration), TIVA (Thermally-Induced Voltage Alteration), CIVA (Charge-Induced Voltage Alteration), LECIVA (Low Energy CIVA), OBIC (Optical Beam Induced Current), OBIRCH (Optical Beam Induced Resistance Change), EBIC (Electron Beam Induced Current), EBAC (Electron Beam Absorbed Current), EBIRCH (Electron Beam Induced Resistance Change), LVP (Laser Voltage Probing), LADA (Laser Assisted Device Alteration). [6], [7], [27], [28], [95]

4.4.1 Passive Attacks—Backside Imaging

Because the chips consist of more layers and observation of the top layer usually does not disclose enough information, there are also different possibilities to be performed, e.g., X-ray or backside imaging.

Backside imaging complies with the definition of the semi-invasive attacks because, except decapsulation, there is no need for further destructive tasks. When accessing the chip from the backside, the first layer is the thick silicon substrate ended with transistors. Moreover, silicon is nearly transparent to photons with wavelengths >1100 nm [148]. In recent chips, use of highly doped silicon that is less transparent to IR light [53] requires either more intensive light sources or an IR camera with high sensitivity (or both intensive light source and sensitive IR camera).

This technique has its limits set by the used technology. Because IR light with long wavelength (>1100 nm) is used for illumination, resolution of the observation apparatus is lower. The employed IR camera also influences resolution and overall. It was practically tested that chips built with technology nodes down to 0.8 μm are observable also with this kind of setup [148]. However, the current nodes around 10 nm are thus out of the scope of this technique.

Apart from observing transistors, researchers might be able to read masked ROM contents. This can be convenient especially in cases when the ROM is covered with a shield metal layer preventing standard observation from the front side. [96], [148], [136]

4.4.2 Active Attacks—Fault Injections

The fault injection attack principle is not dissimilar from the UV EPROM erasure mechanism. The original scenario was using UV light to erase the EPROM memory, mostly as a whole. This particular principle is usable not only with EPROM chips but also with other similar types of memories based on floating-gate transistors, like EEPROM and FLASH. Despite the later named memory type was not initially designed for internal state changing by UV light, it was proven that it was possible—it can be seen as a form of unwanted fault injection. [96], [148]

Defending the chips against such attack scenario (not only against UV light) is based primarily on shielding the parts of the chip that could be attacked. With more layers chip layout, the element, i.e., the memory, is usually hidden behind a mesh on the top, and several metal layers in between. Therefore, light or particles do not reach their target. To re-enable this principle with the "protected" chips, there might be the need for partial deprocessing of the chip or FIB milling (milling a hole above the point of interest)—however, this leads to a more demanding invasive attack. To mitigate the protection impact, there is still the backside access available. Nevertheless, it is possible to shield the chip from both sides—this is mostly not the case.

For illustration, another similar scenario might be to copy data from a protected memory, where the anti-copy fuse is implemented as a memory cell. Such fuse has to be located and experimentally erased or set to "1" state. The difficulty of this task depends on the complexity of the chip and on where the fuse is located. It can be located separately from the memory part. It can be directly next to the memory, which is worse for recognition and also for the fuse value manipulation. The worst case is when the fuse is embedded in the memory itself. There exist various approaches to reset such fuses in order to unlock the read/write protection of the memory cells. The sophisticated ones use, e.g., laser for

precise content manipulation. On the other hand, there are also low-cost attempts based on the use of ordinary permanent markers and toothpicks. [96], [148]

As stated in Chapter 4.4, there are many possibilities (using light, particles, etc.) how to introduce a fault to the chip.

4.5 Non-Invasive Attacks

The last class of attacks is the most attacker-friendly one. It does not require very expensive equipment, working with dangerous chemicals, and that deep inter-disciplinary knowledge. On the other hand, the chip has to be seen as a black box, thus discovering a flaw is then almost as searching for a needle in a haystack. Therefore, the common approach is to perform invasive or semi-invasive attacks on the chip. This provides the necessary understanding of the chip's constitution and functionality. Afterward, the much cheaper non-invasive attack can be developed based on the information gained from the previous thorough examination. Moreover, the non-invasive way enables a widespread deployment of the attacks. The unquestionable advantage is that the victim sees no changes to the physical structure, as there is no need for physical tampering. Thus, the chip can be attacked directly in real use. Moreover, if the attack is prepared wisely, it can be performed completely out of sight of the victims.

In general, performing non-invasive attacks means, e.g., playing with supply voltage and clock signals—inserting power or clock glitches; examining input and output values; performing side-channel attacks. Very popular are power analyses—SPA (Simple Power Analysis), DPA (Differential Power Analysis), HO-DPA (High-Order Differential Power Analysis)—or timing attacks. Software-oriented attacks (investigation of communication protocols, implemented algorithms, etc.) also belong to the non-invasive group. [148], [102], [4], [67], [85]

Let us limit the non-invasive attack class with this short description because these attacks are the least relevant for this doctoral thesis. A detailed discussion of this group of attacks is out of the scope of this work.

Chapter 5

Chip Security Improvements

This chapter deals with the physical aspects of security of chip security and provides proposals for the hindering of subset of physical and side-channel attacks on microchips. We intentionally use the word "hindering" instead of "avoiding" or "disabling", because almost every countermeasure has the potential to be overcome at some point in the future. The target is to make the attack as disadvantageous for adversaries as possible.

Recently, we have seen several papers covering especially a split manufacturing process that allows for the construction of reliable and trustworthy devices, at least from the producers' perspective [175], [75], [40], [39], [58], [141]. A very favorite technique providing protection of the chips emerging in recent years is logic locking [49], [140], [180], [177], [82], [131]. On the other hand, there are attacks aiming at the logic locking and camouflaging techniques—SAT attacks (based on boolean satisfiability) [176], [103], [77], [23], [143], [180], [179], [178], SPS attacks (Signal Probability Skew) [180], [179], [178], CP attacks (Circuit Partitioning) [180], [178], SMT attacks (Satisfiability Modulo Theory) [9], etc.

In this chapter, we would like to propose possible techniques for hindering attempts on gaining knowledge from the physical examination of the chips. Methods employing recent technologies like 3D integration, MEMS, integrated energy source, etc. will be introduced.

We will not consider the price aspect in the following chapter, because what is expensive for one use case may be acceptable for another one. At the end of the day, price always influences the final design and many decisions made along the way to the market. As we do not want to present a concrete example where it would be possible to assess adequacy of a particular countermeasure, let us propose and describe various possibilities for increasing the security of microchips, regardless of its price.

5.1 Reverse Engineering Countermeasures

For a long time, reverse engineering attacks had been neglected through deceptive feelings of the inherent security of microchips. With up-to-date knowledge, we know that adversaries can be very well equipped, as some of the attacks might be, for example, of national interest and thus have strong financial backing and a desire for results [126]. Moreover, there are not only cutting-edge chips available on the market, there are many chips produced with older technological nodes, due to financial reasons or even overhauled outdated specimens secretly used in places where nobody expects them [61], [62], [60], [121], [125], [129], [167], [169], [182], [183]. This allows for many amateur adversaries (e.g., up to Level 2, as described

in [112], or up to Level MODL, according to [1]) to perform a cheap, partial reverse-engineering process with success.

Let us assume a scenario where an adversary has several pieces of a chip that is supposed to be partially or fully reverse engineered. During our research, we have proceeded in a similar scenario with high-tech international partner companies.[1] Along the way we have learned what countermeasures these partners recognized as tough, or sometimes even nearly impossible to break. Our proposals presented in the following chapters are based on this apprenticeship.

The RE scenario is about detaching the chip from a system, the decapsulation of the outer packaging, and the further examination of the chip out of operation, which means delayering, acquiring images, and analyzing the acquired image data [160], [98], [100].

5.2 Technological Node

The more advanced the technological node used for the fabrication of a particular chip is, the more advanced equipment is needed for successful deprocessing. Hand in hand with advanced equipment, a deeper knowledge is also required for the personnel operating the appliances. The later chips typically express progress in technology by providing more powerful features that are achieved by utilizing a higher number of smaller transistors, more metal layers, different materials used for conductive lines, and also insulation.

Our recommendation is to employ the latest possible matured technology for the production of security-aware ICs. With this measure, we can trim down the range of potential adversaries to those with access to the appropriate technologies. Needless to say, the price for the necessary equipment for the deprocessing of advanced nodes is at least in the hundreds of thousands of US dollars, but it can easily get into the millions.

This first recommendation is a very basic and rationally expectable one. However, in the following chapters, we will show that it is not always ideal to use the latest technologies everywhere. As we have learned during our visit to the TESCAN Laboratories, the idea that everything can be delayered just with the latest FIB machine is incorrect. Each technological node is specific and thus requires a specific approach, i.e., for older nodes (above 100 nm) it is far from convenient to use the very recent FIB tools, which are focused on the smallest structures around 10 nm.

In other research articles [115], [147], [120], [137], [148], we have found statements that universities and similar institutions usually have some of the necessary equipment available. Therefore, it should be possible to rent these expensive tools at a relatively low hourly basis. First of all, our university is by far not equipped with such machines. We tried this recommended approach by renting a very advanced FIB machine at one of our partner's laboratory. Nevertheless, without a thorough training on how to use all of the required features of that particular machine, with its hundreds of settings, the adversary or researcher is hardly able to achieve high-quality results. For achieving an output of a decent quality, it is necessary to rent the machine together with the erudite personnel. Furthermore, the latest chips are spatially large when compared to the ones produced ten years ago, and so removing all of the layers and acquiring images of the whole chip structure is not a matter of days, but rather weeks, or even months, unless we aim for a very specific area of interest only.

[1] ON SEMICONDUCTOR, TESCAN, PRESTO ENGINEERING

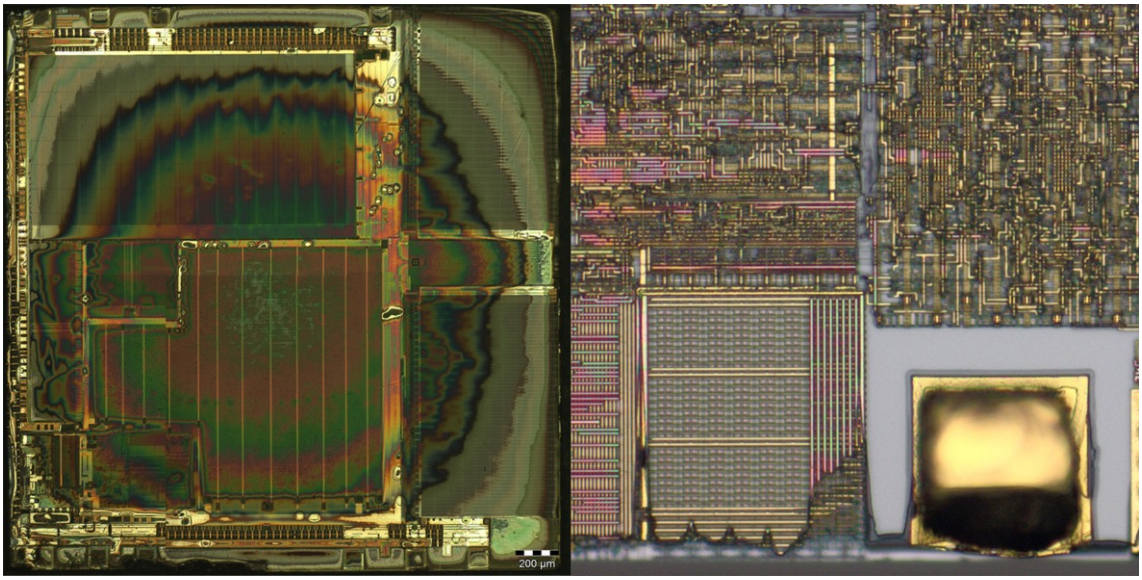


Figure 5.1: Left: Planarization problem after layer polishing. Right: An underetching problem after several layers are removed (right). (Source: author's work.)

5.3 Complex Integration and Camouflaging

2.5D and 3D integration is a substantial contribution to a possible security increase in IC fabrication. These chip composition techniques are emerging especially in relation to split fabrication processes that should assure the genuineness of IC production in offshore foundries [175], [75], [40], [39], [58], [141]. Solving supply chain issues is not the aim of this work, however, so we will refer readers to the above-mentioned papers for more information about that subject.

Our intention with the employment of 3D integration is to hinder delayering and the consequent analysis of the inspected specimen. Delayering is already complicated with current state-of-the-art 2D integration—e.g., avoiding unintended cross planar grinding or underetching is tough enough with the recent nodes; see Fig. 5.1.

Removing all layers of a 2D design down to the silicon is currently still feasible. We also succeeded with the use of chemical etching on biometric passport chips revealing the silicon layer. Nevertheless, transistors express only a part of the IC blueprint, the same importance lies in the interconnections that add semantics to the whole circuit. Thus, obtaining images of transistors is, in a vast majority of cases, not sufficient. Therefore, adversaries have to concentrate on extracting all of the needed information—transistors and interconnections. And this can be aggravated through the use of 3D integration. Let us name several possible measures on how to make reverse engineering more challenging.

5.3.1 Heterogeneous Integration

Heterogeneous integration of dies made with different technologies will certainly make planarization issues much deeper. It would be ideal to utilize a combination of materials at the same planar levels that have very different grinding resistivity. Grinding would require equipment that allows for perfect control over the grinding process (in order to keep grinding absolutely uniform across the whole heterogeneous plane). Wet or plasmatic etching

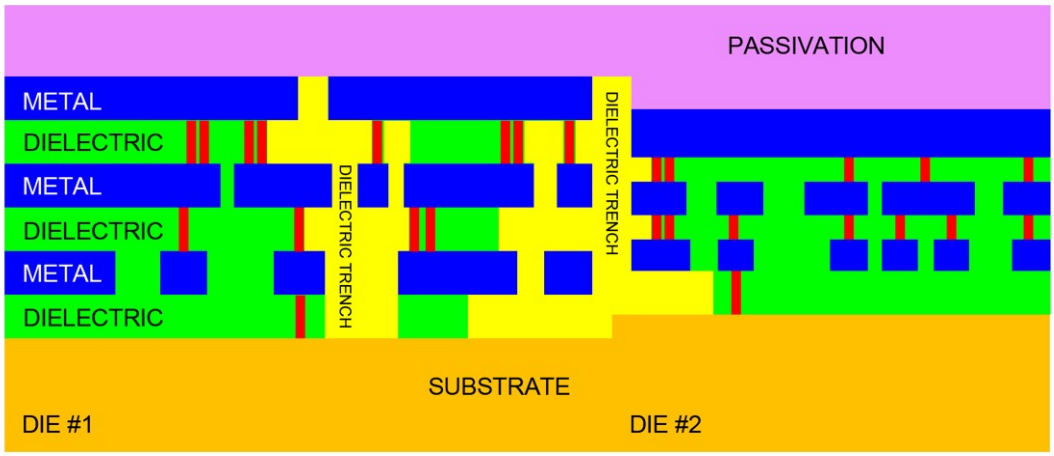


Figure 5.2: Illustration of heterogeneous integration of a chip. (Source: author's work.)

will be even more difficult, especially when the layers will be wisely combined in order to ensure underetching or even direct damage to the adjacent layers. In Fig. 5.2, see the dielectric layer that is a combination of two different materials with a dissimilar resistance to chemicals—the green parts endure for much longer before dissolving, which enables yellow dielectric trenches to initiate underetching. The passivation layer can be of a different thickness across the die to make the decomposition harder from the very beginning.

5.3.2 Unreadable Non-Volatile Memory Types

Although we truly hope that this "must" is fully obvious to anybody interested in security, we have not allowed ourselves to omit mentioning it at least in a few lines.

Vastly used, cheap masked ROM memories can be literally read out after the proper delayering of a device [126], [160], [32], [87]. To deflect an information breach, we recommend the complete abandonment of using masked ROM memory types and types with similar features (readability of stored values, e.g., [126], [31], [32]; impossible to erase/rewrite content). This step will help us to keep the stored information optically unreadable and will also give us the opportunity to employ a defense scenario presented in Chapter 5.4.

The employment of memory encryption might be seemingly enough for protecting the plain content stored in memory cells. Unfortunately, frauds can find a way to decipher the stored information [146], [59]—either finding the right key or reading out the data after it is decrypted by the device itself. Generally, one more step toward security would be to not disclose the memory content at all, regardless of the encryption used.

5.3.3 Cell Camouflaging

Cell camouflaging or circuit obfuscation are known techniques described in several research papers [12], [26], [55], [128], [142], [168], [166], [126], [141], [21]. It is known that this technique is expensive because of the aerial demands, so it is impossible to camouflage the whole IC. Furthermore, the security impact can be of a much lower extent than expected during design time due to existing attack scenarios [75], [142], [128], [55], [166]. Furthermore, it is possible to observe obfuscated cells through a series of cross-section slices with

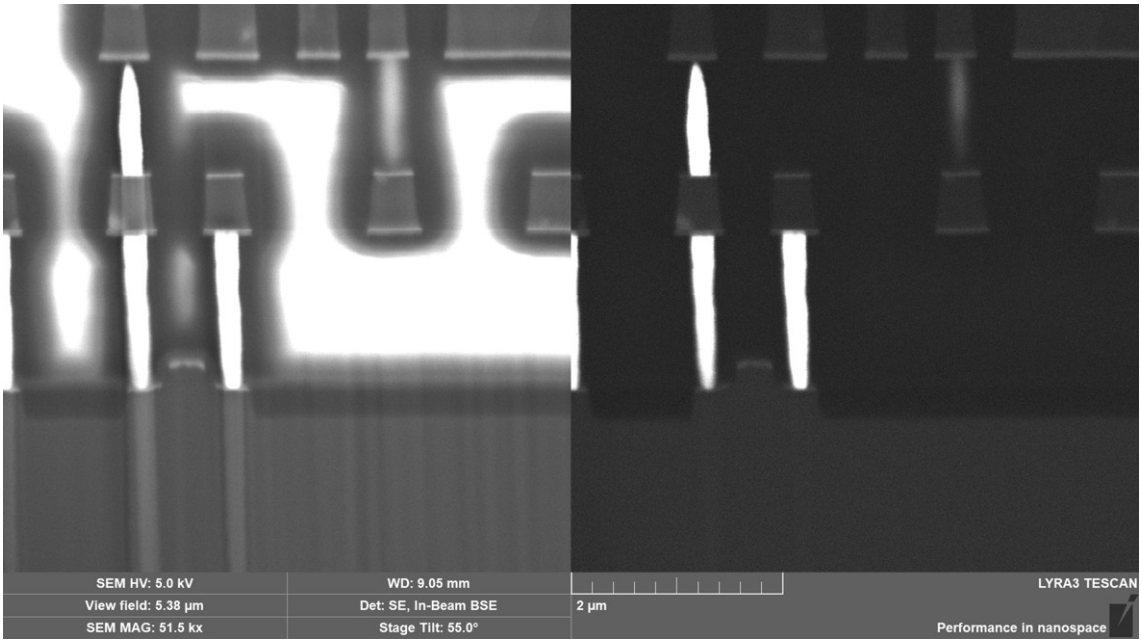


Figure 5.3: A FIB cross-section detail created with the use of TESCAN LYRA3. (Source: author's work.)

a properly set milling step (a detailed FIB cross-section is displayed in Fig. 5.3, step by step cross-section demonstration with 50 nm milling step is displayed in Appendix A). With this approach, it can be determined which contacts are really connected and which are just fake. Such advanced cross-sectioning is achievable with FIB milling or with X-rays [126], [10], [33], [118].

Let us introduce the possibility to disable this cross-sectional analysis of camouflaged cells with the employment of inductive or capacitive contact-less connections, where some of the contacts in the camouflaged cell can be fake without showing any visual difference. This potential enhancement also has its drawbacks, e.g., spatial and power requirements, heat dissipation, and potential side-channel attack support. Camouflaged cells are spacious even with the physical contacts, so there is not much of a difference. With wise design, we might get to the same spatial needs and potentially a similar camouflage effect. The fake contacts will then be visually indistinguishable from the real ones. It is clear that the use of this type of obfuscation in a single die has to be very limited due to its drawbacks [37], [48], [107], [79]. Nevertheless, it would mean one more measure against physical reverse engineering and visual observations.

5.3.4 3D Integration with Dummy Dies

There are many unused or recycled old dies available on the market (which are vastly used by fraudster foundries in fallaciously new integrations [61], [62], [60], [121], [125], [129], [167], [169], [182], [183]). These can be wisely used for increasing the complexity of 3D integrations. Although this artificial complexity bloat will not prevent adversaries from performing decomposition and analysis, the intricacy of the integration can be risen. The time consumed for the determination of the dummy part may help discourage adversaries.

We propose using dies with diverse technological nodes for 3D integration. Each node requires a distinct approach for delayering and analysis. This approach will make reverse engineering more unfriendly. The interconnection between the dummy part of the integration with the truly used segments of the chip will be important. When connected sloppily, an attacker might suspect the fake part quickly. Correct employment of this measure requires the thoughtful placement and linking within the 3D IC.

The disadvantages of this solution are mainly technological. Because thermal management is one of the most important aspects to be dealt with in 3D integration, adding more unnecessary dies into the integration makes the situation worse. When we consider connecting the dummy part electrically to confuse the attackers as much as possible, more power will be consumed, thus more heat radiated into the 3D IC. Therefore, the implementation of this measure has to be carefully judged at the design stage. On the other hand, increasing security is in some use cases so valuable that it might be worth spending the extra effort on camouflaging the design with dummy parts.

5.4 Active Tamper Detection

When adversaries perform reverse engineering, the chip is destined for physical destruction, unless proceeding with X-ray only.

The chip is detached from its package and the power source. The latter does not have to be necessarily the truth in the very near future, due to discoveries and the successful development of micro batteries suitable for direct integration into ICs [17], [83], [114]. Due to the growing complexity of chips, we do not expect batteries to be capable of powering the whole chip for an exceptionally long time. Nevertheless, if we focus strictly on keeping the protective functionality alive only, this might result in a decent duration for active tamper detection endurance, even without an external power source. Moreover, recent endeavors in the field of energy generation can lead us to mechanisms that are capable of refilling the integrated battery and hence allow for the exceptional endurance of active tamper detection. Let us name VEH (Vibration Energy Harvest) based on MEMS (Micro-Electro-Mechanical Systems) [3], [29], [38], thermoelectric generation based on parasitic load of the device [52], [38] and photovoltaic solar power generation [38], [18], [132] which can be sensing the package decapsulation at the same time.

Integrated non-volatile memory is very often targeted because of its content—sensitive data, passwords, configuration, etc. Therefore, our aim is to protect the chip from tampering, intrusion, or analysis of its physical structure revealing the internal arrangement, the memory cells or in some cases even the content or memory cells. In this chapter, we propose the employment of active tamper detection mechanism in order to detect undesirable manipulation with a chip. We also present measures protecting memory content from being disclosed in two different scenarios.

After tamper detection, our target will be to reliably remove memory contents (Chapter 5.4.1) and/or damage the circuit itself (Chapter 5.4.2). With successful tamper detection and consequent memory erasure, the examined chip might be of a significantly decreased value.

5.4.1 Active Tamper Detection with Active Memory Protection

An active tamper detection shield should consist of several layers aimed at the possible ways of intrusion. The partial or complete decapsulation is one of the first steps when targeting

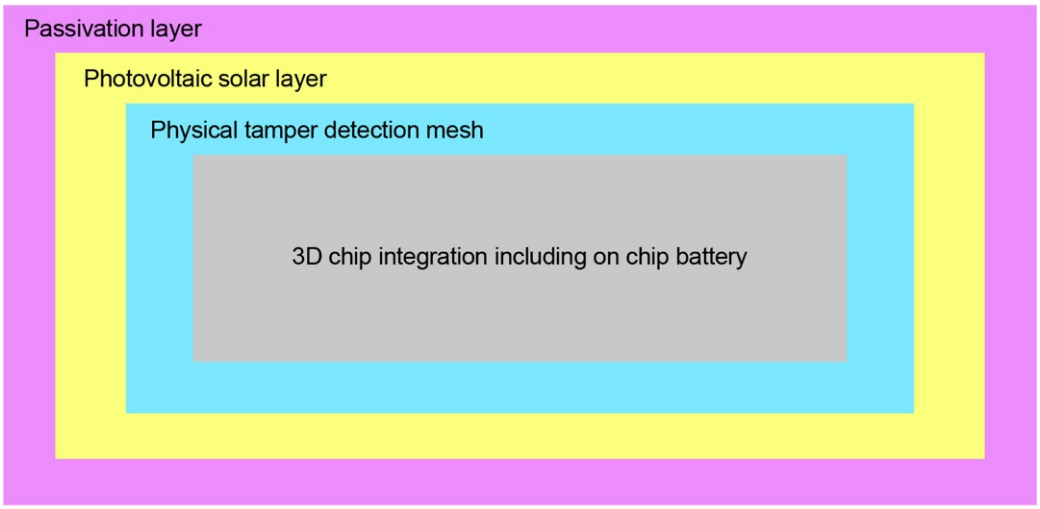


Figure 5.4: Simplified composition of a chip with active tamper detection. (Source: author's work.)

chips with invasive investigation. After opening the package, there should be natural light entering and interacting with the chip's surface. Therefore, the very first detector should be a light-sensing layer on the top of the chip. Ideally, it should be covering the chip's entire surface to ensure that even partial openings trigger the alarm.

The second anti-tamper layer should be a fine-pitch sensing mesh against attempts of penetration into the chip. Even small-sized FIB editing has to be detectable by this layer in order to not allow for any modifications that could possibly lead to the restriction of active shield functionality, memory bus exposure, etc. This layer should be actively powered by the integrated battery to regularly check the integrity of the mesh. Due to the fact that reverse engineering is not a fast process, the check can be set to be run after a certain period. This interval has to be designed with respect to the power demands of the whole active shield circuit as well as the capacity of the integrated battery. It can be expected that the parameters of the batteries will significantly improve over the next few years. Then this active shield use case will be supported even more.

In Fig. 5.4, we provide a simplified view of a chip's structure with respect to the proposed active protection. As there are many attacks led from the "backside" of a chip, we recommend using 3D integration with a back-to-back connection to have a 3D chip with only the frontal part facing all the edges of the packaging. Passivation, photovoltaic sensing, and physical tamper detection layers are used all throughout the chip's structure. A battery is expected to be integrated inside the 3D integration.

As soon as the outer package is (partially) removed, the photovoltaic solar layer produces energy. This should be the signal to immediately remove memory content. Memory content removal should be a battery-powered action.

In cases when the attackers would be able to somehow disable the photovoltaic sensing layer and thus would be able to avoid immediate memory erasure, their next step will be layer-by-layer removal or FIB editing. Once the physical layer tamper detection mesh is touched, the same memory-erase signal will be triggered.

Memory modules should have low energy demand in order to enable this protection scenario with battery-powered memory erasure. When the battery reaches its critical low

level of charge (the minimum charge needed for memory erasure), it should automatically erase the memory content in order to devalue the chip. This low-charge status can occur when the battery is not recharged or in the event of malfunction.

We can more or less rely on battery re-charging mechanisms (TEG, VEH, Photovoltaic, etc.), and thus prolong active shield durability for various scenarios.

Considering our work where we deal with personal e-documents and chips inside those (e-Passports, ID cards, ILR, ...), let us provide a whole scenario for the battery-powered active tamper detection use case. Our theoretical assumption is that we are able to power the active tamper shield with an IC-integrated battery for at least N years. If the document is on the move with its holder, it is automatically re-charging the battery because of the integrated VEH system. If the document is used, it is re-charged as well, with power obtained from the document reader and because of the heat produced by the chip (thermoelectric generation). The worst case is when the document is stored in a drawer and never used during that N year period. In this case, the battery slowly discharges. When it reaches its low charge limit, the chip itself should trigger the command for memory erasure, making the document invalid. When we go even further, there might be a battery status indicator (low, mid, high), based on technology like e-ink (low power consumption, power is used only when switching states), showing to the user whether the passport needs to be recharged in order to remain valid.

Furthermore, due to the very rapid development of technologies, it can be expected that such chips for holding e-documents will be implanted into human bodies soon [172], [163], [122]. It can be as easy as implanting an ordinary RFID chip under human skin, which has already happened. Such a chip will have direct access to a power source in the form of the heat produced by the human body. Under these circumstances, we will not have to think about the endurance of the internal battery that much, because of the constant power that is available. As soon as there is no thermal power, the chip will assume the owner is deceased or there was an unallowed extraction from the body. In these cases, it shall start its memory erasure procedure based on the internal battery power. Aside from that, the implanted chips will have direct access to the biometric characteristics of the holder and will be able to detect counterfeit attempts.

5.4.2 FPGA Employment with Active Bitstream Protection

Protecting a chip against reverse engineering by implementing its key parts inside a fully integrated FPGA circuit is not a novel idea in principle. The concept is based on the fundamental presumption that FPGA is composed of visually similar cells that change their behavior according to the configuration loaded upon power up. However, there are known attacks against such implementations [126], [57], [109] focusing on the reconstruction of the FPGA configuration bitstream, thus essentially gaining a netlist of the circuit.

To avoid these attacks, we have to protect the main memory that holds the configuration information and buses from micro probing, FIB editing, etc. Our proposal is to physically protect the chip in the same way as described in the previous chapter with the active tamper detection shield. Whenever there is an alarm triggered by the active shield, all configurations of the key functionality implemented in FPGA has to be reliably deleted. The attackers then gain a worthless chip with general purpose FPGA and no notion as to its configuration.

5.5 Active Defense Against Microprobing

Targets in a microprobing attack are mostly internal buses or signals that are not freely accessible via standard contact pads. When connecting to these internal signals, adversaries can reveal signals and data that are meant to be shielded from the outside world. Before employing microprobing, the chip has to be at least partially decapsulated and then appropriately tampered. First, an analysis should take place in order to figure out the areas of interest to figure out where exactly to place the needles. After a successful analysis (this means destroying a few samples), the aimed contacts have to be exposed for the micro-needles.

Our aim is to protect the chips from tampering attempts, intrusion, or the analysis of its physical structure that reveals its internal arrangement. In Chapter 5.4, we proposed the employment of active tamper detection in order to detect the undesirable manipulation with a chip as well as measures protecting the data processed inside the chip from being disclosed.

The countermeasures proposed in the previous chapters have the following stages—the first stage is the active tamper detection mechanism and the consequent stage is then the reliable removal of the memory contents (Chapter 5.4.1) and/or damaging the circuit itself (Chapter 5.4.2). With successful tamper detection and consequent memory erasure, the examined chip might have a significantly decreased value for adversaries.

5.6 Disabling Backside Observations

Backside imaging can be considered an easy method of almost directly accessing the transistor layer with further scanning possibilities realized by photon-emission microscopy, laser voltage probing, laser voltage imaging, IR imaging, thermal emission imaging, etc. [15], [22], [150], [157], [165]

The typical first step towards such a backside observation is to decapsulate the back side of a chip, either by using wet etching or via the polishing technique. It is usually not necessary to remove the whole packaging, thus polishing is very suitable. Furthermore, when examining chips connected as a flip chip (the backside of the chip is facing the package), it is very convenient to consider backside access.

Subsequently, there might be some obstacles in the form of various pads placed below the silicon part of the chip. These pads, whatever they are made of, have to be removed. Consequently, the silicon substrate has to be thinned down according to the chosen scanning technique (100 μm–50 nm) [15], [22], [150], [157], [165]. After preparation as stated above, backside observations can take place.

Because the recent chips are becoming very complex in terms of layer count and the advanced level of materials used for the metal and dielectric layers, the results of many observation techniques may become meaningless in such a tangle of metallic and nonmetallic structures. Therefore, the idea to inspect transistors directly from the backside while preserving the whole chip structure above it was introduced [15], [22], [157]. This allows active observations with specimens in full operation, while only the backside is exposed.

Our proposal for disabling the techniques using backside access is to employ 3D integration, as mentioned in Chapter 5.3, so that there is no real backside of the chip (see Figure 5.4). Ideally, there should always be the frontal part of the chip facing the outer world, from all angles. In this chip layout, there is no backside exposed or easily accessible. One can object that one of the chips in the 3D layout can be removed and thus the backside

of the other chip might be exposed. Nevertheless, the removal of one of the chips from the 3D layout makes the active backside observations of a specimen under operation practically impossible, because the chip setup is designed to operate as whole, not separately.

A further proposal in regard to designing security-oriented 3D chips is that these should work only when correctly interconnected, even when using dummy dies. In case a chip is taken out of such a 3D layout, the chip setup should be able to detect it due to missing signals, different delays of signals, etc. Thus, the removal of a dummy die should not allow the adversary to perform an attack from the backside.

In back-to-back 3D design, *through silicon vias* (TSV) are very likely to be in place for interconnecting the particular chips. These vias can ensure the integrity of the whole setup—the physical properties of particular TSVs can be checked by the chip, and thus the chip can recognize tamper attempts. In such cases, the observed chip can either completely refuse operation or it can intentionally operate in a different mode to confuse adversaries.

Backside protection mechanisms can be combined together with mechanisms proposed in Chapter 5 to provide as complex protection as possible. Let us name the active tamper detection mechanisms and the use of FPGA for critical functionality. These mechanisms can serve as a fuse for incidents when backside protection fails.

5.7 Side-Channel Attack Countermeasures

This chapter is devoted to providing further proposals on hindering a subset of possible hardware-oriented attacks. Here we will focus on improving the physical security of microchips against a subset of so-called semi-invasive attacks that are widely used against cryptographic hardware and smartcards.

5.7.1 Side-Channel Attacks

Side-channel attacks, belonging to non-invasive attacks [148], [54] are among the favorite attack types due to the relative simplicity of their execution compared to invasive attacks. This does not imply that the performance of such attacks is in all cases effortless. In these scenarios it is not necessary to decapsulate the observed specimen at all. The specimen can be powered on, encapsulated in its original package, and measurements of examined quantities, e.g., heat radiation, power consumption, radiation in general, acoustic analysis, can be carried out. However, it might still be convenient for some of the observation schemes to fully or partially decapsulate the chip, i.e., for more precise measurements of the formerly mentioned quantities. Then we can question whether we are still in the group of non-invasive attacks, however, this is not to question.

5.7.2 Active Defense Against the X-Ray Observation Technique

Lately, X-rays have become a serious technology used in the observation of advanced chips [118], [33], [34]. As chips are turning into unimaginably complex devices, the classic method of invasive reverse engineering is becoming harder. With the recent progress of X-ray technology, it seems that classical reverse engineering is no longer practical compared to X-ray imaging. We have to consider all chip structures disclosed at any time, even without being physically penetrated. Protecting a chip against reverse engineering by implementing its key parts inside a fully integrated FPGA circuit was already presented in Chapter 5.4.2.

In fact, X-ray exploration is basically a kind of non-invasive reverse engineering. Thus, similar measures, as used for anti-RE, can be taken in order to increase the chips security.

In order to avoid these attacks, we have to protect the main memory that holds the FPGA configuration data and buses against information leakage. We propose using active tamper detection with active memory protection that was presented in Chapter 5.4.

Protection against X-rays or ionizing radiation has to also employ a protection mechanism against these non-invasive observation techniques. Either radiation detectors have to be placed inside the chip's structure [123], [93]; according to recent research in physical chemistry, it is also possible to turn X-ray radiation directly into electricity through the use of nanomaterial. This might be used as a sensing technique for triggering proposed memory erasure procedures in a similar way as in 5.4.

Regardless of the practical implementation of X-ray sensing, the described setup should ensure that when radiation exceeds a given threshold, an alarm is triggered. Whenever there is the alarm triggered, all configurations of the key functionality implemented in FPGA has to be reliably deleted. The attackers then gain a worthless chip with general purpose FPGA.

5.7.3 Passive Defense Against X-Rays

For hindering radiation-based observation techniques, it is possible to use the methods presented in the section devoted to reverse engineering, such as:

- Cell camouflaging.
- Inductive or capacitive contact-less connections.
- Key functionality implemented in FPGA.
- Visually unreadable memory cells.
- Increasing complexity with 3D integration.
- Increasing complexity with dummy dies.

Ideally, it should be possible to allow the attackers to display the physical representation of the chip. As stated before, protection against X-ray imaging is not dissimilar to the protection against physical reverse engineering.

Among others, it is possible to use materials that are used for radiation hardening in general, especially in the space industry; for example, borophosphosilicate glass [181]. The chip package can be constructed from materials that will make X-ray scanning difficult (however, this research field is not covered in this thesis). Therefore, it would be needed to decapsulate the chip first. At this point, active tamper detection presented sooner in this chapter can be used for package intrusion detection.

5.7.4 Power, Thermal, and Timing camouflaging

The chips are very often analyzed by reading power consumption, thermal emissions, or time spent within the performance of some operations [4], [67], [85]. The better control over the input parameters, the better starting point for the analysis.

As a prerequisite to the ability to influence values obtained from such analyses, the chip has to be able to generate truly random events. The second precondition would

be to place several inductive and/or capacitive contact-less connections (as mentioned in Chapter 5.3.3) into the chip layout. Not only do we contribute to the physical camouflaging, with this implementation we can also influence side-channel outputs. Some parts of the important functionality can be physically realized in multiple traces—direct connection (shortest path), with longer conductive lines (i.e., through the use of an extra buffer), and and with contact-less connections. Unfortunately, the design phase will get to a new level of complexity due to variable delays, consumption, thermal properties, spatial requirements, production price, etc. for each function implemented with this countermeasure. However, if the designers manage usability of the worst trace of each function implemented as proposed, then the chip will respond to the same inputs in a pseudo-random way. We used "pseudo-random" here because the implementation in hardware will always give us only a limited number of interconnections, thus only pseudo-randomness. Choosing the particular trace for the actual computation of the multiple-trace function should be based on the random event generator, so that the choice of the trace would not be predictable for the attacker. With using few contact-less connections in the IC, we may introduce a portion of noise into the measured side-channel signals. With multiple-trace functions, we introduced variable timing for the same computational tasks. These countermeasures bring more ambiguity to the signal interpretations and measurements.

We recommend going even further with the implementation of multiple-traces of chosen operations. The final operation should consist of several segments, whereas each segment should have several implementations with various traces. In every segment, the trace should be chosen independently on the fly based on the random event generator. This would give us more combinations for one functionality implementation with hard-to-distinguish and hard-to-map side-channel signals.

One more step ahead for application of the proposed principle would be to employ an FPGA and reconfigure, or partially reconfigure, the FPGA in truly random times. Thus, the same operation will use random, different FPGA cells and lengths of interconnections. This can lead to quite a complex solution that is primarily limited by the size of the FPGA.

Chapter 6

Biometric Passports

The chips we are dealing with are used in the implementation of the Czech biometric passports—this revision of chip was used for issuing new passports until end of 2014; considering the standard 10-years validity range, the chips are still in active use. Information about this implementation is presented in this chapter in order to gain a broader overview of the analyzed piece of hardware—what to expect inside and what the use cases are. This knowledge gives analysts at least a rough view on what to expect under the hood.

6.1 RFID Technology

Radio Frequency Identification (RFID) is currently a widely used technology, that is massively used simply for all kinds of identification actions—chain supplement, personal identification cards, access cards, etc. The RFID chips for simple purposes are equipped with only a few bytes of ROM memory, on the other hand, modern trend is to integrate various functions with sufficient amount of writable memory, like in the case of the e-Passports (all the electronic passports are labeled with an international logo—see the red circle in Fig. 6.1). In this case, we can talk about cryptographic functions and r/w memory modules accompanied by memory modules that are readable only for the tag itself. No information from these memory cells can be retrieved out of the device.

RFID technology is based on two main devices—RFID tags (also known as the RFID transponder) and the RFID reader that usually has a writing ability, so the term "RFID reader" can be misleading. RFID tags can be either active or passive. The tag is commonly connected to an antenna. Active tags have onboard power supply—usually a battery—and active transmitter. Analogically to the active tags, the passive tags have no integrated power source and no active transmitter. The biggest difference between active and passive tags is the price and the operating range—passive tags' range is given in centimeters or tens of centimeters (e-Passports incorporate passive RFID tags). In contrast to that, the active tags can communicate for up to kilometers. [144], [92], [14]

All wireless technologies bring many advantages, but also disadvantages—in the scope of travel documents, the biggest issue is security. It is clear that all wireless transfers can be eavesdropped or exposed to other known attacks. That is why all communication transferring sensitive information has to be securely encoded.

For the e-Passport chips the standard-compliant with ISO 14443 with modulation A or B was chosen—frequency for transmissions is 13.56 MHz with a short-range (max. 15 cm). [130] This manner of use of biometric passports can prevent at least some of the

Figure 6.1: Left: A Czech biometric passport specimen with a circled international logo of electronic passports. Top right: RFID chip without and with an antenna. Bottom right: The data page with labeled MRZ. (Source: [64])

attacks like "man in the middle" attacks—due to the short distance between the tag and the reader. [92], [14]

6.2 Passport Chip Memory

The memory is logically divided into two main regions—one is accessible from outside of the chip (via wireless communication), the second one provides security by hiding its content—the hidden content is available only for internal functions of the chip.

The part of memory available for reading provides sixteen separated data groups (labelled as DG1, DG2, ..., DG16—see Fig. 6.2). Each group holds different data. Dissimilar types of protection are used over the groups of the stored data. The data groups DG1, DG2, DG3, and DG5, are important within the scope of the biometric passports because these groups are used for storing information related to the identity check. [68]

Data Group 1
: DG1 stores exactly the same information as those presented at the data page of the passport (see the bottom right part of Fig. 6.1)—basic personal information like name, date of birth, place of birth, sex, date of expiration, etc.

Data Group 2
: This data group is dedicated to a digital form of a facial photograph used for facial recognition. Size limit is set to 15 kilobytes.

Data Group 3
: The most recent security element of passports—fingerprint(s)—is stored in the DG5. Size limit is set to 15 kilobytes per fingerprint.

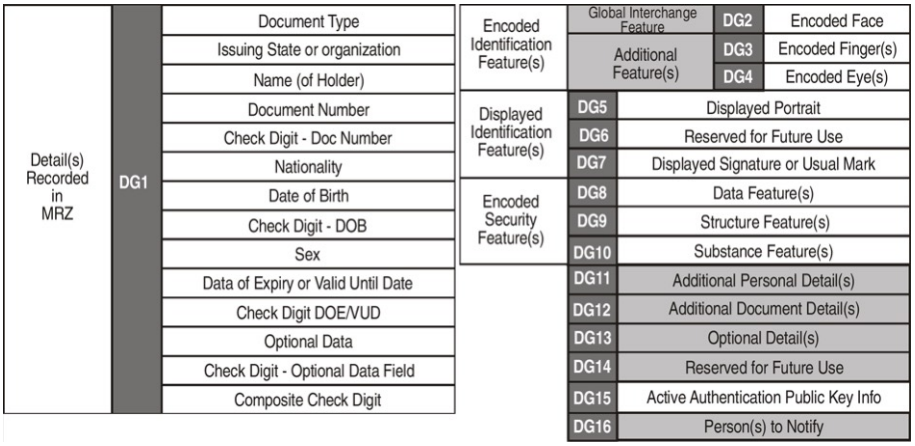


Figure 6.2: Memory data groups of e-Passport chips. Please notice the description of DG1–DG5. (Source: [68])

Data Group 4

Data group 4 should contain encoded iris data, but this feature has not been used yet in the Czech implementation.

Data Group 5

The last important data group (with respect to the biometrics) stores a photo of the owner that is depicted on the data page. This data group is not used in the Czech implementation.

6.3 Introduction to Biometrics

Techniques based on biometric features are being widely deployed, especially in the spheres, where a higher level of security or a precise identification is desired. However, all the technologies are becoming affordable for more ordinary purposes, as well. We can expect massive use of biometric-based products in the following decades.

A proper biometric feature should be unique for each person, and it should be invariable in time (usually from a specific age); given in the simplest possible way—it is an unambiguous identifier of a person. Moreover, some of the biometric features are well proven and have been practically used for a long period of time, i.e., fingerprints in criminalistics. On the other hand, many of the biometric features have been explored relatively recently. As it is not possible to give an exhaustive overview of biometrics, let us focus on the features that are important for contemporary passport implementation—2D facial photo and fingerprints. The use of the iris can be expected in the near future. [47]

6.3.1 Facial Photograph

Facial photograph of an applicant is employed as a basic security element. This type of security is well-known also from older types of documents. In classic paper documents, the facial photo primarily serves for visual identification by officers. Despite the officers' training and their ability to recognize a person even if there is some change in an applicant's appearance (mustache, haircut, glasses, etc.), the case of similar individuals (twins,

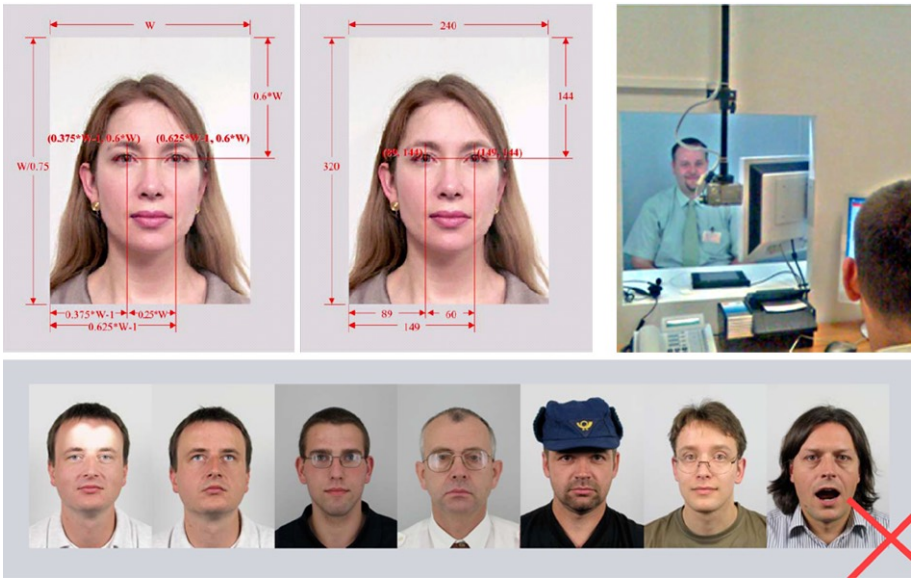


Figure 6.3: Top left: An example of an ideal facial photo with measures. Top right: The Czech endpoint station from the officers' view. Bottom: Unacceptable facial photos. (Sources: [63], [104])

siblings or even doubles) could lead to identity mismatch. If the facial photo is treated from a biometric point of view, not just as a picture of a person, the face contains information that is invariant in time and can be measured, e.g., the distance between eyes, the position of chin, the position of nose. These factors can affect the recognition process by providing additional information to the officer. Nonetheless, twins will still look similar. That is why a completely different security component is needed (see Chapter 6.3.2). [47]

Picture Data Storage

The picture data holding the facial photo is taken according to specifications in ISO19794-5 that defines conditions for the acquirement of this type of data: format, scene, picture properties, etc. The picture data is stored on the chip twice (DG2 and DG5—see Fig. 6.2), both in JPEG/JPEG 2000 format.

The first occurrence is encoded and stored in DG2 in full color, resolution of 240×320 with max. size of 15 kilobytes. Although smaller in resolution than the second representation, this image is used for a biometric identity check. The second picture is designated for laser engraving with the following properties—gray-scale, 60px distance between eyes, resolution of 620×796, stored in DG5. [104], [68], [101]

6.3.2 Fingerprints

With respect to the facts introduced in the second paragraph of Chapter 6.3, the need for new reliable means of identity verification has been solved by introducing fingerprints. It has been proven that even fingerprints of monozygotic twins are significantly different. That means the two identities of twins can be undoubtedly distinguished by matching the

Figure 6.4: Difference between a rubber stamp fingerprint (left) and a fingerprint from a real finger (right). (Source: Martin Drahanský, FIT BUT course Biometric Systems)

corresponding fingerprint with its stored digital representation[1]. Even so, there still exist possibilities for counterfeiting fingerprints. Nevertheless, the fraudsters have to face the problems with tricking the fingerprint scanners, because the scanners are being more often equipped with sophisticated liveness detection—especially when a security risk is expected. Sometimes it is simply impossible to cheat the fingerprint checking, i.e., because of the presence of an officer. Adopting this measure naturally does not result in an absolutely perfect protection against unwanted actions[2]. Nonetheless, the security level has been rapidly increased with the incorporation of the fingerprint check. [47]

A potential attacker can use finger fakes to circumvent the fingerprint reader. Thus, securing automated and unsupervised fingerprint recognition systems used for access control is one of the most critical and most challenging tasks in real-world scenarios. Basic threats for a fingerprint recognition system are repudiation, coercion, contamination, and circumvention [42]. A variety of methods can be used to get unauthorized access to a system based on automated fingerprint recognition. If we neglect attacks on the algorithm, data transport, and hardware (all these attacks demand good IT knowledge), one of the simplest possibilities is to produce an artificial fingerprint using soft silicone, gummy, and plastic material or a similar substance [42], [161], [154]. One example of the use of an artificial finger is shown in Fig. 6.4, where the fingerprint from a rubber stamp is displayed in comparison with the fingerprint from a real finger. For a really big number of sensors, there is no difference between them, i.e., the artificial fingerprint is processed and recognized as a concrete enrolled user from the database. To discourage potential attackers from presenting a fake finger or even worse, to hurt the authentic person to obtain the real fingerprint, the system must be augmented by liveness detection component [42], [46], [80]. To prevent false acceptance, we have to recognize if the finger on the plate of the fingerprint sensor (also referred to as the fingerprint scanner) is alive or not. There exist the following liveness detection methods [42]: perspiration, spectroscopic characteristics, ultrasonic technology, physical characteristics—temperature, hot and cold stimulus, pressure stimulus, electrical properties, bio-impedance, pulse, blood oxygenation, and some other experimental and not very reliable methods.

[1] Of course, not only with the digital representation of the fingerprint, but also with a paper record of that fingerprint.

[2] Absolute security does not exist.

The second often neglected problem are skin diseases and their influence on fingerprint recognition [45], [44], [41], [43]. These skin diseases attacking fingers or generally hands could be divided into three main groups [45], [44], [41]: histopathological changes, skin discoloration, combination of histopathological changes and skin discoloration. Histopathological changes mean that the structure of papillary lines is changed and the biometric system is not able to detect the separate papillary lines and valleys among them. Most of the sensors are based on physical principles, which do not allow acquiring a fingerprint with a histopathological skin disease. The second group contains skin discoloration, i.e., only the color of the skin is changed, but the structure of papillary lines is kept unchanged. Most of the sensors for fingerprint acquirement are not prone to this type of skin diseases. The last category combines both previous types. This category is very difficult for almost all of fingerprint sensors because the combination of change of the structure of papillary lines and change of their color is often resulting in structure and color, which is not recognizable as a fingerprint for further processing. Closer description of these skin disease categories, including concrete disease examples, can be found in [45], [44] and [41]. Furthermore, basic information about how to evaluate the quality of a fingerprint is also presented in [45], [44] [41], i.e., how to recognize if the skin is affected by such a distortion, which does not allow to acquire and/or process the acquired fingerprint.

Fingerprint Data Storage

Fingerprints are taken in compliance with ISO/IEC FCD 19794-4 and ANSI/NIST-ITL 1-2000 IS standards. The quality of the stored fingerprint has to be marked with NFIQ (NIST Fingerprint Image Quality) equal to 3 or a better grade (according to the last information obtained from STATE PRINTING WORKS OF SECURITIES, state enterprise, fingerprints of the quality levels 4 and 5 are stored in the recent chips as well). In Fig. 6.2 can be seen that a DG3 has been designed to hold fingerprint data. The maximal data size of one fingerprint is 15 kB in compressed format WSQ (Wavelet Scalar Quantization) specified in document IAFIS-0110 (V3), precisely according to the Gray-scale Fingerprint Image Compression Specification 1997. [101], [64]

6.3.3 Proposal for Further Use of Biometrics in Passports

With respect to the latest results in the sphere of biometrics, it is convenient to incorporate more biometric features into one device to ensure the quality of automatic processing of personal identities and to prevent frauds in this area. Proposed features for the future use are primarily based on iris recognition, veins of fingers recognition, and combinations of these features with time-proven fingerprints (see Fig. 6.5). Use of the iris recognition has already been prepared in current passports; however, the real application is still not common. A correctly implemented combination of aforementioned biometric features should be robust enough to provide all demanded properties of a passport system. Moreover, scanning of veins of a finger during the fingerprint scanning process should also provide liveness detection at the same time—that is a very important aspect in the fake fingerprint detection. [42], [80], [90]

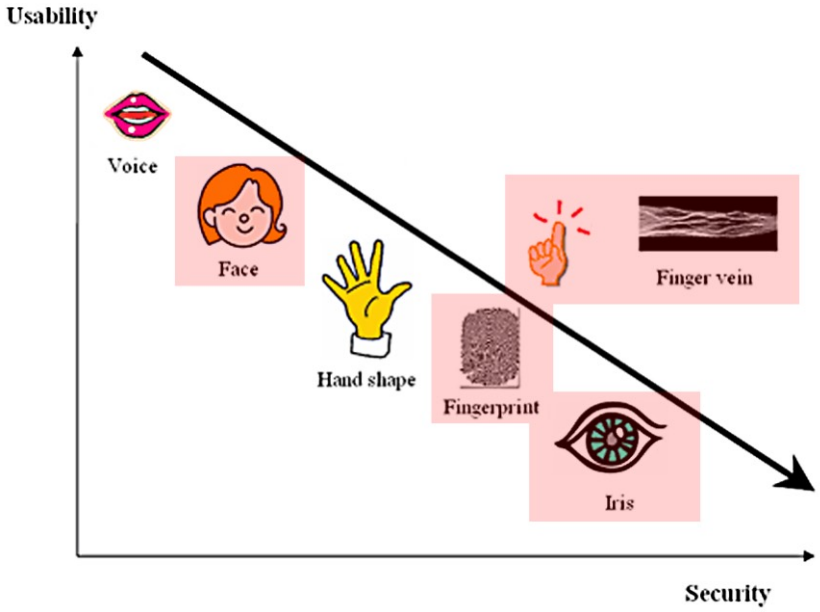


Figure 6.5: A simple comparison of biometric features with highlighted features proposed for future use in the e-Passports. (Source: http://www.hitachi-ics.co.jp/product/english/about_fv.htm)

6.4 The Czech Implementation

The Czech electronic passport was introduced as a second device of this type in the EU. The passport system architecture can be seen in Fig. 6.6. Since that time, new types of security improvements have been introduced. However, due to the backward compatibility of all solutions across the world and given minimal requirements of the ICAO, the former threats will still be present.

Despite particular rules being set by either the ICAO or consequently by the EU, there is still enough space for country-specific modifications. This results in a variety of solutions across the world. The solutions are different, but (mostly) compatible at the same time. The necessity for the variability of local solutions rises from the fact that each country has its effective law and implemented related technologies that have to be incorporated into the local passport implementation.

6.4.1 Legislative Framework for Passports in The Czech Republic

Relevant legislative background for the implementation of the mentioned security technologies were established by the European Union as a reaction to the 11th September 2001 terrorist attacks[3]. To be more precise, the preparation of such a technology started much earlier in 1981. The original proposals were set by the ICAO organization, and the EU regulations proceeded from these recommendations.

[3]The consequences of this event were undoubtedly a very strong argument for employing more sophisticated travel document technology.

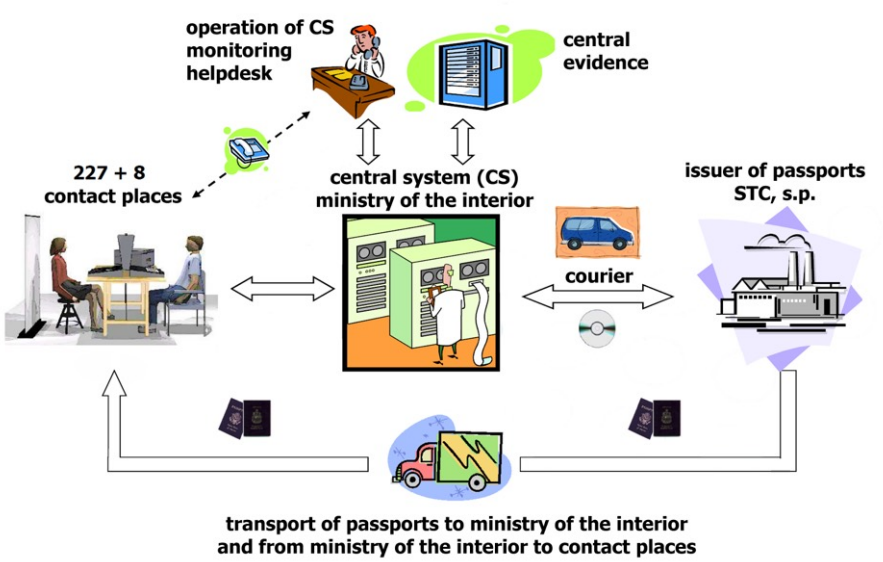


Figure 6.6: The Czech passport system architecture. (Source: Ministry of the Interior of the Czech Republic).

Let us consider a coherent approach of the EU (2003) as the first important document for the EU e-Passports implementation. A brief sequence of milestones with respect to the Czech Republic implementation came into being as follows ([51], [50]):

20th June 2003
A coherent approach in biometrics implementation for biometric documents and data for the EU citizens, third-country nationals, and information systems.

13th December 2004
The council regulation Nr. 2252/2004 concerning standards for security features and biometrics in passports was issued by the Council of the European Union.

28th February 2005
The EU Commission Decision C(2005) 409 established the technical specifications on the standards for security features and biometrics in passports and travel documents.

15th June 2005
Government of the Czech Republic issues the ruling Nr. 740 that approves the process of implementation of the European Council regulations in the Czech Republic.

23rd December 2005
Signature of a contract with a provider of the e-Passports, STC, s.p. (STATE PRINTING WORKS OF SECURITIES, state enterprise).

1st September 2006
Launch of the first stage of the project—testing of the whole process with real data.

1st April 2009
Launch of the second stage of the project—implementation of fingerprints (as the second country in the EU, after Germany).

1st December 2014
e-Passports in the EU have to support Supplemental Access Control (SAC)—the third generation of electronic passports.

6.4.2 Security

The main goal of the whole e-Passport project is to preserve the privacy of the personal data and prevent forgery of the travel documents. Different measures are used with respect to the importance of a particular aspect.

During the process of biometric passport implementation from assignment to the final product, several security aspects were treated, especially:

- Process security in general.
- Security of involved buildings.
- Mechanical and optical security elements.
- Public key infrastructure (PKI).
- Fulfillment of international standards and recommendations.
- Digital communication encryption.
- Incorporation of biometrics.

Mechanical and Optical Elements

Security elements of this type are often used not only in passports but also on, e.g., banknotes and other types of personal documents. Although we do not aim at this type of security, let us mention at least some of them: serial numbers, fluorescent elements, relief stamping, engraving, guilloches, holograms, laser perforations, mechanic perforations, watermarks, etc.

Basic Access Control (BAC)

A very simple mechanism used for protection of information stored in DG1; and DG5 (see Chapter 6.2). The BAC technique is based on two crucial principles—the first: data can be read-only in case the passport is opened on the data page. If not, the RFID chip is shielded—a communication cannot be technically established. The second: the MRZ contains information which is used for the transmission password derivation. Actually, the data in DG1 and DG5 are the same as [63] information on the passport data page (see the right bottom part of Fig. 6.1), that is why in the case the attacker has the ability to open the passport on the data page and read the MRZ, the information from the data page and thus also from the DG1 and DG5 is not secret anymore.

The keys for the BAC are derived with the use of SHA-1[4] from the MRZ, precisely from the passport serial number (9 characters), owner's date of birth (6 characters), and the date of expiry (6 characters). The result of the hash function is truncated to 16 bytes and divided into two passwords (key A: 0th–7th byte; key B: 8th–15th byte) for 3DES. A key for the main communication is then established via 3DES encoded messages. [64], [63], [104], [130], [92], [14]

[4]It is possible to use SHA-1 or SHA-2 (SHA224, SHA256, SHA384, SHA512).

Active Authentication (AA)

The active authentication serves as a protection against passport cloning. A couple of keys (private and public) are generated during the process of personalization of a new passport. The private key is stored in the part of memory that is inaccessible from outside of the chip. It is provided only within the hardware functions of the chip. The public key is freely available in DG15. [68]

The principle is then based on the asymmetric cryptography. Random data is generated and sent to a passport chip by an RFID reader. The data are signed internally with the private key stored in the chip and sent back to the reader. In the last step, the reader verifies the compatibility of the key pair and emits a result about the authenticity of the private key. [104], [64], [92], [14]

Extended Access Control (EAC)

The aforementioned BAC is definitely too weak to secure the sensitive biometric data—the fingerprints (DG3), in the future also the iris (DG4). Therefore, a new security specification was created. The EAC was specified in technical report BSI TR-03110 (Advanced Security Mechanism for Machine Readable Travel Documents—Extended Access Control). The EAC has been used in the Czech Republic since April 1, 2009. It was brought to light together with incorporation of the fingerprints [104], [64], [130], [92], [14]. Two cryptographic mechanisms are being used within EAC:

Chip Authentication (CA, based on Diffie-Hellman)
: It is an alternative to Active Authentication (protection against chip cloning). In contrast to the Active Authentication, the CA does not suffer from so-called challenge semantics. The challenge semantics can cause tracking of the owner's transfer in a specific case. That is why Germany did not include AA into their implementation of e-Passport. After the DH process, a cryptographically strong shared secret is available for encoding the following communication. [130]

Terminal Authentication (TA, based on PKI)
: Only approved terminals have permission to access the data groups with biometric data. The terminal has to be equipped with a valid certificate of a particular country to access the data. Each terminal is set to a specific self-destruction time period. The length of this period depends strictly on conditions of use of each terminal (from one shift to one month max.). Each terminal is labeled with unambiguous ID and can be blocked. [104], [64]

Supplemental Access Control (SAC)

Supplemental Access Control introduces improvements into security of the e-Documents. In particular PACE (Password Authenticated Connection Establishment) that has been introduced in order to overcome the weakness of BAC. It is capable of providing cryptographically strong session passwords on the basis of asymetric cryptography (DH). [71]

6.4.3 Introduction of New Security Principles

Generally, it is always important to employ contemporary standards, e.g., cryptographic standards, to ensure resistance to attacks against algorithms, protocols, and hardware.

Use of new techniques is also recommended in the area of travel documents by the ICAO. Nevertheless, the introduction of technologies capable of handling such new versions of passports—with new algorithms and security principles—takes a certain time. In some countries less, in some countries more. Till the time the old passport security mechanisms as, for example, BAC or manual identity check will be supported as regular principles; it will be possible to perform attacks against these poor mechanisms at least at places where the more secure mechanisms of current passports (fingerprints, iris) are not implemented.

The best solution to this situation would be to prepare a completely new revision of passports with employing only contemporary secure mechanisms. Nevertheless, it is important to ask how all countries would be able to adopt new technologies necessary for handling the new passports. This is the biggest issue that cannot be easily solved, and so the backward compatibility will always open possibilities for attackers and fraudsters.

Although there exist newer editions of the e-documents specifications and recommendations for implementation issued by ICAO [74], [72], [73], we still have to reflect the original documents from years 2006-2008 that were used with the examined implementation. Needles to say, the passports issued in 2014 with the examined chip remain valid for 10 years. From the end of 2014, the Czech biometric passports use a different chip.

Chapter 7

Microscopic Analysis

A definition of the word analysis in [66] says: "The study of such constituent parts and their interrelationships in making up a whole." A comprehensive analysis of chips would surely consist of software and hardware analysis. A combination of these two approaches leads to very complex work with good potential to reveal different types of errors or vulnerabilities present in the examined device. Each type of analysis—software and hardware—has to be taken as a completely stand-alone discipline. Moreover, even the sub-sets of HW or SW analyses can be considered stand-alone disciplines. Hardware analysis requires inter-disciplinary knowledge, practical experience, and a lot of expensive devices thus is globally more demanding than the software analysis. Let alone needed work with very dangerous chemicals.

Microscopic analysis is being used on a wide range of devices, from very simple integrated circuits representing some basic functionality, e.g., logic gates, adders, multiplexers, to very complex chips containing more separate units on a single chip, e.g., smartcards, cryptographic devices, MCUs, CPUs, GPUs.

General-purpose chips like CPUs, GPUs, DSPs, FPGAs, etc. cannot be analyzed in the same manner as specific purpose chips because they do not express their behavior (functionality) directly in the hardware, but in the form of general instructions used to form different programs[1].

Microscopic analysis can be used for various purposes—fault analysis, system-level inspection, manufacturing process analysis or simply for reverse engineering of the circuit. Nature is always the same. First, the samples have to be prepared for observation. Usually with performing the following steps—x-ray examination of the chip in a package, removing the package, deprocessing[2]), scanning with the use of an appropriate type of microscope and finally processing the data—layer images, cross-section images, EDX (Energy-Dispersive X-ray spectroscopy) analysis, etc.—in a desired way [160], [159].

In this chapter, we will go through the whole process of microscopic analysis as we did in our experiments. We will deal with all parts from obtaining the chips up to the final analysis of the acquired images. All of these steps have to be mastered in order to be able to succeed with microscopic analysis performed from scratch[3]. We spent sizeable time periods with each of the disciplines and our knowledge gained during the process of learning, and consequent experiments are described in the following lines. We have to mention that for

[1] Therefore, these chips are not among the primary aims of this thesis.
[2] Sometimes called delayering.
[3] This was exactly our scenario, building all the knowledge at our faculty completely from scratch.

decapsulation and deprocessing, we spent many hours in laboratories of our partners from the commercial sphere, where we were allowed to work with the required equipment.

7.1 Obtaining Chips

The process of obtaining the chips is usually not difficult, especially not difficult when comparing it to the following steps like decapsulation and deprocessing. For the most part, we can obtain stand-alone encapsulated chips, or even bare chips for experiments either by buying them or by getting them from manufacturers. However, sometimes it is necessary to obtain the chips from PCBs (Printed Circuit Boards), from plastic cards or from some other more exotic setups.

7.1.1 Obtaining Chips from PCB

There is more than one possible technique on how to get separate chips [97]. The often-used method is to unsolder the chips from the PCB. It is almost exactly a reversed process to the original manufacturing process (soldering chips onto the PCB). When using this method, it is recommended to keep in mind the following rules to preserve the chip functionality (in cases of such demand) after detaching it from the PCB [152]:

- The temperature of the soldering iron tip should be max. 250 °C.
- The increase or the decrease in the soldering iron tip temperature should not exceed 25 °C/s.
- The soldering process duration should be 6 seconds max.

The new types of chip packages are usually better protected against simple unsoldering, because the leads are hidden below the package body (BGA, FlipChip), or there are other ways of attaching them to the PCB (DCA—Direct Chip Attach, WLP—Wafer Level Packaging, 2D and 3D packages) [97]. Thus, it can be very uncomfortable or even impossible to get the necessary heat to the exact place with iron solder tip. In certain cases, we can use a heat gun instead. It is recommended to use the heat gun with a thin head that aims the heat to the very narrow space in order to achieve the required efficiency and to protect the chip from damage. The principle of working with a heat gun is the same as that of working with the iron solder. One has to be very careful as the heat gun hits a larger area, and this could be undesirable in many situations. Overheating of the specimens is mostly undesirable. [98]

This is definitely not the only way to acquire separate encapsulated chips. Due to modern requirements, the pitch of chip packages leads is still decreasing, and hence, the contemporary classic packages with bare leads on sides contain many very thin leads. We can call this a fine pitch package—the pitch between the leads is smaller than 1 mm. Due to that, we can simply use a sharp knife to cut out the chosen chips or, rather, the packaged chips from the PCB. There is no need to use heat with this method. Because of no heat needed, this can be considered a safer technique.

Figure 7.1: Initial phase of the microscopic analysis—obtaining the chips. Left: RFID chips in typical plastic cards. Right: Boiling nest with acetone bath. (Source: author's work.)

7.1.2 Obtaining Chips from Plastic Cards

Many RFID chips and security-oriented chips are stored in plastic cards. This is also the case for biometric passports or other personal documents. First, parts of the plastic cards with the chips inside have to be cut out from the rest of the plastic cards (see Fig. 7.1).

Then, extraction of the encapsulated chips from the plastic cover is carried out. In the early stages, we used only acetone bath in a beaker at normal room temperature. The extraction took several minutes or even tens of minutes. The plastic card was slowly becoming pliable, and it was possible to peel the plastic layers off one after another.

We improved the process with the use of a boiling nest (displayed in Fig. 7.1). The time of extraction got shortened to 1–3 minutes per piece. We simply heated the acetone to its boiling temperature (slightly above 50 °C), the plastic compound was then almost immediately removable. We highly recommend using plastic gloves for the extraction of the chips because of two reasons:

- The chips can be relatively hot, and the plastic gloves help to mitigate the heat interaction between fingers and the chips.
- The plastic compound becomes very sticky after the boiling process, and it is difficult to clean it from fingers afterward.

Out of these two points, the first one also represents the only drawback of this improvement. The temperature of boiling acetone is higher than the normal room temperatures. Fortunately, the boiling temperature is not high, and the damage of the chips is therefore not likely.

7.2 Chip Decapsulation

After obtaining the chip, the next step is to remove the chip packaging. For this purpose, two main approaches are commonly presented—etching and grinding (sometimes stated as

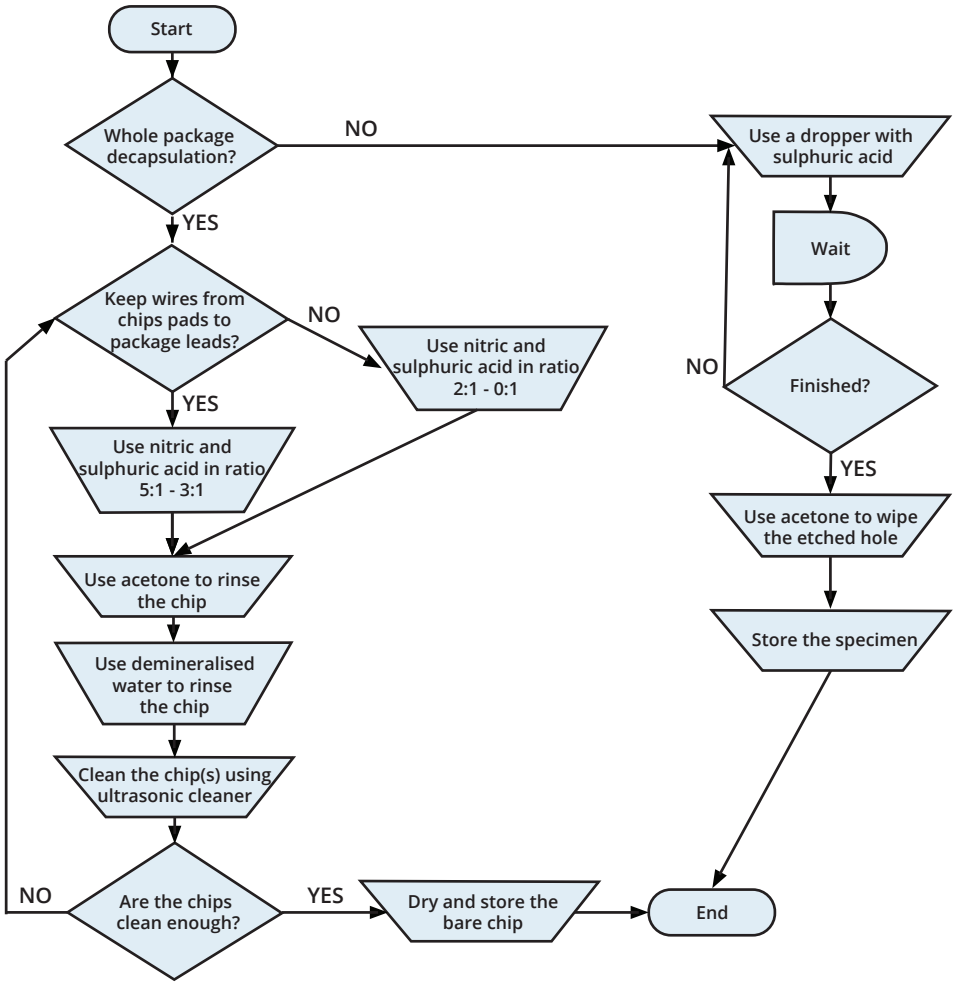


Figure 7.2: Illustration of the presented decapsulation process. (Source: author's work.)

polishing). Both are specific, and it is always felicitous to have the capability of performing both.

There are several books dealing especially with the chip packages and the related topics, e.g., [152], [25], [162], [13]. It is a very wide area, with many important aspects of knowledge. For my work, it is important to extract the information, especially regarding managed package destruction—the decapsulation.

It can be said, very briefly, that there are three main types of common chip packages with respect to their material—metal, ceramic and plastic. The most important category for us is the plastic one. These packages are widely used (also because of their low price) and are often suitable for most of the ordinary integrated circuits. The metal and ceramic packages do not occur as often as the plastic ones.

We decided to follow the mainstream packages in this thesis, the plastic packages. Up to now, almost all the investigated chip packages appertain to this group of packages. The effective approach of obtaining bare chips out of plastic packages with the use of sulphuric and nitric acids was performed within the scope of the Brno University of Technology in cooperation with Faculty of Chemistry and is described herein. [98]

7.2.1 Chemical Approach for Removal of Plastic Packages

The chemical approach is usually more convenient compared to the other ways of removing plastic packages, because of its simplicity, low time requirements, low price and availability of the chemicals involved. The fact is also that the chemicals are chosen in a manner that they react only with the plastic compounds—the danger of damaging the sample surface is then minimal because the surface of the chip is protected by a passivation layer. There exist different variants of the chemical approach that make use of the same acids, but mostly in a different step order or acids ratio [24]. However, some of them are totally dissimilar, i.e., another technique that is based on using resin instead of acids (briefly described in [137]). It has a significant drawback—it takes a lot of time. In fact, it is not recommended to follow this approach in a chemical lab, because resin vapors can foul up the exhaust. Moreover, the temperature of resin during the etching process should be really high, approx. 350 °C. This temperature may cause serious damage to the chips. On the other hand, the use of dangerous acids is completely avoided in this scenario. The process can be performed somewhere outside a building to avoid the necessity of using any exhaust. In that case, the work can be done without a proper chemical laboratory. Considering the price of resin and the fact that storing resin is safe, this approach may be sometimes more convenient than the others, especially in cases when the high temperature is out of the discussion. [98], [160], [159], [88]

Plastic packages differ one from another. Due to this fact, a certain package provides unique features suitable for a particular purpose—physical endurance, heat distribution, chemical composition, price, etc. Nonetheless, the composition is usually considerably similar, and that is why the same process can be used to remove almost all types of such molding compounds—the described process should be taken as a template because the diversity of the plastic compounds requires flexibility in performing the procedure. For a better illustration of the presented decapsulation process, see Fig. 7.2.

As mentioned above, in our case, there is a need for cooperation with a chemical facility, because we cannot avoid working with chemicals. Let us assume that all the next steps take place in a properly-equipped chemical laboratory. At least a fume cupboard, a chemical sink, and personal protective equipment should be available for the etching process. Passed safety training is also a must. Our setup in the chemical laboratory is displayed in Fig. 7.3.

It is strongly recommended to proceed very cautiously while observing the chemical laboratory rules to avoid any injuries or damage to the laboratory equipment or to the chips.

Required Equipment

This part provides a list of items needed to complete the process of decapsulation [97]. Without these, we cannot recommend carrying out the described procedure, because of potential danger—in the better case the samples could be destroyed, in the worst case, serious injury could occur.

- Sulphuric acid (H_2SO_4, concentration of 96% or more).
- Nitric acid (HNO_3, concentration of 96% or more).
- Acetone (C_3H_6O).
- Demineralized water or distilled water (in the worst case tap water).

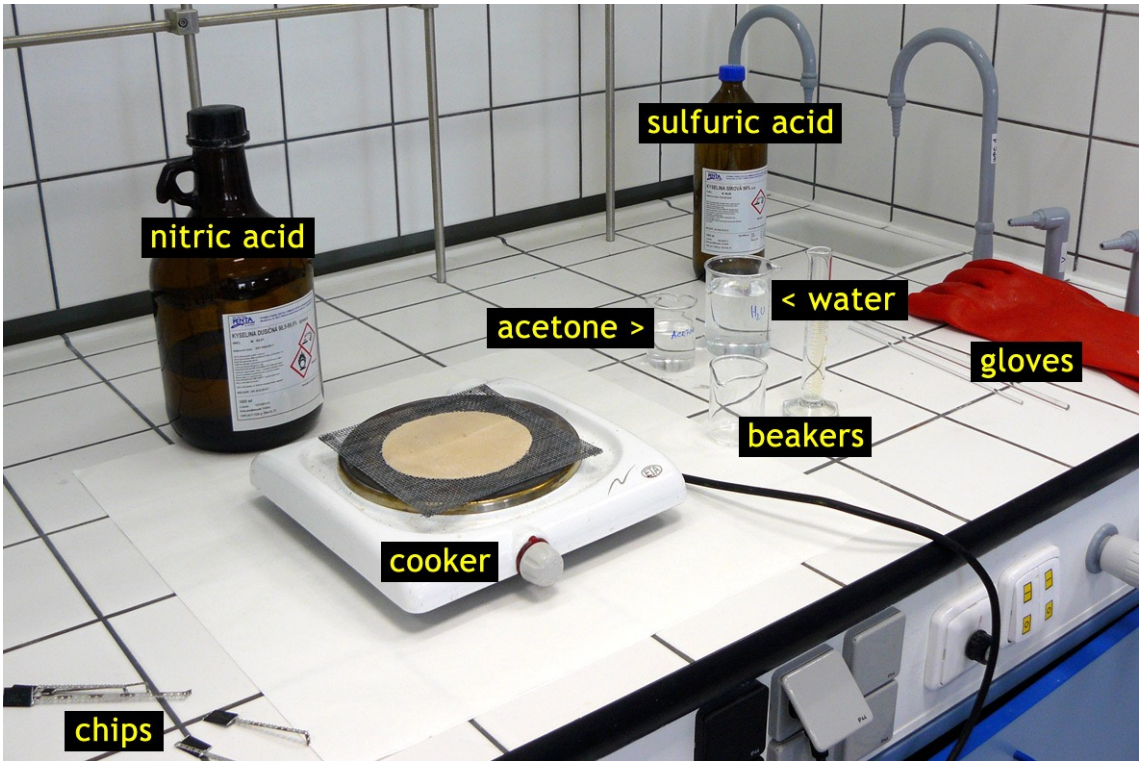


Figure 7.3: Chemical laboratory with setup prepared for chip decapsulation. (Source: author's work.)

- Ultrasonic cleaner.
- Cooker.
- Fume cupboard.
- Beakers (at least four pieces).
- Tweezers.
- Filter paper or nitrogen flow.
- Personal protection equipment (gloves, glasses, protective clothes, etc.).
- A microscope or at least a magnifier.

It is always convenient to use an X-ray (see Fig. 7.4) to get more information on the chip inside the package prior to the decapsulation process.

The Ratio of Nitric Acid to Sulphuric Acid

Before we describe the process of chemical etching, let us introduce this very important part. Prior to the beginning of the etching itself, it has to be decided what is expected as a result. According to our needs, the correct ratio of the acids mixture has to be chosen. The result is affected not only by the ratio of the acid, but also by the time duration of the

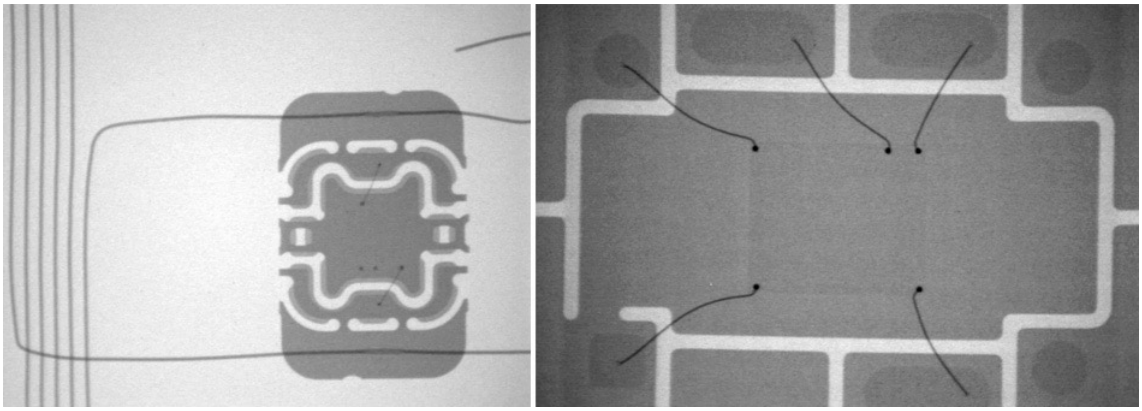


Figure 7.4: Information obtained from X-ray investigation of smartcards. Left: A chip position inside a contact-less plastic card can be revealed without destructive decapsulation. Right: A closer look at the position of a chip and its bonding inside a contact plastic card. (Source: author's work.)

active etching (the time period when the specimen is inside the acid bath) and also by the temperature of the acids.

As the compounds can be significantly dissimilar, the time of etching can differ significantly as well. We encountered the etching duration from tenths of seconds to tenths of minutes. The longest plastic package etching was performed on a type of RFID chips—about 40 minutes.

The ratio is always stated as nitric acid to sulphuric acid (the recommended approx. temperature is mentioned in brackets) [97]. Length of active etching has to be determined experimentally for each type of package.

- 5:1–3:1 (ca. 90–94 °C)—preservation of wire bonds from the lead frame to the chip.
- 2:1–1:1 (ca. 90–94 °C)—faster, cheaper, and a more aggressive decapsulation.
- 0:1 (up to ca. 270 °C)—to etch very resistant molding compounds, very aggressive.

Whole-Package Decapsulation Process.

First, four beakers should be prepared. Two beakers should be half-filled with the acids in the correct ratio (see Chapter 7.2.1). The third beaker should be half-filled with acetone and the last one with demineralized water. The cooker must be placed into a fume cupboard because of the production of dangerous vapors. Other recommended equipment should be placed nearby the fume cupboard or even inside if there is enough room for all the items. Only the beakers with acids need to be placed on the cooker in order to reach the desired working temperature. The other steps of the whole process should take place out of the cooker.

The major etching should be performed in the first beaker with the acids. The acids will become non-transparent soon due to the presence of etched molding compounds. The recommendation is to inspect the level of decapsulation during the process often periodically. When the process is almost complete, it is better to use the second beaker with the transparent acids to do the fine etching. It is necessary to monitor the progress permanently at this point. The transparent acid is ideal for this purpose (for illustration, see Fig. 7.5).

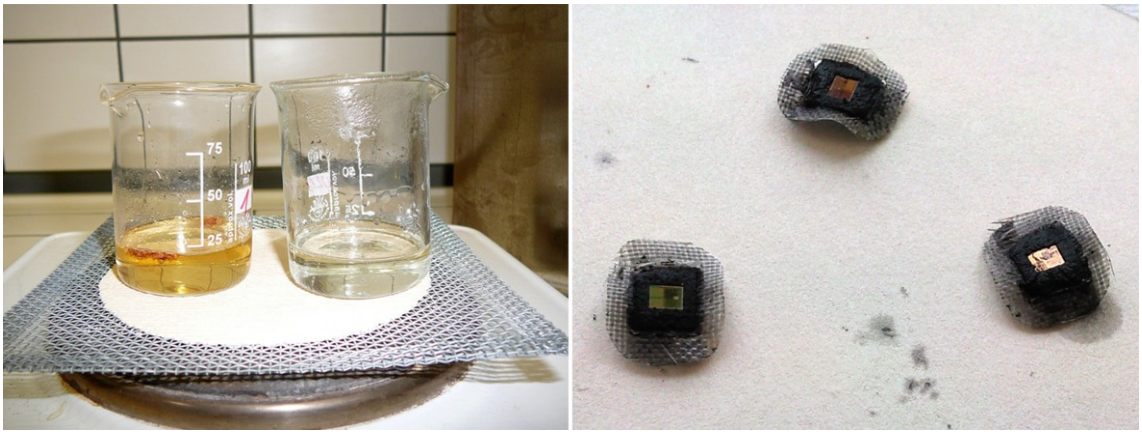


Figure 7.5: Chips decapsulation in the chemical laboratory at the Faculty of Chemistry at Brno university of technology. Left: Two beakers with acids with the decapsulation process is progress in the left one. Right: The result of the decapsulation process—bare chips on their original pads. (Source: author's work.)

We do not recommend leaving the specimens in the acid bath longer than is necessary—the lower layers could be damaged by the acid because these layers are usually not protected along the chip edges and thus so-called underetching can occur.

When the chip is bared, it should be washed in the beaker with acetone. In cases where the chip surface is bigger than 5 mm^2, we recommend using demineralized water first, acetone and demineralized water again due to safety reasons—a bigger amount of the acid left on the chip can react with acetone. Then, the specimen should be rinsed with flowing demineralized water, and it should be put inside another beaker with demineralized water afterward.

The beaker with the demineralized water and the chip or with more chips, as the case may be, should be placed into an ultrasonic cleaner for up to two minutes in the case where there is only one chip in the beaker or for ca. thirty seconds in the case when there are more chips in that beaker. With little exaggeration, the chips act as emery paper to each other. Then a final check should take place. If everything seems to be all right, the chips should be dried with filter paper or nitrogen flow. The nitrogen flow has to be used very cautiously, because the chips are lightweight and it is very difficult to find them on the floor, especially the smaller ones. If there is any problem with the cleanliness of the chip surface or similar, the entire process can be repeated [97]. When the process is finished, we should store the chips into some safe box.

Etching of Specific Packaging Part

A useful approach for some use cases is to etch out only a specific part of the chip package. The entire chip is preserved, and hence, it can be attached to other components as usually. This is needed when some measurements have to occur, or when it is necessary to preserve the whole structure of the chip. The chip surface can be observed via the etched aperture. This is the preferred method for observing the new, unknown chips because the risk of damaging the specimen is smaller. In fact, we have used this method just for the testing of

the process itself. However, the manufacturers use this method relatively often to perform specific checks of the chip surface or of the wire bonds connections, etc[4].

Because of the manual dosing of acid drops to the specific area via a dropper, this process is more demanding. First, a little hollow is made in the surface of the chip, approx. in the center of the area designated for etching. Then, the drops of sulphuric acid are manually applied to the chosen location. Nitric acid is mostly not used in this approach. There has to be a short period of inactivity between the applications of the acid drops to allow reaction with the surface. It is undesirable to apply more acid than is imperative, because the reaction is not under perfect control in that case, and therefore, the result cannot be predicted or even guaranteed. [97]

7.2.2 Grinding and Polishing of Plastic Packages

The grinding and polishing approach is suitable in cases where we cannot apply the chemicals, or the chip package allows the use of a grinder with its advantages, i.e., the chip is placed deeper in the package and there is a gap above the chip—in that case the wirebonds can be easily preserved also with use of grinder. It is necessary to proceed cautiously to avoid contact of the grinder with the surface of the chip or with wirebonds. For this reason, it is recommended to use an X-ray to inspect the starting situation and then the current level of decapsulation periodically. [98], [160], [159], [88]

7.3 Chip Deprocessing

Each conventional chip is a composition of different oxide and metal layers above one transistor layer that is formed on the silicon substrate. Nowadays, we can also encounter various 2.5D or 3D layouts, these are way beyond our capabilities, and thus, we focus on conventional layouts in this thesis. The layers across producers are made of various compounds with respect to the desired functionality and needed properties of the whole layout.

The chip deprocessing (delayering) involves three main techniques: dry etching (plasma etching), wet etching (using different types of chemicals) and polishing. As stated before, each particular type of chip is unique with an individual composition of layers. That is why an appropriate procedure has to be prepared according to the information available or obtained about that concrete chip. In some situations, it is possible to simply assume the common composition and materials present in the chip layers. Especially in QA departments of producers, where the employees are familiar with the used technologies of the producer. Simply put, oxide and metal layers are interleaving and the common materials are known. On top of each chip, there is some kind of passivation that is usually removed with dry etching. Every now and then, this basic knowledge can be enough to proceed successfully with the trial and error approach.

In any case, the best practice is to start with a cross-section analysis, including the thickness of layers measurement and materials analysis. The thickness of a layer can be as thin as ca. 30 nm, and the situation is evolving every day to still thinner layers. Only after the information is harvested from the cross-section analysis, the correct procedure can be created and carried out.

[4]This information was provided by personnel in QA laboratories of Onsemiconductor, Rožnov pod Radhoštěm.

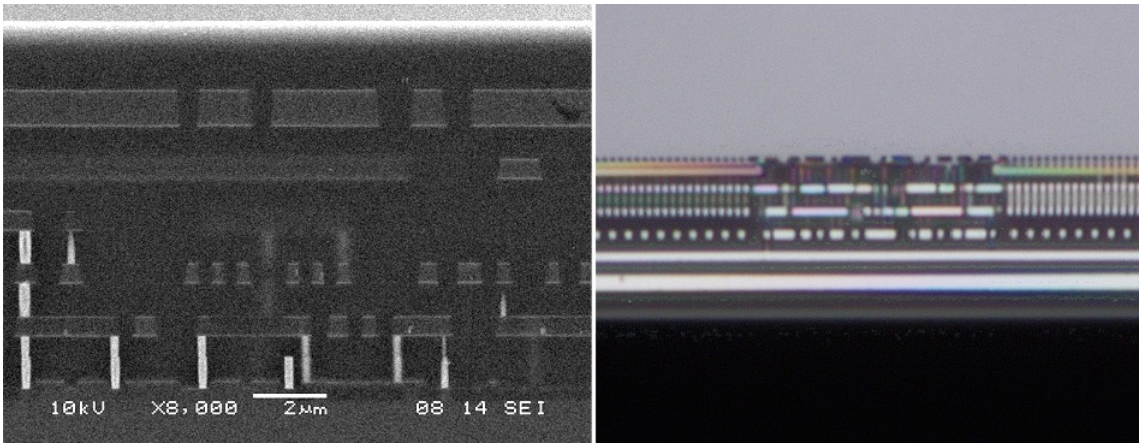


Figure 7.6: Images taken during the process of cross-section analysis. Left: Detailed image of NXP P5CD080 V0B chip acquired by an electron microscope. Right: The same cross-section displayed with an optical microscope. (Source: author's work.)

Although a suitable chemical procedure can be found for each layer, it is sometimes not possible to choose wet etching because of the high risk of damaging other layers, i.e., an underetching problem (we presented these problems in Fig. 5.1). In these cases, it is still possible to make progress with a grinder and a decent level of skillfulness. Using an ordinary grinder for removing the layers may result in the destruction of the processed specimen because the layers are very thin and so it is difficult to determine the correct time of active grinding. In addition, it is also very difficult, and with contemporary production nodes really impossible, to maintain planarity across the whole chip surface without special grinders[5]. The procedure carried out with the proper equipment is then called parallel polishing. Unfortunately, we were unable to reach such a special grinder within our partners. Thus, we had to abandon the grinding approach in our work, except for the preparation of cross-sections. [99]

One basic principle is valid for all types of procedures in this sphere—it is necessary to periodically and often inspect the current progress. The structures are miniature, indeed, and the progress of the employed technique can be faster than we could ever imagine.

7.3.1 Cross-Section Analysis

Virtually all laboratory grinder may be suitable for preparation of a specimen cross-section. It is convenient to prepare more cross-sections of one chip type always in a different position in order to gain more information about that chip. First, we use an optical microscope to inspect the process of chip polishing, see Fig. 7.6. When the results are satisfactory, we use an electron microscope to observe the layers in more detail (displayed in Fig. 7.6 and Fig. 7.7). In Fig. 7.7, thickness measures of the layers are shown in Angstrom units (1 Å = 0.1 nm). These labels are created manually by the researcher.

Displaying cross-section immediately after grinding provides the image with hard-to-distinguish borders of some layers and low distinction among these layers. Thus, to make the observability better, there are various etchants used for finishing the sample before scanning. For oxide layers, NH_4F, CH_3COOH, H_2O, and HF can be used; for diffusion

[5] http://www.alliedhightech.com/Equipment/multiprep-polishing-system-8

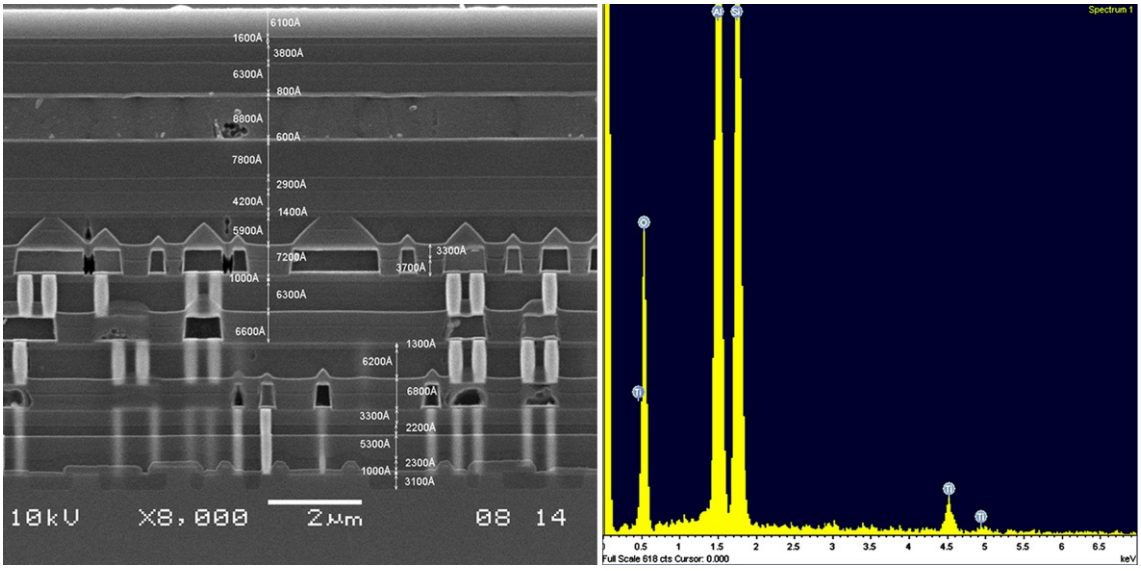


Figure 7.7: Outputs of cross-section analysis and material analysis. Left: Detailed cross-section of NXP P5CD080 V0B chip with measurements of layers given in Angstrom units (1 Å = 0.1 nm). Right: Energy-dispersive X-ray spectroscopy (EDX) of a top layer of NXP P5CD080 V0B chip. The EDX analysis was performed in order to determine the composition of the top-level layer. (Source: author's work.)

semiconductor layers, HNO_3 and HF can be used [174]. Recent FIB devices (i.e., TESCAN FERA3) allow for some limited use of this wet etching directly in-situ (in-situ chemistry is in rapid development), however only with some of the chemicals supported. Therefore, for the best results, the etching procedure has to be carried out in a chemical lab. We had the chance to create in-situ cross-section with TESCAN LYRA3 FIB, without chemical etching in-situ (see. Fig. 7.8).

After the cross-section is made and measurements are taken, the next step is to employ one of the spectrometric techniques to acquire the elements' composition. For example, an electron microscope equipped with an X-ray detector can provide such information—see Fig. 7.7. A precise deprocessing procedure can be prepared according to the composition and thickness of each layer.

7.3.2 Chemical Deprocessing

We would like to introduce a method based on the chemical approach in the following text. The method can be split into two variants. Let us name the variants as *reduced process* and *complete process*. As the names reveal to readers, the reduced process can be understood as a subset of the complete process.

Prior to a presentation of the variants of the chemical process, let us introduce the main difference between the variants. The difference is in performing or non-performing the cross-section analysis. Given in simple terms, the reduced process is experimental. For many conventional chips, the reduced process is faster, less demanding, and cheaper with keeping outcomes sufficient for further analysis. It depends on the expected outcomes and the number of sample chips available for the process of whether we chose the complete process over the reduced one or not.

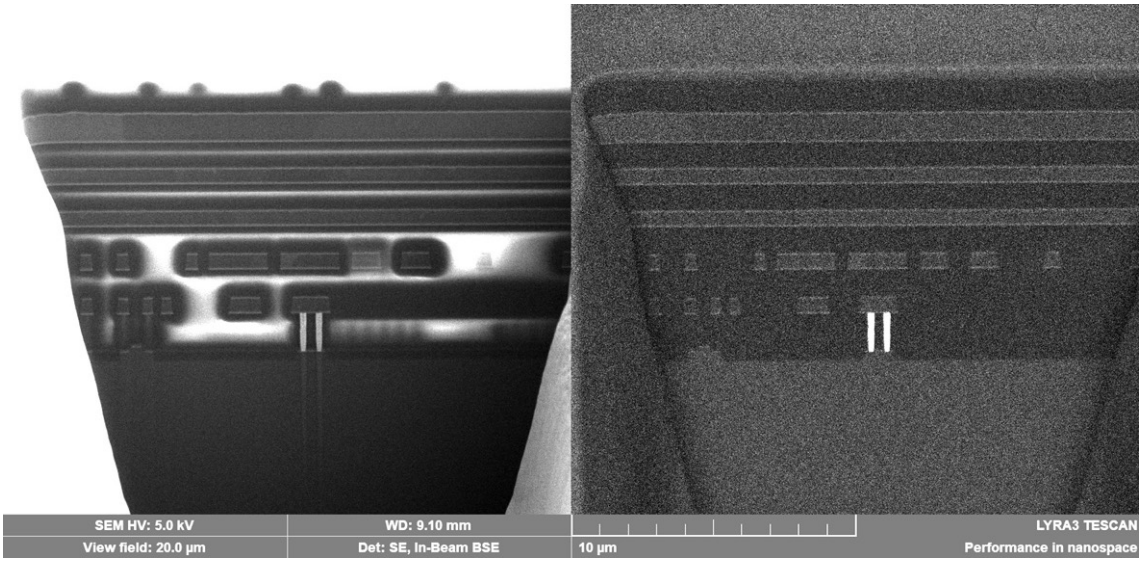


Figure 7.8: Cross-section created in-situ with FIB milling, TESCAN LYRA3. (Source: author's work.)

Reduced Process

The reduced process moves directly to the delayering of the chip without any knowledge of its composition. For removal of a specific layer, see Chapter 7.3.2. This may seem to be hazardous. It is sufficient enough for many analysis processes to apply this approach because of a typical composition scheme employed by chip producers. This method can be considered to be a type of trial and error method. The researcher simply tries to use removal methods for particular layers in a common or in a predicted order. The method is feasible due to the fact that in the case of a wrong removal method choice (we still talk about the chemical approach), virtually nothing happens—the wrong method simply does not react with a non-corresponding layer. The researcher has to be still very careful because an incorrectly chosen removal method can react with the appropriate layer edge in lower levels that are usually exposed on the border of the chip. The current top-level layer can be partially under-etched, which causes problems with lower layers removal.

For the performance of experiments, it is favorable to use this faster method. The basic assumption for success with this approach to the chemical decomposition process is that a sufficient number of chips is available—it can happen that a chip gets damaged because of wrong layout or composition predictions. The process relies on the fact that the first layer is mostly a passivation layer and so an appropriate removal method should be applied—usually plasma etching. A conductive layer composed of aluminum compounds usually follows. An insulant layer (oxide compounds) should be present to separate the conductive layers. These conductive and insulant layers then usually alternate down to the silicon layer.

We have successfully applied the presented reduced method to different RFID chips (goods tags, RFID cards). For some chips, it was necessary to experiment with the incorporation of totally different removal methods at some point of the decomposition process, because of the presence of a different layer type (layers containing titanium, wolfram, etc.). It is strongly recommended to properly inspect the deprocessing results after each particular procedure to avoid any unwanted damage to the chips. [99]

Complete Process

The complete process is recommended basically for all experiments unless we intentionally decide to go for the reduced process. This approach incorporates the analysis of layers' composition and thickness measurement of the layers as described in Chapter 7.3.1. For the complete process, it is necessary to have at least two specimens of the chip. Although succeeding with just two samples is idealistic indeed. At least one specimen is needed for the cross-section analysis and the second one for the main deprocessing. As mentioned in the previous sentence, the first one has to be sacrificed for the material composition analysis (cross-section analysis) of particular layers and their thickness measurement. The second one can be deprocessed according to the information gained from the first step. After removal of each layer, we have to acquire images of the layer with an appropriate microscope before we proceed to remove the next one. However, we strongly recommend preparing more samples than just two. Performing such analysis with only the two samples is not very common and not likely to be successful in real life. The chips can be easily damaged or lost and so relying on just two pieces is not advisable at all. The recommendation is to proceed with 10 or more samples, to have space for potential errors or to be able to prepare each sample to a different level of deprocessing—one with preserved passivation, one with removed passivation, one with a removed first metal layer, etc. [99]

Deprocessing of Common Layers

With the outcome of the cross-section analysis, an exact sequence of steps can be prepared in order to get to the coveted deprocessed chip. The form of the whole decomposition process naturally depends on our objectives. Usually, pictures of bare transistors and interconnections are desired. Although, the process can be completely different for some other use cases—just to get rid of the first passivation layer to get the ability to connect microprobes to bonding pads, to obtain pictures of all conductive layers, etc. [99]

According to the information given previously in this chapter, the decomposition process consists of a specific sequence of particular steps. The most common layers and matching chemical procedures are described below. The following information is derived from deprocessing of various simple RFID chips, MIFARE Ultralight C, MIFARE Classic 1 kB chip and NXP P5CD080.

Passivation

The very first layer on the top of the chips is mostly a passivation layer—a kind of protection against mechanical and electromagnetic effects of the environment. To remove this layer, it is recommended to use plasma etching, so-called dry etching. The whole process takes about 45 minutes with old plasma etcher TESLA 214 VT that was available at our disposal. The actual plasma etching lasts only 4 minutes out of the mentioned time period. The length of the plasma etching process depends on the type of plasma etcher and should be adjusted according to the particular machine performance. The rest of the time is devoted to preparing conditions necessary for performing this procedure.

Conductive compounds

Conductive layers that are made of aluminum compounds can be taken away by the application of *phosphoric acid etching mixture, PEWS 765-140-57-36*[6]. The recom-

[6] http://www51.honeywell.com/sm/em/common/documents/2.6_europe_msds_p_8.pdf

mended working temperature is 50 °C; the common time of the bath should be from 2 to 6 minutes, depending on the layer thickness. There exist also other commercial etchants designated for etching conductive layers, i.e., KMG Mix attaque phosphorique (mix of acetic acid, nitric acid and phosphoric acid), recommended working temperature is also around 50 °C. More of such commercial products can be found.

Dielectric compounds

A special chemical mixture is also available for removing oxide compounds—insulating material. Precisely, the mixture consists of ammonium fluoride and hydrofluoric acid in ratio 7:1. The working temperature is 30 °C; the common time of the bath should be from 2 to 6 minutes.

Deprocessing a chip down to silicon

To deprocess a chip completely down to silicon, there exists one reliable method employing hydrofluoric acid (HF), 50% concentration. It is usually used at ambient temperature. The HF acid reacts with metals and also with oxides. This acid is not selective and is often used just for this purpose to remove all layers above the silicon. Duration of the etching depends on the chip composition and the amount of material above the silicon. [30]

7.4 Layers Scanning

This part is undoubtedly the most flexible of the whole process because each researcher has to employ locally available equipment. The general possibilities are always the same. It is possible to use optical, confocal (CLSM—confocal laser scanning microscopy) or electron microscopes. All of the mentioned technologies are suitable for scanning of the chip layers. However, optical technology has reached its limitations, especially with regard to the contemporary technological nodes. The most advanced optical microscope can provide sufficient magnification for structures up to 0.25 μm [159], [160]. The confocal microscopy could be a very good alternative with better capabilities than the pure optical microscopy but still cannot provide the same magnification and details as electron microscopes. That is why SEM (Scanning Electron Microscope) microscopy is often the most common for scanning layers with large magnification. Besides the named types of microscopes, there exist various other types—-scanning capacitance microscopy (SCM), transmission electron microscopy (TEM), scanning microwave impedance microscopy (sMIM), etc. However, detailed overview of types of microscopes with their capabilities and limits is out of the scope of this thesis.

It is nothing extraordinary that the combination of high magnification and the size of the chip requires multiple image tiles of each layer. The images have to be acquired one by one, in the best case automatically by special control software. After that, stitching of the tiles has to take place to get a complex overview. Readers can easily imagine that the final image of one particular layer can contain billions of pixels, i.e., chip used in e-documents contained more than 1.6 billion pixels after the stitching process.

At Faculty of Information Technology at Brno University of Technology, our capabilities of scanning are given by optical microscope Olympus BX61 with motorized stage equipped with objectives 5×–100×; and by electron microscope Phenom Pure G2. We also cooperated with Onsemiconductor, Rožnov pod Radhoštěm laboratories, where we performed part of decapsulation and deprocessing procedures. Thus, we used their local equipment for

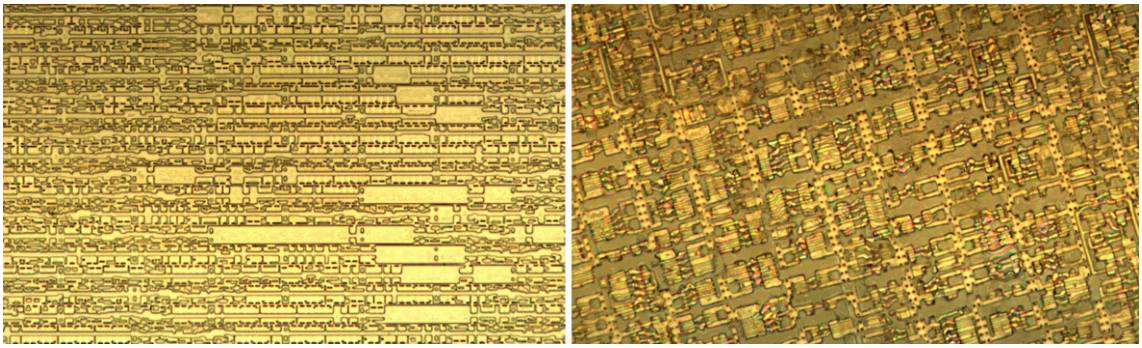


Figure 7.9: Chips produced with the use of technological node observable with an optical microscope—the images were acquired with optical microscope Olympus BX61. Left: A detail of MIFARE Classic 1 kB. Right: A detail of MIFARE Ultralight C. (Source: author's work.)

periodic inspection of the processes. Further, we cooperated with Presto Engineering, Inc. and TESCAN Brno, s.r.o., where we were able to scan the biometric passport chips.

7.4.1 Scanning with an Optical Microscope

The optical microscope is equipped with a function for automatic tiles scanning and subsequent tiles stitching. However, according to our experience, it is better to ask the software to perform tiles scanning without stitching. The stitching can be performed separately with better results, for more information about stitching, see Chapter 8.4.

Optical microscopes suffer from a problem of non-uniform focus (due to a smaller depth of focus) within a single image. This problem occurs because of the used technology. System of lenses cannot auto-handle improper height positioning of the specimen on the laboratory glass. Thus, we had to acquire each tile several times with different focus settings in order to be able to merge more images into one to get proper focusing across the whole tile. Unfortunately, our microscope is not equipped with automated axis-z that would give us better autofocus results. Therefore, we performed scanning manually when using our optical microscope . This manual work with MIFARE Classic 1 kB chip consumed two to three hours per each layer without post-processing (merging images to get proper focus, tiles stitching). A complete processing of the whole chip in this way took more than six days. We can rather recommend seeking a workplace with automated stations capable of doing these basic tasks automatically.

We used our optical microscope only for overview images of the more recent chips. Thus, we have not suffered from the mentioned drawbacks of our station. However, there still exist chips that are fully observable with use of the optical microscopes, see Fig. 7.9.

7.4.2 Scanning with an Electron Microscope

When looking at our electron microscope Phenom Pure G2, there is no automatic scanning at all. So, the microscope is again usable for manual inspections and experimental work, but not for the whole chip batch scanning. Therefore, we cooperated with TESCAN Brno, s.r.o., where they have cutting-edge electron microscopes with automated scanning.

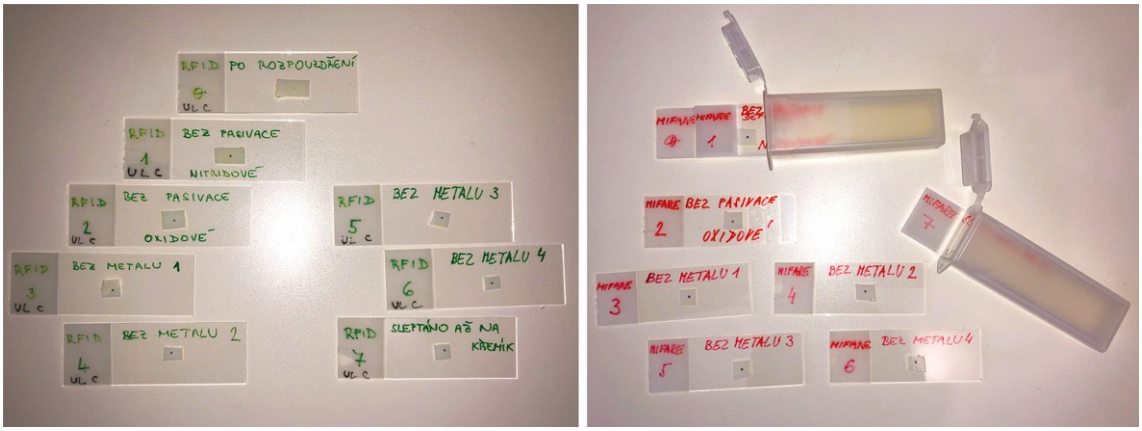


Figure 7.10: Chips deprocessed layer-by-layer and prepared for observation. Left: MIFARE Ultralight C samples. Right: MIFARE Classic 1 kB samples. (Source: author's work.)

The problem with focus described in the previous section does not appear with electron microscopes. Every single point of an acquired image is focused. That is undoubtedly a big advantage of this technology over optical technology.

With our electron microscope, we focused mainly on the semiconductor layer with areas holding transistors. The data was stored in tiles with resolution 1024×1024 in jpeg format. The highest resolution of a single tile that Phenom Pure G2 can provide is 2048×2048. Unfortunately, the speed of acquisition with the highest resolution is not suitable for fast manual scanning. That is why the majority of images taken by this microscope is stored in the resolution 1024×1024. Format jpeg was chosen above bmp because of sufficient quality, better disk space utilization and storing speed. The size of data is influenced by the resolution of the images and also by the fact that electron microscopes provide data in gray-scale only.

7.4.3 Processed Chips

The information stated in this chapter was verified practically on various simple RFID tags and on smartcard chips, namely MIFARE Ultralight C, MIFARE Classic 1 kB; partially on MIFARE DESfire EV1 and NXP P5CD080 (SmartMX family). The last mentioned is described in detail in Chapter 8. The MIFARE samples were deprocessed and a sample of each layer is stored as depicted in Fig. 7.10. After investigation of the chips, we realized that the MIFARE Classic samples were not original, but very popular clones produced by Shanghai Quanray Electronics Co., Ltd. The chips were labeled as QR2217.

Fig. 7.11 depicts differences between the decapsulated MIFARE Classic 1 kB with a preserved passivation layer (see the left part of the figure) and the same specimen after five steps of the decomposition process (see the right part of the figure)—4 minutes of plasma etching (passivation); 4 minutes of etching—use of PEWS (aluminum compounds); 4 + 2 minutes of etching—use of ammonium fluoride and hydrofluoric acid in ratio 7:1 (oxide compounds); 5 minutes of etching—use of PEWS (aluminum compounds); 3 minutes of plasma etching (removing the last layer above transistors). All phases of MIFARE Classic deprocessing, including transistor field detail, can be seen in Appendix B. Demonstration of other processed chips—MIFARE Ultralight C and MIFARE DESfire EV1—is shown in Fig. 7.12.

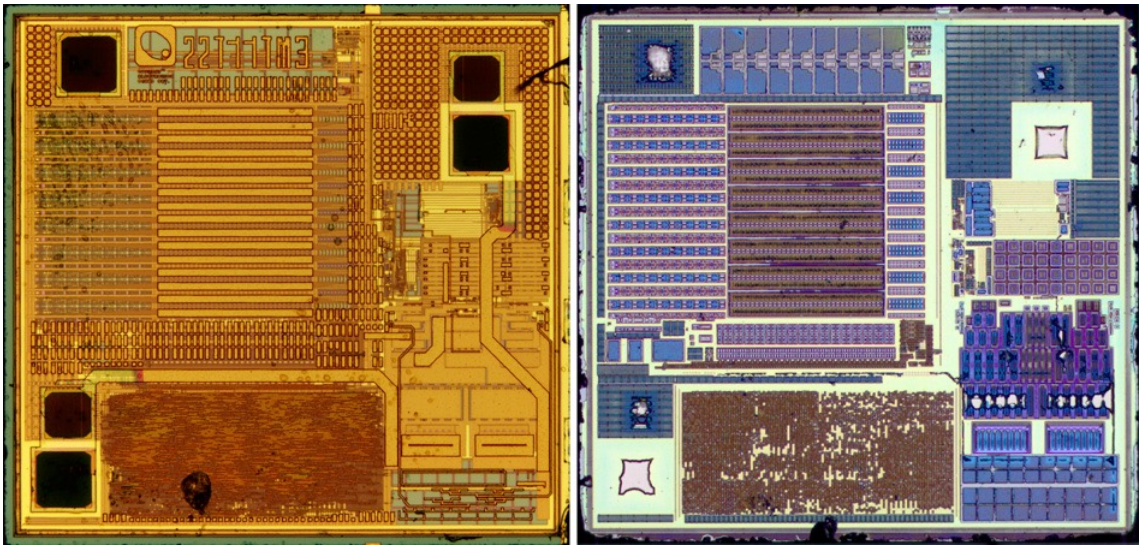


Figure 7.11: Two images of MIFARE Classic 1 kB. Left: The decapsulated chip without any application of decomposition. Right: The same chip after five steps of deprocessing. (Source: author's work.)

7.5 Analysis of The Images

It is a big difference if the analysis is performed by the manufacturer for a concrete reason—i.e., after unexpected outcomes of tests during production process—with lots of data regarding the chip structure available for the analysts; or if the analysis is carried out by an independent facility providing security or functionality audits on the devices (or even circuit reconstructions). Complex analysis with no informational support from the manufacturer is hardly feasible with recent chips within a short period of time. Hence, we always try to get as much information as possible and perform goal-directed actions.

The analysis process also depends on the form of image data—separate tiles scanned with/without overlapping, a complete image of each chip layer, etc. Demonstration of such analysis process can be found in Chapter 8—starting with obtaining the chips, decapsulation, deprocessing, image scanning, image stitching, information gathering, understanding of the basic structures, etc.

The image data for analysis is in bitmap format by nature of the scanning—the meaningless bitmap pictures have to be mined for information either completely by hand (not recommended for the recent chips) or with use of some more or less automatic software. The software can provide many functions applicable to these kinds of data—finding conductive layers (reconstruction of interconnections), (semi-)automatic functional blocks recognition, blocks annotation, image filtering (or preprocessing) for assuring better results of the algorithms, scheme reconstruction, etc. However, the software available publicly is usually limited to some kind of technology or technological node, i.e., rompar for parsing masked ROMs[7]; or not maintained properly, e.g., Degate[8], pr0nsweeper[9]. Although Degate tries to be a complex software kit accompanying the researcher up to the scheme reconstruction, it

[7] https://github.com/ApertureLabsLtd/rompar
[8] https://degate.org
[9] https://github.com/JohnDMcMaster/pr0ntools/tree/master/capture/cf

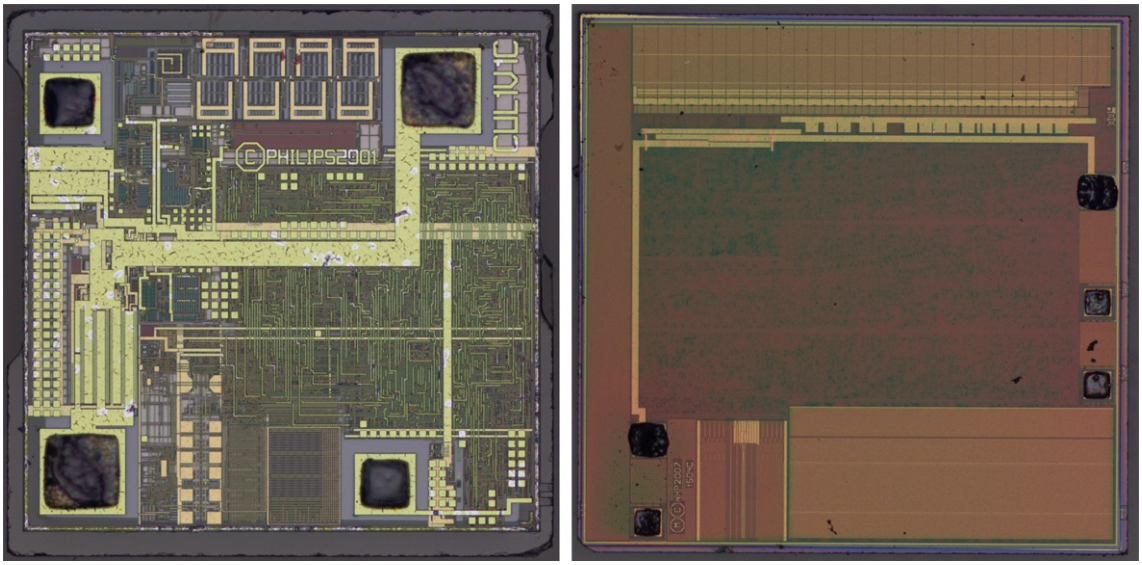


Figure 7.12: Demonstration of other decapsulated chips. Left: MIFARE Ultralight C chip. Right: MIFARE DESfire EV1 chip. (Source: author's work.)

is usable rather for smaller chips with a lot of manual work needed and a lot of patience with its instability.

On the other side, there are commercial software packages that are not publicly available. There exist only presentations and marketing materials promoting the quality of these products—i.e., ChipJuice[10]. Chipworks Inc., a company based in the US (operating web TechInsights.com[11] with analyses of various products), represents one of the biggest players on the market. Their software solution can be considered state-of-the-art in chips analysis—although we have no real possibility to verify the statements—because it implements many techniques and provides possibilities of connections with designers and other software. [160], [159], [88]

[10] https://www.texplained.com/about-us/chipjuice-software/
[11] https://techinsights.com

Chapter 8

Analysis of The Czech Biometric Passport Chip

In this chapter, we will present results of low-cost physical analysis of the chips used in the Czech biometric passport implementation. Because the passports hold sensitive information like digitally stored fingerprints, facial image usable for automatic face recognition and several other pieces of personal data, a certain level of security is expected. Our aim is to detach the chip from its outer package (from the plastic card), decapsulate it and try to perform deprocessing with consequent image data analysis.

These chips were obtained from STATE PRINTING WORKS OF SECURITIES, state enterprise (STÁTNÍ TISKÁRNA CENIN, státní podnik[1]), encapsulated in plastic cards, the same as used in real passports implementation. We had to pass through all the steps beginning with the extraction of the chips from the plastic cards and ending with image processing and data analysis.

It was clear from the very beginning that the biggest issue will be the integration density and the number of layers. The times of MIFARE Classic being observable under optical microscope are gone for good (i.e., size of edition of MIFARE Classic 1 kB die examined by us is 0.9×0.86 mm with technological node approx. 0.25 μm). The size of the examined chip is 2.76×2.76 mm and when used together with technological node 0.14 μm, it ensures the enormous number of transistors and interconnects inside.

8.1 Extraction and Decapsulation

The plastic cards used for this particular biometric passport revision are produced in compliance with the recommendations of ICAO (International Civil Aviation Organization) [70], [69] issued in 2008 and 2006. Although there are more recent editions [74], [72], [73] of the documents (these are used in the very recent implementation), we intentionally mentioned the ones that were used as the foundation for the samples that we received.

Extraction from the plastic cards was performed as described in Chapter 7.1.2. We encountered no extraordinary issues when carrying out this task, which is why it was described so briefly in this chapter.

We approached the decapsulation of these chips experimentally, based on our previous experience with other RFID chips extracted from smartcards. Moreover, we had enough

[1] https://stc.cz/en/

samples of the chips and we were not aiming to preserve the wire bonds. Thus, we decided to go for the simplest method.

We used sulphuric acid (H_2SO_4, 96% concentration), heating it inside the fume cupboard to 278 °C. We inspected the degree of decapsulation every minute. It turned out that the compound was very resilient. During the process we had to use 3 separate beakers with the acid bath, because of contamination of the acid with the molding compound. After we were satisfied with the cleanliness of the chip's surface, we rinsed it in an acetone bath and with demineralized water afterwards. The quality was periodically checked with an optical microscope. Finally, we used ultrasonic cleaner to clean the chip. The whole process took 15 minutes, 12 minutes out of the whole process length was the duration of acid bath. The process was verified and confirmed with a second specimen. After the process verification, we successfully decapsulated the other 12 specimens in the same manner.

8.2 Deprocessing

The next step, after obtaining the chips from plastic cards and removing their packages, was to deprocess the samples into separate layers. Because of the size of the die and because of the expected complexity, we decided to go for the cross-section analysis first.

8.2.1 Cross-Section Analysis

The cross-section (see Fig. 8.1) was prepared according to our recommendations given in Chapter 7.3.1. The cross-sectioned sample was glued to a glass pad for easier manipulation. The prepared specimen was immediately scanned with an electron microscope in order to provide more details about the layer count and composition. 5 standard metal layers were identified, plus one special layer called LIL (Local Interconnect Level) what was directly above the polysilicon layer. This layer enables extra interconnections among polysilicon structures and also with the M1 layer. In other words, it allows for more connections among the transistors. The metal layers are made of aluminum (Al). LIL layer is made of tungsten (also called wolfram, W), the same as contacts and vias. The structure of the chip gained from the cross-section analysis is as follows (top to bottom):

- Top passivation layer is made of standard silicon dioxide (SiO_2) with a layer thickness of 1.2 μm.
- After passivation, on the top of M5, there is a small barrier layer. It is 53 nm thick, made of titanium nitride (TiN) or Ti/TiN stack.
- The thickness of the M5 metal layer (Al) is 1.31 μm.
- Barrier (TiN or Ti/TiN stack) between M5 and M5/M4 is 60–70 nm thick.
- Dielectric M5/M4 thickness is 925 nm. Vias are made of tungsten (W) with a diameter of 390 nm.
- Barrier (TiN or Ti/TiN stack) between M5/M4 and M4 is 60–70 nm thick.
- M4 layer (Al) is 481 nm thick.
- Barrier (TiN or Ti/TiN stack) between M4 and M4/M3 is 65–75 nm thick.

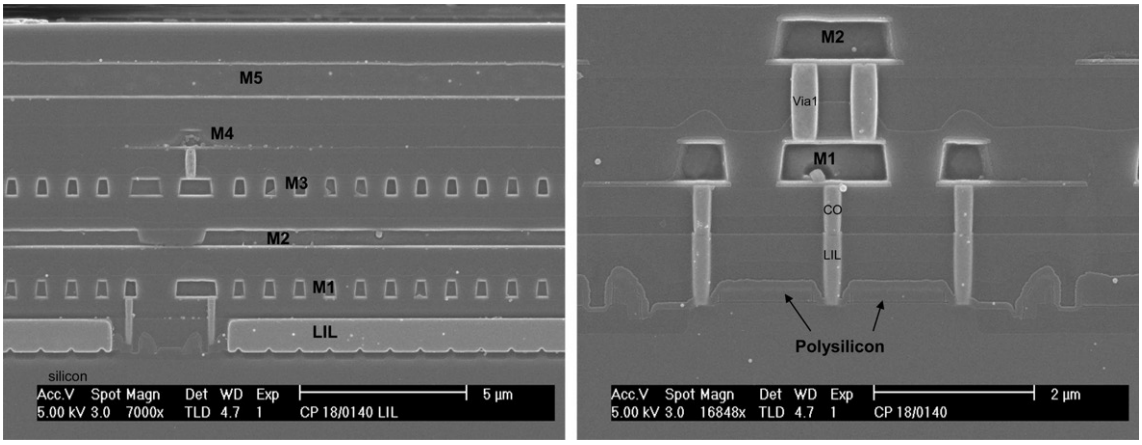


Figure 8.1: Cross-section of the NXP P5CD080 V0B chip. Left: Complete chip structure. Right: Detailed view displaying part of the chip from the silicon level to the M2 layer, also depicting details of the LIL layer and Vias. (Source: author's work.)

- Dielectric M4/M3 thickness is 864 nm. Vias are made of tungsten (W) with a diameter of 352 nm.
- Barrier (TiN or Ti/TiN stack) between M4/M3 and M3 is 85–95 nm thick.
- M3 layer (Al) is 475 nm thick.
- Barrier (TiN or Ti/TiN stack) between M3 and M3/M2 is 70–80 nm thick.
- Dielectric M3/M2 thickness is 923 nm. Vias are made of tungsten (W) with a diameter of 340 nm.
- Barrier (TiN or Ti/TiN stack) between M3/M2 and M2 is 90–100 nm thick.
- M2 layer (Al) is 486 nm thick.
- Barrier (TiN or Ti/TiN stack) between M2 and M2/M1 is 75–85 nm thick.
- Dielectric M2/M1 thickness is 920 nm. Vias are made of tungsten (W) with a diameter of 340 nm.
- Barrier (TiN or Ti/TiN stack) between M2/M1 and M1 is 75–85 nm thick.
- M1 layer (Al) is 483 nm thick.
- Barrier (TiN or Ti/TiN stack) between M1 and M1/Si is 75–95 nm thick.
- Dielectric M1/Si thickness, including LIL and the last barrier, is 1582 nm. Vias are made of tungsten (W) with a diameter of 290 nm.
  - LIL (W) layer thickness is 990 nm.
  - Barrier (TiN or Ti/TiN stack) between M1/Si and Si is 100–150 nm thick.
- Shallow trench isolation (STI) thickness is 326 nm. This layer prevents current leakage between semiconductor elements and is used with CMOS nodes starting from 250 nm and smaller.

8.2.2 Removing Layers

When we had cross-section analysis done, which gave us information about the chip composition, we advanced to chip deprocessing. We did not investigate dielectric materials in very much detail, because we detected standard silicon and oxygen elements with use of EDX. This was enough to assume standard silicon oxides.

First, we deprocessed one sample completely down to silicon through the use of hydrofluoric acid (HF, 50% concentration, ambient temperature, 40 minutes of etching), which reacts with metals and oxides as well. This acid is not selective and is often used just for this purpose of removing all layers above the silicon.

Then we continued with one-by-one layer removal performed on a different specimen. The first step was to remove the top passivation (SiO_2) with a dry etching technique. Then we continued with the removal of the barrier right above M5 layer (TiN). Dry etching with the ultrahigh frequency (UHF, 2450 MHz) option turned on was used for this task. The UHF allows for high density plasma generation. After the first barrier, we had to remove the M5 layer. Commercial etchant *Mix attaque phosphorique* from company KMG at temperature 45 °C was successfully used for this purpose. As it is known from the cross-section analysis, another barrier layer follows M5. However, we could no longer use the UHF option with the plasma etcher—it was out of operation shortly after the first barrier removal. The only available option was to use polishing to get rid of the barriers. We had Buehler Ecomet grinder-polisher at our disposal. Unfortunately, without special planar-oriented features. The polishing method was not performed successfully, just because of our inability to maintain planarity across the whole chip surface (see Fig. 8.2). Without the prospect of getting the UHF plasma etching fixed within a reasonable time and with problems during polishing attempts, we were unable to acceptably remove the barrier under M5. The UHF plasma etching would likely be the only reliable method available to us. This was for us a show-stopper with the idea to get the whole chip deprocessed layer by layer.

At this point, we have to state that in case we had the UHF plasma etching fully operational, we would have been able to reach the originally specified goal—having all layers separately deprocessed and scanned. It has to be concluded that a fully operational laboratory is an inevitable precondition for the successful performance of complete deprocessing such (or more advanced) chips. The final results of deprocessing for the scope of this doctoral thesis are the three layers available for further scanning—the top-level, M5 and silicon layers.

8.3 Image Scanning

Although it was not possible to get all layers deprocessed, we were able to scan at least the top-level layer and the silicon layer, as that is the most interesting one. When starting observations of the samples after transportation to and from the chemical labs of our partners, we discovered several damaged ones—see Fig. 8.2. It is necessary to take such losses into account as well. This occurred even though maximum care was expended. The chips are extremely fragile.

For the main scanning, we were allowed to use TESCAN MIRA3 SEM in our partner's laboratory, TESCAN Brno s.r.o. This cutting-edge SEM allows high-speed scanning with low-noise imaging, thus providing very good output image quality, which is important for further image processing.

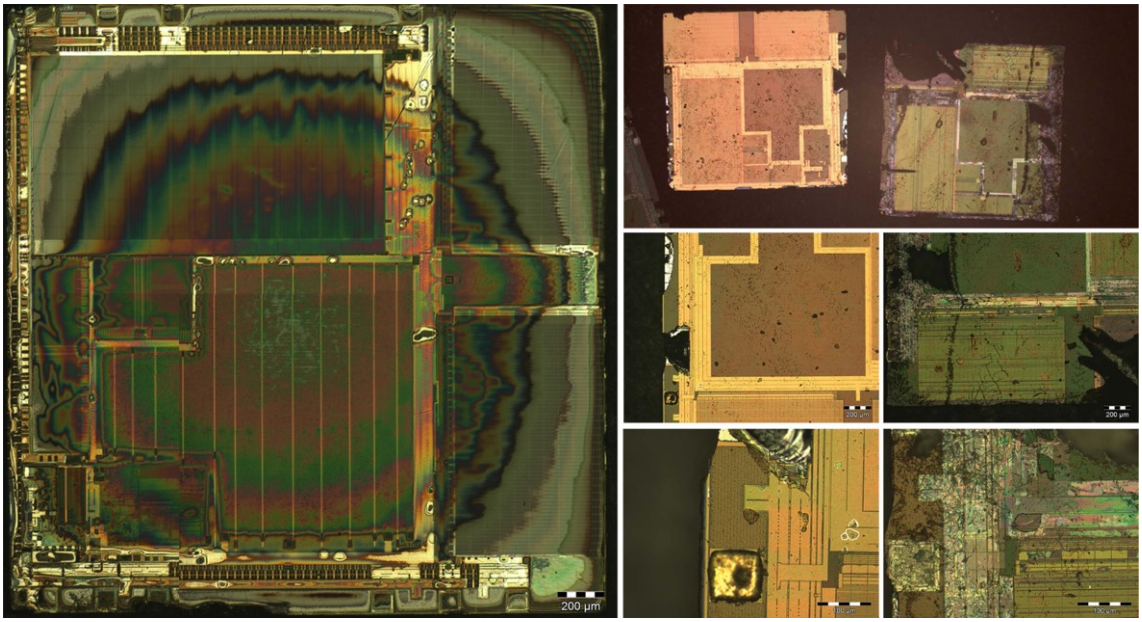


Figure 8.2: Various damages induced to the examined samples. Left: Planarization issue caused by a grinder during performance of parallel polishing. Right: Damages caused by transporting between laboratories and manipulation of the samples. (Source: author's work.)

For achieving a reasonable scanning speed, we decided to go for a 768×768 resolution of each tile with a 4152× magnification and overlapping between tiles set to 10%. These settings resulted in 3422 tiles, 59 rows×58 columns. The scanning duration of this particular setup was 15 hours and 12 minutes. There was also overhead with preparation of the sample and searching for the best setup for the scanning procedure, lasting almost two hours. We intentionally did not use the built-in stitching software found within the MIRA3 system in order to avoid processing errors that may occur with stitching. Such an error would cause the need for re-scanning the sample. And because we had access to the machine only for a very limited time-frame, we instead gathered the data in the form of separate tiles for later processing.

Side by side the transistor layer scanning, we manually scanned the top surface of the chip as well. For this overview scanning, our local optical microscope, Olympus BX61, was used with the employment of a 10× objective lens. This setup resulted in a 3×3 matrix of tiles with about 60% overlapping. Although the chip's surface was not clean enough, it provided sufficient quality for the needed overview of the surface and position of bond pads. Results of scanning can be observed in the following section.

8.4 Image Stitching

Due to significant number of tiles needing to be stitched in the silicon-level layer, manual stitching of the overall image would mean a lot of time spent with GIMP, Photoshop, or similar software. Nowadays, there exist various applications for panoramic image stitching, however, we considered a recommendation from discussions within our university and that

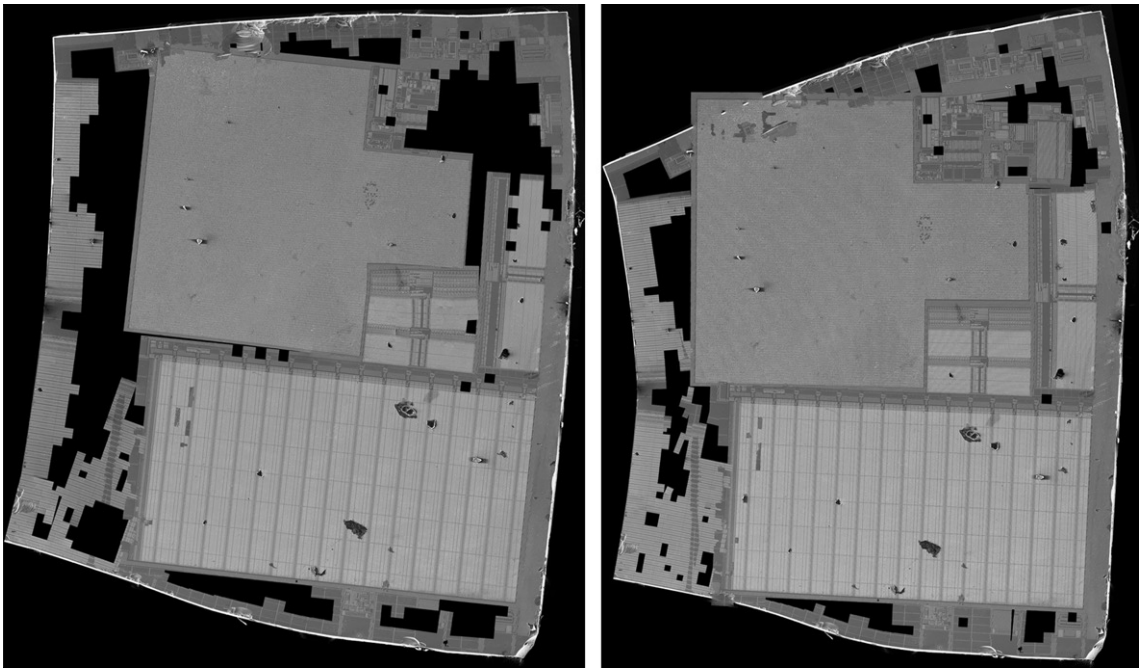


Figure 8.3: Faulty results of stitching attempts performed with Microsoft Image Composite Editor 2.0. (Source: author's work.)

was Microsoft Image Composite Editor 2.0[2]. The user interface of this software allows for the automatic detection of the scene from imported images; this automatic detection did not work with our data and the computed image was always significantly distorted. The next attempt was to manually input the parameters of the scene—number of columns and rows, image order, horizontal and vertical overlapping, angular range, and space for searching for matches. Although we tried various tweaks in regard to the settings, we have not received a satisfactory output; see Fig. 8.3.

After receiving a result not corresponding to the initial idea, we decided to search for an alternative software that could be able to carry out the task acceptably. We came across the software ImageJ (Image Processing and Analysis in Java)[3], and consequently to its bundled version, FIJI (a distribution of ImageJ that includes many plugins contributed by a community). Later, we also discovered the software called MIST (Microscopy Image Stitching Tool)[4] that was also available as plugin for ImageJ/FIJI. This tool was promising from the very beginning, as it was under the auspices of NIST (The National Institute of Standards and Technology), described in [20] and also still actively developed on GitHub.

Actually, we succeeded with the first attempt using the FIJI package, even without the MIST plugin. For this computation, we used the plugin accessible via the "Plugins" menu, choosing "Stitching and finally Grid/Collection stitching". The settings of the very first attempt is shown in Fig. 8.4. The obtained result was already acceptable for further analysis, although there were 9 missing tiles in the image, partially displayed in Fig. 8.4, due to not enough matching elements found (displayed as black squares). Few tiles were not

[2] https://www.microsoft.com/en-us/research/product/computational-photography-applications/image-composite-editor/

[3] https://imagej.net/

[4] https://pages.nist.gov/MIST/

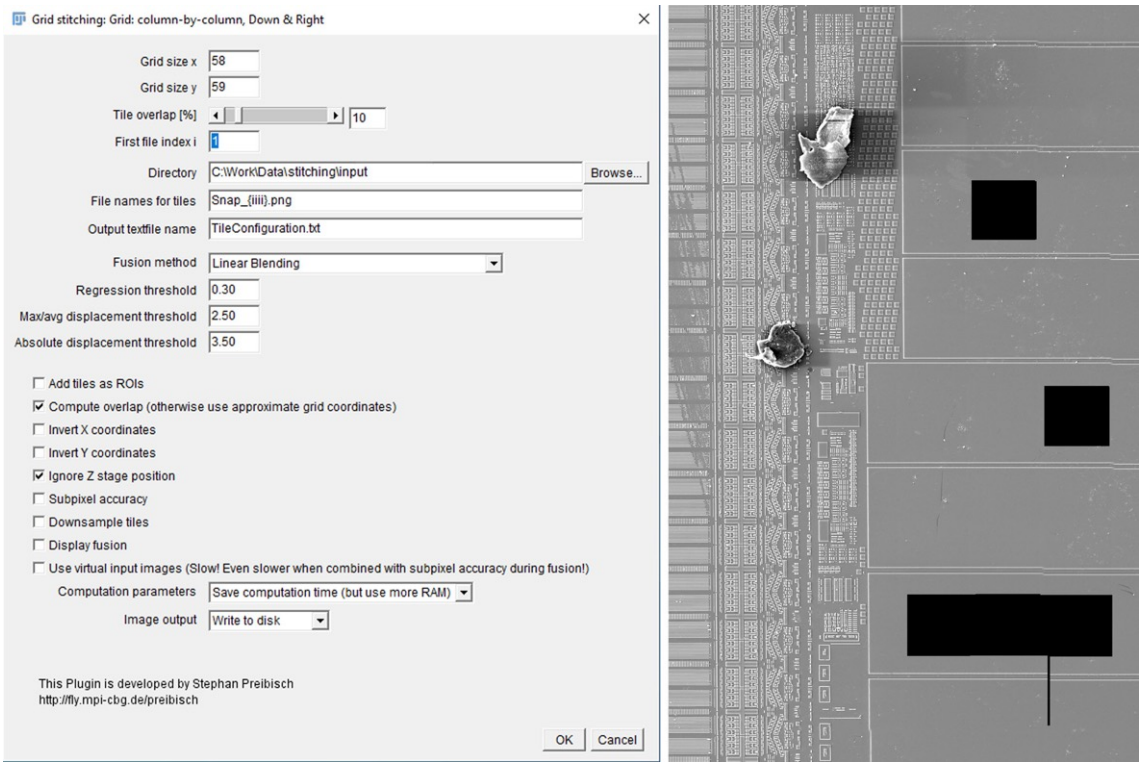


Figure 8.4: The first stitching setup with use of FIJI. Left: Settings screenshot of the very first attempt of stitching performed with FIJI Grid/Collection stitching plugin. Right: Problem of missing tiles after otherwise successful stitching with FIJI Grid/Collection stitching plugin. (Source: author's work.)

absolutely precisely aligned. However, compared to the previous results, this outcome was almost immediately usable. The missing tiles were placed into the overall image in Adobe Photoshop and the image was exported. Total resolution of the image was 40365×40612 pixels. This computation took 2 hours, 46 minutes, and 15 seconds and occupied 10.5 GB of RAM in total (including the RAM occupied by the OS). It is important to have enough free RAM memory for the FIJI software, otherwise the software reports an error during the computation and the computation is thusly terminated. We experimented with the Grid/Collection stitching plugin settings further. We repeated the settings of the initial computation, just turned on also Subpixel precision checkbox. The computation occupied 17.9 GB RAM memory (including the OS demands). Surprisingly, the duration was only 2 hours, 29 minutes, and 35 seconds (less than without sub-pixel precision). This outcome was also missing the 9 tiles, however the overall quality of the stitching was the best achieved across all experiments with FIJI and MIST. This image was updated with the missing tiles and later used as the final output for further analysis. It was possible to read from the logfiles that the set overlapping (10%) was actually corrected by the algorithm to approximately 11% and the longest part of the processing was just the overlapping and positioning calculation. We also manually checked several tiles and the overlapping among them; the value was varying from 82 pixels (10.7%) to 98 pixels (12.8%).

The next steps in experiments with the built-in Grid/Collection stitching were with a switched off calculation of overlapping, thus relying on the value given to the algorithm through its user interface. These results were very fast—5 minutes and 10 seconds for standard precision; 6 minutes 25 seconds for sub-pixel precision—the same amount of RAM was required; 10 GB and 18 GB, respectively (including OS demands). The output was—as expected—not precisely aligned and the whole grid structure was visible across the whole stitched image. Although several settings of the overlapping percentage value were tested, we were not able to stitch the whole image properly. It seems that uniformity of the overlapping is not precisely achieved neither with such an advanced SEM machine. On the other hand, the result is definitely sufficient for fast preview. Another plus of this method was no missing tiles in the final image, the tiles are simply "blindly" placed according to the given settings.

After experiments with the built-in plugins, we decided to test the MIST tool that is available as a plugin for FIJI. Installation is easy, through the integrated updater in the FIJI software package. Generally speaking, MIST-tool-based calculations were even faster than Grid/Collection stitching computations with overlapping calculation switched off (based on the settings we entered, the calculations took anywhere from 4 to 25 minutes, RAM consumption was oscillating around 7 GB), moreover the results were visually better than the ones from Grid/Collection stitching with overlapping calculation switched off. However, Grid/Collection stitching with overlapping calculation switched on was anyway the best calculation setup from the stitching quality perspective.

A summary of quality issues in particular setups is stated in the following lines:

- Grid/Collection stitching, overlapping calculation, standard precision:
  - the second best output quality,
  - problems in section with EEPROM (see the left part of Fig. 8.5, the stitching of this part of the chip was evidently problematic for all algorithms and setups),
  - very few artifacts across the whole chip image.
- Grid/Collection stitching, overlapping calculation, sub-pixel precision:

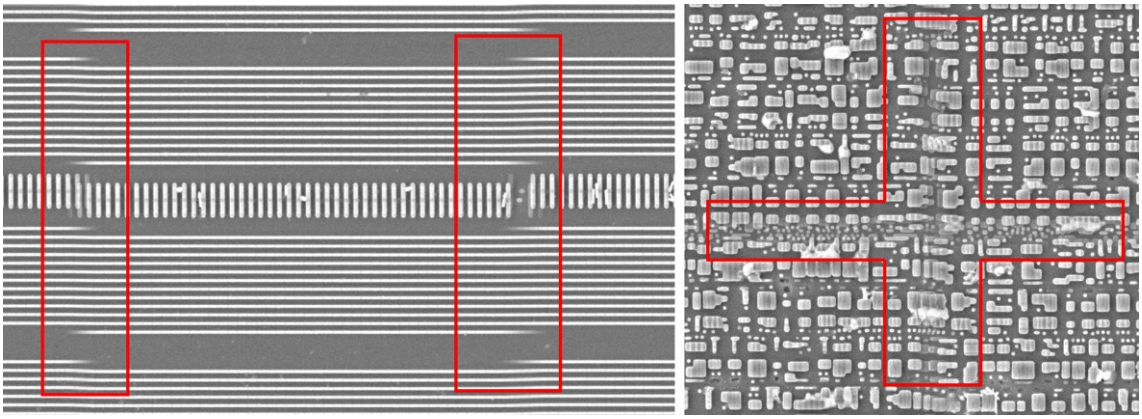


Figure 8.5: Faulty parts of the stitching performed with FIJI Grid/Collection stitching plugin. Left: Improperly stitched part of EEPROM after stitching with overlapping calculation turned on (this part of the chip was stitched improperly with all tested algorithms and settings). Right: Visible seams after stitching with overlapping calculation switched off. (Source: author's work.)

  - the best output quality,
  - problems in section with EEPROM,
  - virtually no other artifacts across the whole chip image.
- Grid/Collection stitching, no overlapping calculation, regardless precision:
  - clearly visible stitching grid (see the right part of Fig. 8.5),
  - artifacts all around the chip form a clearly observable stitching pattern.
- MIST, various settings:
  - no stitching grid visible,
  - however, there are many artifacts in essentially all parts of the chip (EEPROM, RAM, ROM, transistors, etc.)—see Fig. 8.6,
  - we were unable to tune settings to get results similar to Grid/Collection stitching with overlapping calculation switched on and sub-pixel precision.

After completing the experiments, we can recommend the MIST plugin for fast previews and for manual visual analysis of the chip. However, for software processing of the image data, we strongly recommend using Grid/Collection stitching with overlapping calculation switched on and sub-pixel precision.

The FIJI tool offers also other stitching options possibly suitable for our task, however, these are in menu under label deprecated and thus we have not tested them with our data set (e.g., 2D Stitching, Stitch Grid of Images, Stitch Sequence of Grids of Images, Stitch Collection of Images, Stitch Directory with Images). All of them were only quickly evaluated with the result that Grid/Collection stitching feature is the most suitable option for our purposes.

The used version of FIJI was 20170530. The stitching jobs were carried out under OS Windows 10 Pro 64-bit, with CPU Intel Core i5-4690, 32 GB DDR3 RAM, and a 160 GB SSD SATA hard drive. With MIST the plugin, there exists a possibility to speed up the

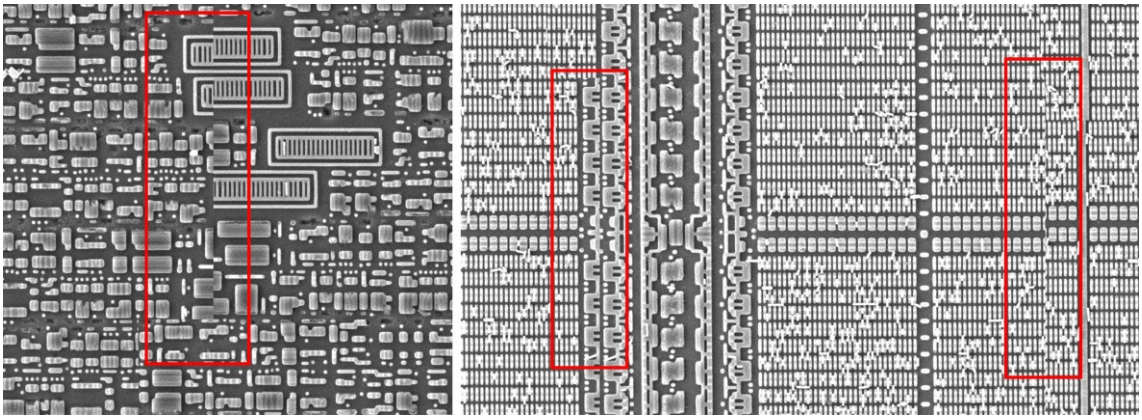


Figure 8.6: Faulty parts of the stitching performed with the FIJI MIST plugin. Left: Improperly stitched portion of transistors. Right: Wrongly stitched ROM memory. (Source: author's work.)

computations by employing CUDA; however, this was neither needed nor tested within this thesis.

Finally, there was also the need to stitch the manually taken tiles displaying the top-level layer of the chip (the 9 tiles acquired by optical microscope Olympus BX61). We did this manually with Adobe Photoshop, because there were only 9 tiles in the whole matrix (actually, we used only 6 of them, because the overlapping was more than 60%), each tile had a resolution of 4140×3096. The manual stitching was chosen because of different color tone of the images. As such, we could control the settings of each individual tile to fit perfectly to the overall image, see Fig. 8.7. The final image resolution was 6400×6400 pixels.

As a result of this portion of the work, we have several versions of the big image displaying the silicon layer of the chip. We also have the single stitched image depicting the top layer of the chip (for both results see Fig. 8.7).

8.5 Analysis

We first had to determine the identity of the chip itself. A few labels were discovered in the chips surface images, mark "T035B" together with "PHILIPS" copyright and another mark, "017", in top left corner of the chip (see Fig. 8.8, note the mesh around the labels, the mesh is present across the whole chip), right between the bond pads. This gave us the first hint as to what to look for. With the help of Google, we found out that the chip was NXP P5CD080 V0B. [117]

At the beginning of the observation, we realized one subjective detail—the MIFARE Classis chip looks ridiculously small and simple compared to the NXP P5CD080 V0B, belonging to NXP SmartMX family.

8.5.1 Datasheet Information

Based on the chip identification, we were able to gather the chip's specifications (see Fig. 8.9): [117]

- CPU—Secure_MX51.

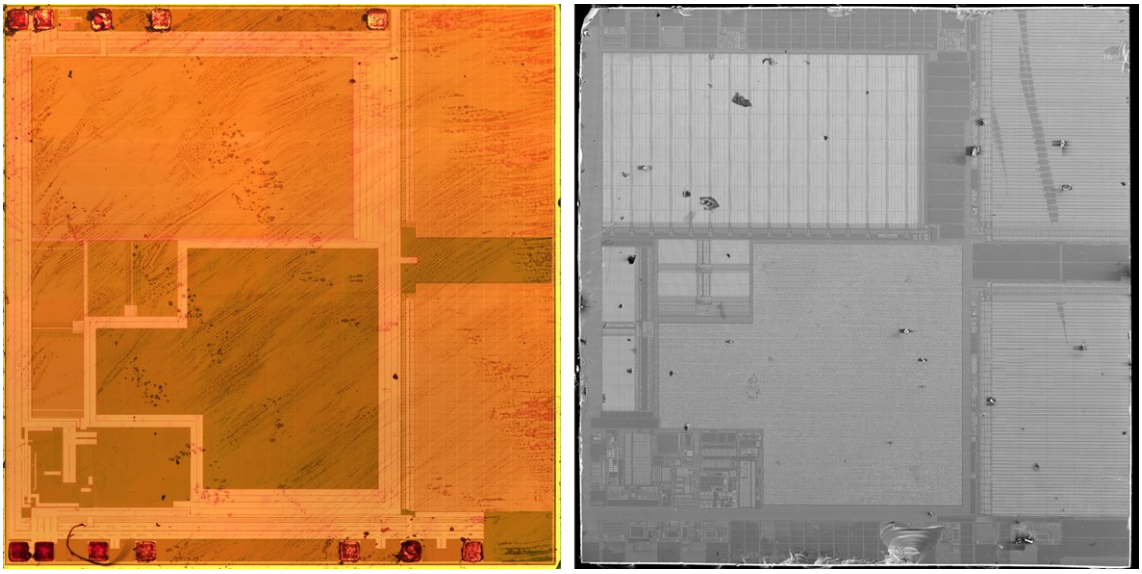


Figure 8.7: Final results of stitching. Left: Top-level layer stitched manually with Adobe Photoshop—9 tiles in total. Right: Silicon-level layer stitched with FIJI Grid/Collection stitching plugin, overlapping calculation switched on and sub-pixel precision—3422 tiles in total. (Source: author's work.)

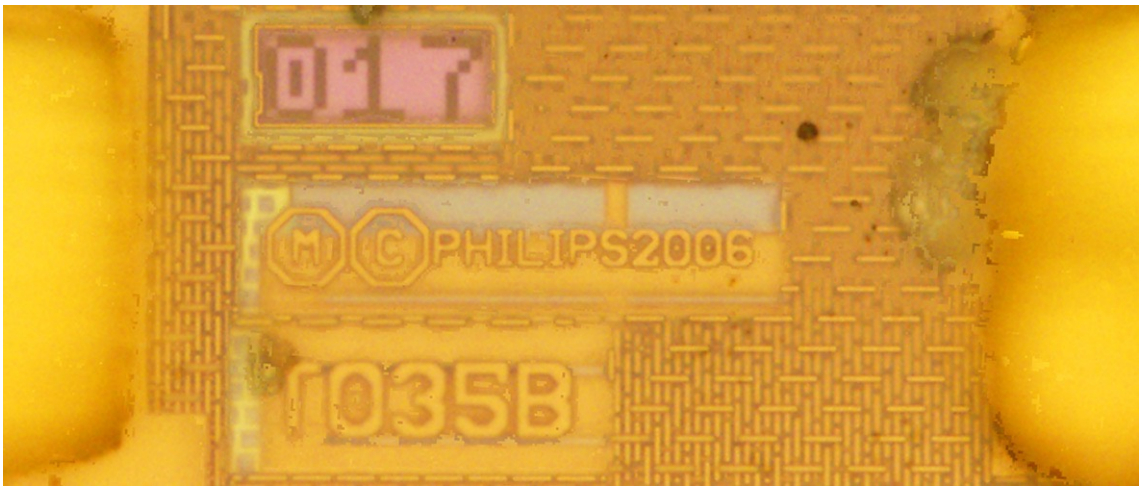


Figure 8.8: Label on NXP P5CD080 V0B chip surface. (Source: author's work.)

- ROM—224 kB[5]:
  - 200 kB application ROM (204800 bytes).
  - 24 kB test ROM (24576 bytes).
- RAM—6 kB:
  - 3.5 kB standard RAM (3584 bytes).
  - 2.5 kB accessible by FameXE engine (2560 bytes).
- EEPROM—80 kB (81920 bytes).
- PKI Crypto-engine FameXE.
- Exception sensors:
  - Low and high supply voltage.
  - Low and high clock frequency.
  - Low and high temperature sensor.
  - Light sensors (included integrated memory light sensor functionality).
  - Single fault injection (SFI) attack detection.
  - Active shield (and associated sensor).
  - Power-up and power-down reset.
- Other modules:
  - DES-Engine.
  - AES-Engine (key lengths 128/192/256-bit).
  - Memory management unit (MMU).
  - Universal asynchronous receiver-transmitter (UART).
  - CRC-Engine.
  - 2× 16-bit timer.
  - True random number generator.
- Interfaces:
  - Contact interface, ISO 7816.
  - Contact-less interface, ISO 14443.

Regarding the physical security of the chip, there are several measures implemented to avoid or hinder physical attacks. According to the [117], there are protection mechanisms against the following security incidents:

- Inherent information leakage.
- Physical probing.

[5]In datasheets and overviews, it is usually presented as 200 kB of ROM. Only application ROM is counted; test ROM is not included.

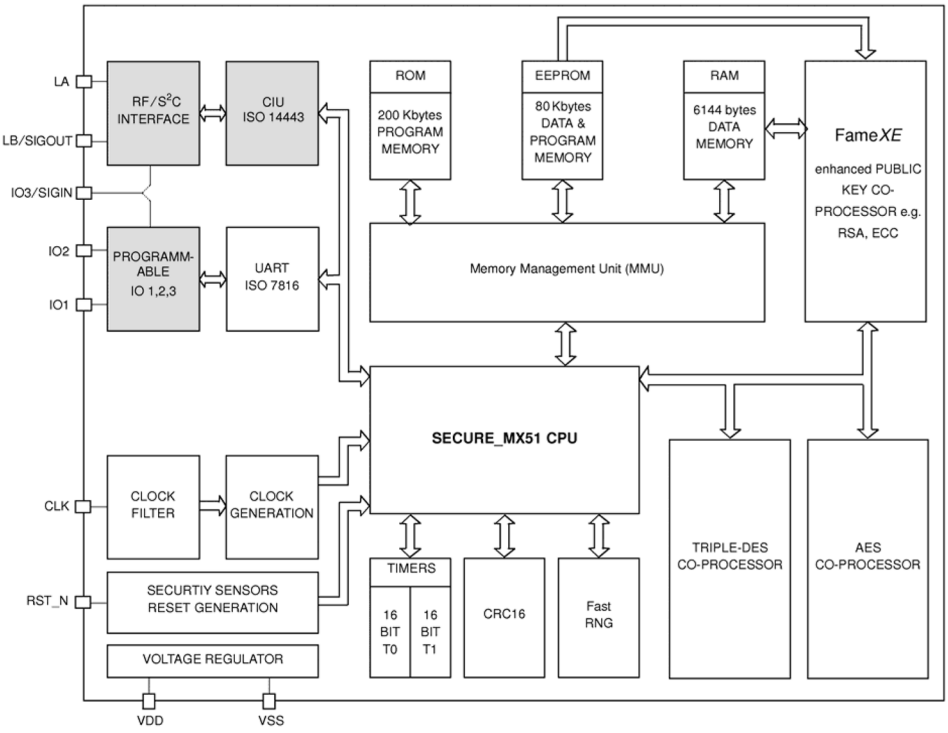


Figure 8.9: Scheme of chip NXP P5CD080 V0B. (Source: [117])

- Physical manipulation.
- Malfunction due to environmental stress.
- Forced information leakage.
- Abuse of functionality.
- Deficiency of random numbers.

There are also several other measures implemented for ensuring the security of the data and functions provided by the chip—separate CPU modes with memory access control mechanisms, multi-application support, strict data separation, etc. For a complete list, please see the datasheet [117]. The chip is also compliant with Smartcard IC Platform Protection Profile [117], [8].

Out of the obtained information, one security-related concern has presented itself. Behind the copyright symbol on the chip's surface, there is the year 2006; the datasheet, and also security reports from the year 2007. This means that in the Czech implementation of e-Passports, technologies older than 10 years are still in active use. Although this chip has not been used for production of new passports since the end of 2014, there are valid passports traveling around the world holding the examined chip.

8.5.2 Bond Pads Identification

First of all, we need to identify the bond pads and project them from the top layer to the silicon layer. We know that the chip supports two interfaces; contact (ISO 7816) and

Figure 8.10: Bondpads in the top edge of chip NXP P5CD080 V0B. (Source: author's work.)

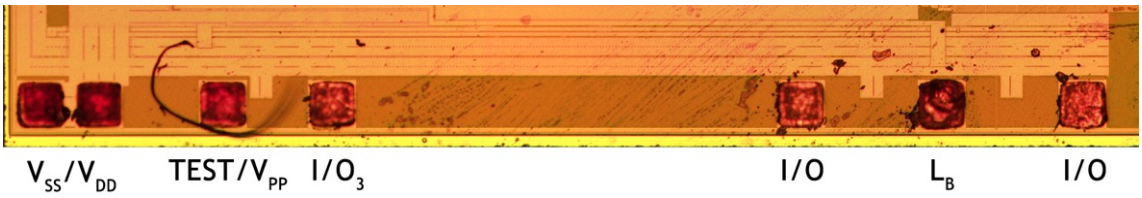


Figure 8.11: Bondpads in the bottom edge of chip NXP P5CD080 V0B. (Source: author's work.)

contact-less (MIFARE and ISO 14443A). It can be read out from the datasheet [117] that we should observe the following contacts:

- V_{DD}—positive power supply (contact interface, ISO 7816).
- V_{SS}—negative power supply (contact interface, ISO 7816).
- CLK—clock (contact interface, ISO 7816).
- RST—reset (contact interface, ISO 7816).
- IO1—input/output 1 (contact interface, ISO 7816).
- IO2—input/output 2 (contact interface, ISO 7816).
- IO3/SIGIN—input/output 3/signal in (contact interface, ISO 7816).
- LA—antenna coil connection A (contact-less interface, MIFARE and ISO 14443A).
- LB/SIGOUT—antenna coil connection B/signal out (contact-less interface, MIFARE and ISO 14443A).

The contact-less interface requires two pads for the antenna connection. These connections should be ideally on the opposite side of the chip, because of the bonding of the antenna—position of antenna pads can be found in the datasheet. It would not make sense to place these pads next to each other and then connect one of them to the opposite side to the antenna pad (antenna pads can be found in the datasheet [117]). So, we tried to find two pads with similar structures around them on both sides of the chip. Surprisingly, there were just two pads like this. Moreover, when trying to track connection visible from the top layer, both lead to the area of the chip, where we expected communication interface—non-transistor structures with present capacitors. So, the first two pads were marked. Later, when we identified I/O pads, we estimated also that from the locality perspective, the LB contact could be located somewhere near to the I/O pads.

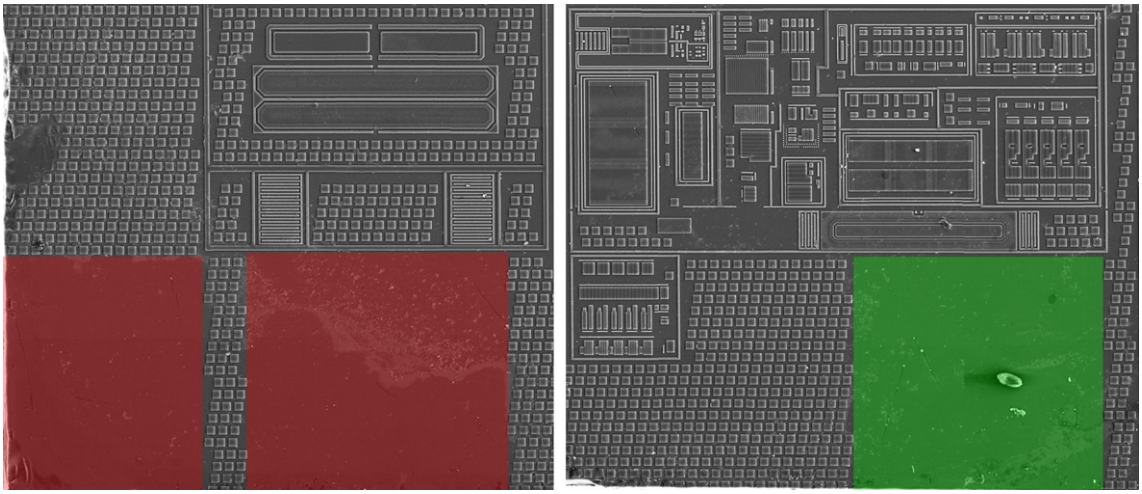


Figure 8.12: Bond pads and their surroundings present in the NXP P5CD080 V0B chip. Note the small rectangles scattered around the pads; these are so-called dummy features. Left: Couple of V_{DD}/V_{SS} bond pads. Right: I/O bond pad of the chip. (Source: author's work.)

The rest of the pads belong to the contact interface. Altogether, we identified 12 bond pads on the chip's surface (see Fig. 8.10 and Fig. 8.11). What is interesting is that the two top-left pads and two bottom-left pads are very close to each other, whereas the rest of the bond pads maintain a certain mutual distance. The two pairs of close bond pads could be suitable for power transmission (V_{DD}, V_{SS}) not to overload a single pad with the power needed for the full-power performance of the contact interface. When observing the top-level layer, traces leading from these two couples of pads go around the whole chip. This is also a sign of power routes because these are usually in the top metal layers (the top metal layer is the thickest one; thus it is suitable for power distribution across the chip) routed all around the chip.

Based on similarities of the surrounding elements, we experimentally determined the three I/O pads (see Fig. 8.12, note the small rectangles scattered around the pads, these are so-called dummy features that are placed on the silicon layer to fill in the empty spaces in order to provide support for the upper layers). From the remaining bond pads, two were conspicuously similar in the top; thus both were marked as CLK/RST (we do not know which one is which). Then, there is still one pad remaining, although all pads officially stated in datasheet have been marked. We would expect this pad to be devoted to VPP (according to ISO 7816) or to TEST I/O. The producers reserve very often one or more pads just for testing purposes.

8.5.3 Chip Segments Identification

The next step leading to understanding what is under the hood means identifying segments of the chip. Then it is possible to make a decision regarding how to further examine the parts in order to gain more information about the IC. As was already stated, we have two layers of the chip at our disposal—the top layer and the most bottom layer. A combination of these two viewpoints can give us at least some information about the chip.

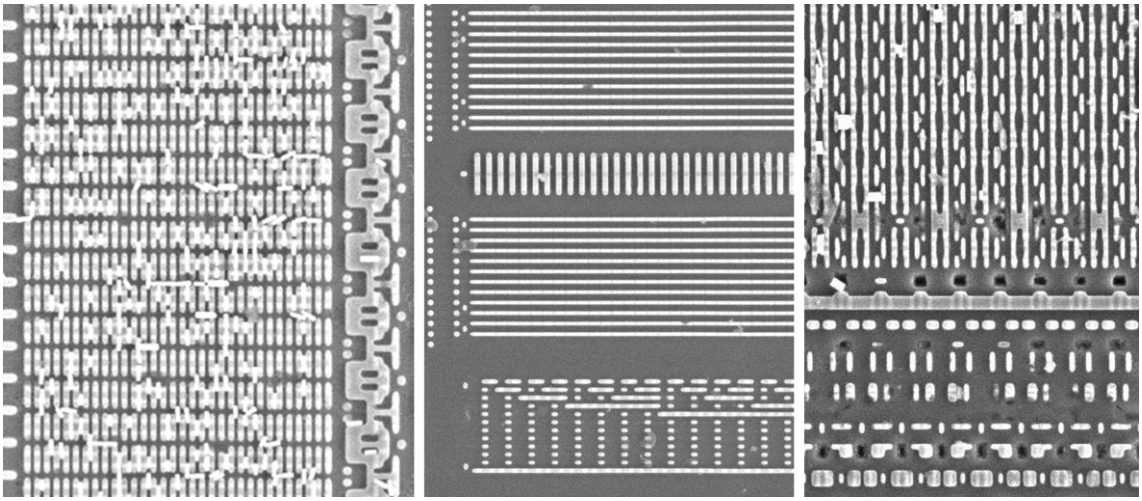


Figure 8.13: Memory elements present in the NXP P5CD080 V0B chip layout. Left: ROM memory. Middle: EEPROM memory. Note the address bits at the bottom. Right: RAM memory. (Source: author's work.)

It was clear at first sight, where the logic can be found. The transistor structures are obvious. Then, there are few sectors that evince repeated, regular patterns. These sectors are very likely the memory parts. Based on the size of each type of memory read out from the datasheet [117], it is possible to distinguish among them—for illustration, see Fig. 8.13.

The ROM memory should contain 224 kB. We were able to identify 224 lines in the presumptive ROM sector, each line holding 1024 bytes. Originally, we thought the bright particles scattered around the ROM structures could reveal the content of the ROM. However, after deeper investigation, we realized that these particles have to be considered remainders of structures from the layer above, basically telling us nothing about the actual content.

The RAM memory is expected to be as close as possible to the CPU. It was stated in the datasheet [117] that only a certain amount of RAM is accessible for the FameXE coprocessor. This RAM could be part of a single block of RAM cells; however, we identified in the image, that this part of RAM is also physically separated. This hypothesis about memory block assignment was confirmed after the precise counting of cells in each block of RAM. We clearly identified 1792 bytes in each of the two rightmost blocks of RAM, that corresponds to the application RAM size—3584 bytes. Consequently, the rest of the RAM cells must have been the part accessible for FameXE. It was again confirmed by counting of the cells in the blocks—4×640 bytes mean the perfectly fitting 2560 bytes of RAM for FameXE.

After determining ROM and RAM, we wanted to confirm EEPROM in the same way, with the counting of the cells. EEPROM is divided into two halves placed on the right, each half containing 40960 bytes. There were 32 segments found in a vertical direction and 320 in a horizontal direction. That would mean 4 bytes per each of the 32 vertical segments—when looking at the vertical segment; it is divided into four similar parts.

The chip was segmented into parts—memory sectors and transistors. Just the last part was not assigned, the violet segment in the bottom left part of the chip displayed in Fig. 8.14. When zooming in to this part, elements resembling capacitors are present in here. Thus, we deduced that this must be the segment responsible for wireless communication (the wireless

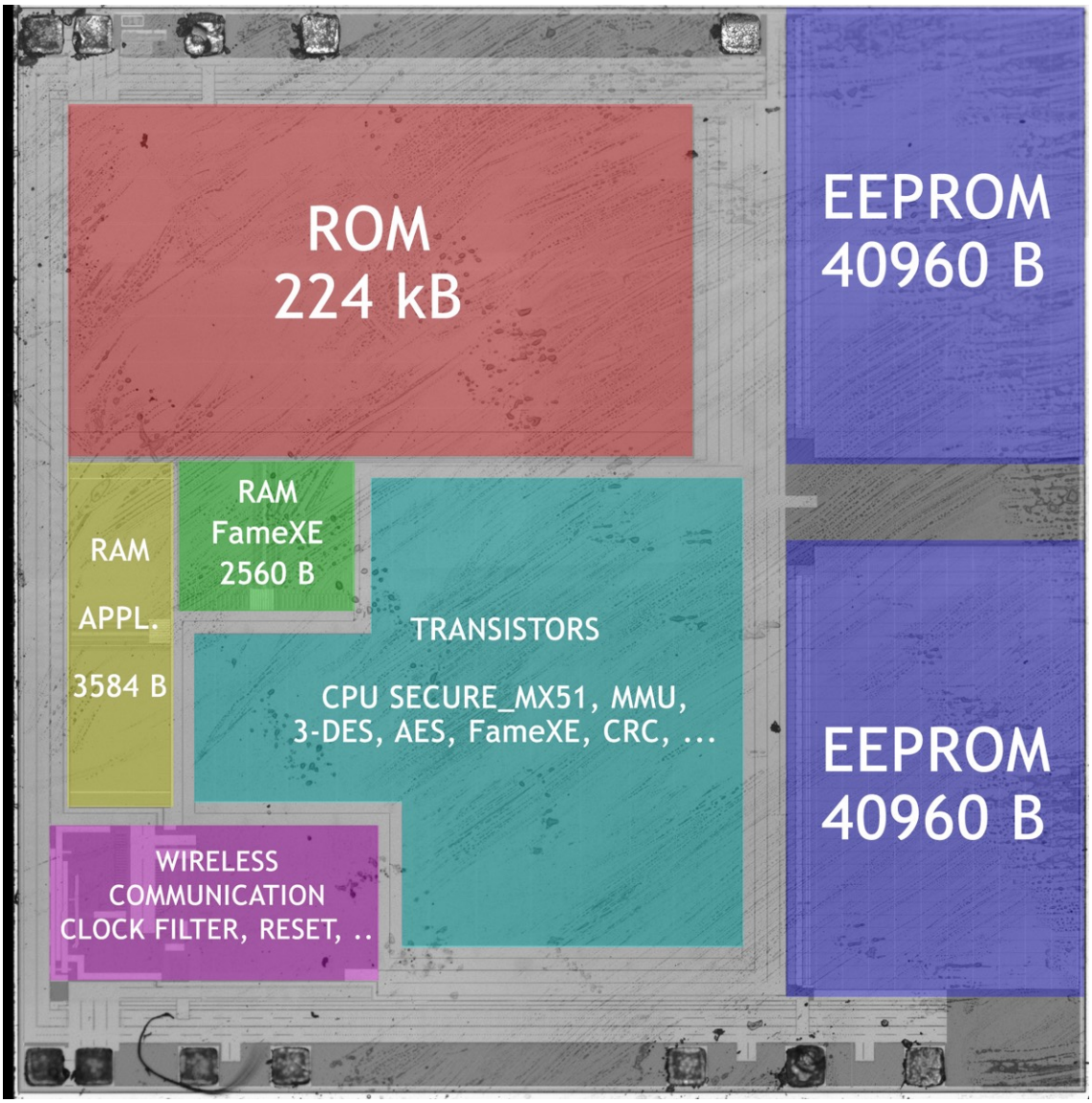


Figure 8.14: Segments of the chip NXP P5CD080 V0B (Source: author's work.)

interface also provides power for the chip). Another hint supporting this hypothesis is the fact that antenna inputs LA, LB both lead to this part.

We looked into the transistor filed and checked whether any pattern, segmentation or anything else helping the analysis could be read out just with the availability of the silicon layer. However, the field seems to be continuous, and so for gaining more information about the chip, it is necessary to scan and align images displaying the higher layers.

8.5.4 Future Work

The next step of the analysis is crystal clear, finishing the chip decomposition. We need to obtain all metal layers and the LIL layer. These layers have to be scanned, stitched and aligned in order to enable the further analysis. The metal layers should provide insight into the segmentation of the transistor field. The local connections realized in LIL, M1 and M2 layers should allow us to distinguish among components on the chip—CPU, co-processors, etc. It is known that the CPU works in several modes with restricted or unrestricted memory accesses—this can lead to another target in investigation. Last but not least, audit of the chip's support for running MIFARE Operating System should be performed as well.

Very interesting will be to see the layers above the ROM memory. We would like to see whether we can harvest further information about the ROM structure and its content based on the information obtained from the higher layers. We would like to investigate whether there is any observable border or splitting between boot ROM, test ROM and application ROM parts. The next point of interest will be the mechanism of disabling access to the test ROM after production and testing phases. Moreover, the data stored in ROM could be encrypted, the potential data encryption would be an another target.

Chapter 9

Software tools for Microscopic Analysis

Finally, after all the steps needed for obtaining the image data of the chips, experiments with data processing can be carried out. We will mainly focus on the transistor field, as there are elements that can be recognized and grouped. We would like to compare several approaches and see the results in the form of computation demands, quality of gained output, and also further applicability of the results. We cannot fully reconstruct the cells without interconnection layers; however, we can focus on the elements recognition to lay down the cornerstone for further work.

First of all, because of limited amount of data sets available to us, we have to rule out any form of learning algorithm. These approaches need big data sets for training and separate data sets for evaluation. Moreover, various forms of learning approaches would be very probably the first direction to go for most of the researchers. That is the second strong argument for us to test approaches without the presence of the learning. The algorithms should ideally work with various data sets, just with some fine-tuning of parameters, without the need to retrain again and again.

In our previous work, we worked mainly with data from optical microscopes. Very simple ICs, i.e., 4-NAND chip were scanned, and thus, we were successful also with the analysis of the images based on edge detection and color separation [100], [97]. This simplistic approach is not possible with SEM images of more complex chips, because these are coming in gray-scale only.

The top layer of the chip usually provides only a very minimal portion of the information—the overall preview, positions of the bond pads and rough segmentation of the chip. This segmentation can give us a clue about the placement of specific chip components, e.g., memories, transistors, communication interfaces. However, the top-level is very often obfuscated with a kind of mesh, especially when dealing with security-related chips. So, we will go directly down to the silicon, where the structures are in their bare form.

As it is known, the chips are designed with the use of software (EDA—Electronic Design Automation), that repeats the placement of small cells, does optimization of the placement with respect to the length of interconnection, etc. This fact is then used in the reverse engineering software that tries to recognize the repeated placements of similar elements and their interconnections. Another fact worth mentioning is that there have to be always power and ground lines available for the transistors, usually organized line by line or in

matrices. Last but not least, we are dealing with CMOS technology, so we can expect pMOS and nMOS transistors close to each other.

For aiding the reverse engineering, there exist commercial software[1] that claims to be capable of recognizing elements on the chip surface, it aids the cells recognition and interconnects tracing. Because it is strictly commercial, we were unable to test any of these software products. There exist some publicly available software packages, like degate.org[2] or pr0nsweeper[3]. However, it turned out that the projects were designed years ago and are not maintained (slightly more information about these publicly available software packages can be found in Chapter 7.5). For some use-cases with newer chips or for even a complete analysis of old chips (that means very simple chips), it is possible to use also general-purpose graphic editors, e.g., GIMP[4], Inkscape[5], Adobe Photoshop[6], Adobe Illustrator[7], to trace the connections and mark the chip parts by hand. This is for sure not suitable for the extent of the NXP P5CD080 V0B chip. We used gimp and Adobe Photoshop only for observing the big images.

Our aim is thus to detect similar elements across the chip, or at least in the transistor field. The further step would be to group these together into bigger similar blocks. This would help the researcher in location of particular logic cells. We would like to evaluate various approaches to this task, from simple template matching, through exploitation of image descriptors to shape or contour detection.

9.1 Setup

In order to achieve sustainable, maintainable, and extendable work, we will use very popular Python programming language with OpenCV library, a standard in image processing. We will test the solution on platforms Microsoft Windows and Linux in order to achieve portability. We have two dedicated machines for development and tests, each with a different operating system:

- Intel Core i5-4690 with integrated GPU, 32GB DDR3 RAM, 160 GB SSD SATA drive, MS Windows 10 Pro 64-bit (primary station):
  - Python 3.7.3 (numpy, scipy, matplotlib, imutils, guizero, statistics, PIL, argparse, statistics, multiprocessing, json, time, ...).
  - OpenCV 3.4.2.16 (used fro SIFT, SURF, ORB calculation); 4.1.0.25 (used for template matching and shape detection).
  - SQLite 3.21.
- Intel Core i5-4690 with integrated GPU, 16GB DDR3 RAM, 250 GB SSD SATA drive, Ubuntu 18.04.2 LTS 64-bit:
  - Python 3.6.8 (numpy, scipy, matplotlib, imutils, guizero, statistics, PIL, argparse, statistics, multiprocessing, json, time, ...).

[1] https://www.texplained.com/about-us/chipjuice-software/, https://www.chipworks.com
[2] https://degate.org
[3] https://github.com/JohnDMcMaster/pr0ntools/tree/master/capture/cf
[4] https://www.gimp.org/
[5] https://inkscape.org/
[6] https://www.adobe.com/products/photoshop.html
[7] https://www.adobe.com/products/illustrator.html

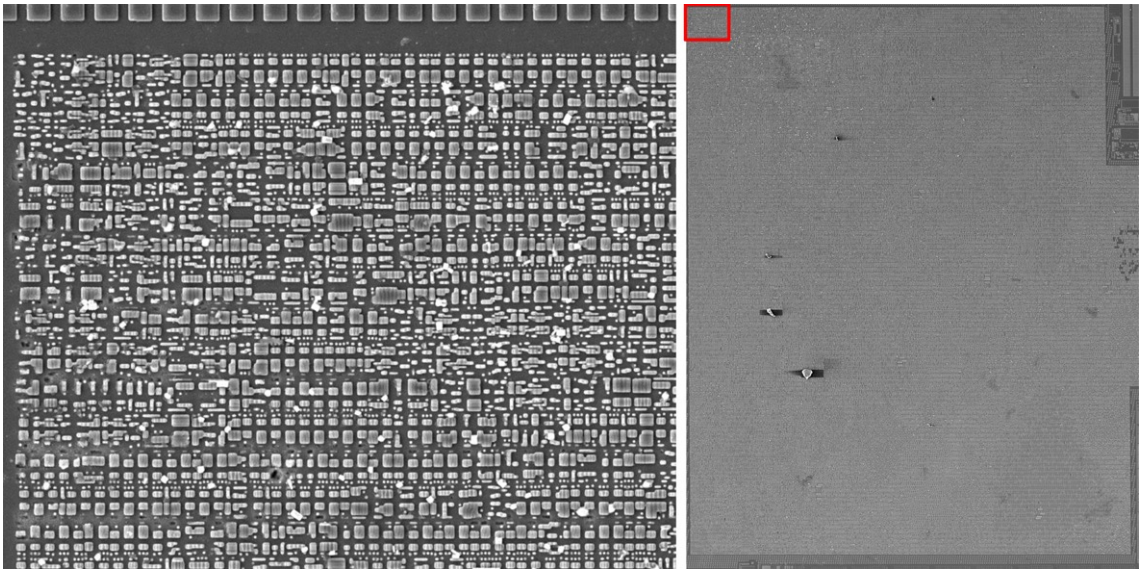


Figure 9.1: Model images used for experiments. Left: Transistor field cut-out taken from chip NXP P5CD080 V0B. This image was used for most of the development and tests. Right: Image with almost the whole transistor filed of the chip NXP P5CD080 V0B. This image was used for verification of developed methods and various measurements on bigger data. Note the red rectangle in top left corner; this highlight denotes exactly the part shown in the left part of this image. (Source: author's work.)

- OpenCV 3.4.2.16(used fro SIFT, SURF, ORB calculation); 4.1.0.25 (used for template matching and shape detection).
- SQLite 3.21.

Because of OpenCV limits, there is a restriction for single image size—an image must not exceed 2ˆ30 pixels, that is 1 073 741 824 pixels. This limit was exceeded with our stitched image that holds 40365×40612 pixels (1 639 303 380 pixels). Thus, we down-sampled the image with GIMP editor to 32000×32079 pixels (1 026 528 000 pixels), which fits into the limit. Image quality did not suffer from this operation.

Using the full image for development and testing purposes would be too impractical (computationally demanding and lengthy). For the purpose of faster computations and comparisons among various settings, we created a cut-out (1150×966) image from the transistor field, see Fig. 9.1. This model image was used across the examined methods described in the following sections.

9.2 Template Matching

Template matching approach is relatively straightforward. We have to determine a template (a subset of an image in our case), and this template is matched against the selfsame image to find similarities. To find all possible similarities, we take the template window, place it at coordinates 0, 0 (x, y) and take the first template from the image. This template is compared against the whole image itself. Then we move the template window with a certain step to the right to the next position and repeat the comparison part. As soon as we move the template window to the end of the first line, we move it with the set vertical step to

the next line and repeat the whole processing of a line. The whole image is scanned in this way.

There are three critical parameters to be tuned for each setup, influencing a number of matches and computation time:

- template size (including template shape—square vs. rectangle),
- template shift step,
- similarity threshold.

We performed several tests with template windows sizes 30×30, 40×40 and 50×50 (note, we always used square windows; the window size has to be determined according to the size of elements we want to pair). Each template size was run with template shift 10% (i.e., 5 pixels shift for 50-pixel wide window) and later with template shift 50% (i.e., 25 pixels shift for 50-pixel wide window). With each setup, we tested threshold settings from 0.75 to 1.0 with step 0.05. All relevant outputs are displayed in Fig. 9.2, Fig. 9.3, Fig. 9.4, Fig. 9.5, Fig. 9.6, Fig. 9.7. The tests were performed with use the model image (see the left part of Fig. 9.1) and TM_CCOEFF_NORMED mode in the function `cv2.matchTemplate(IMAGE, TEMPLATE, METHOD)` that is expressed in Equation 9.1, where letter I denotes image, letter T template, and letter R result:

$$R(x,y) = \frac{\sum_{x',y'}(T'(x',y') \cdot I'(x+x',y+y'))}{\sqrt{\sum_{x',y'} T'(x',y')^2 \cdot \sum_{x',y'} I'(x+x',y+y')^2}} \tag{9.1}$$

Our implementation was optimized in order to utilize all available CPU cores with Python multiprocessing `pool.map_async`. Originally, we calculated coordinates of all possible templates in the whole image, stored them in an array and then run the `pool.map_async` on this complete pool of tasks. However, it was not possible to efficiently monitor the progress of the computation. Printing out a message to standard output after processing of each template window while the CPU was fully utilized was too big overhead with a very negative impact on computation time—I/O operations are apparently very time-consuming. That is why we joined tasks into bunches, we pre-calculated coordinates of templates for a single line (in other words all templates with the same y coordinate) and run `pool.map_async` on this line. Then we printed out a message with information about the processed line and time duration of this single line computation. This approach imposed acceptable overhead on the computation and allowed us to monitor the progress of the whole computation—the wider the image (and generally also the bigger the image), the more negligible the overhead is.

After all, these computations can be sped up seriously by implementation with CUDA, as it is a perfectly suitable task for parallelization. The computation length can be one of the most significant drawbacks of this method. Running this kind of matching on the reference cut-out image was acceptable from the computation length perspective (varying from half a minute to 37 minutes, based on settings of the particular calculation). RAM consumption is, in this case, negligible, around 200 MB. This might not seem too much, but when running the same computation on a bigger image displaying substantial part of transistor field of the chip (see the right part of Fig. 9.1), the computation duration with 50 pixel template width and 50% shift gets to circa 55-60 days (duration calculated after computation of 3 rows out of 604). 12 GB RAM consumption is, in this case, totally

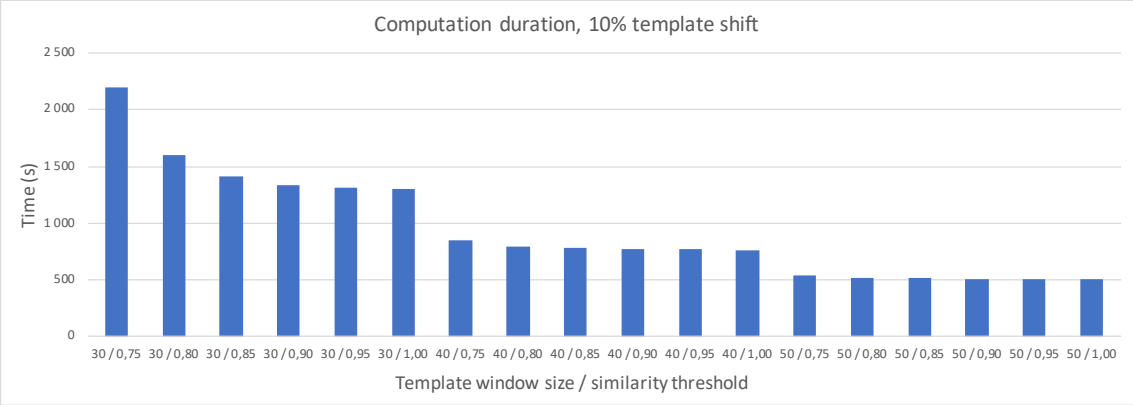


Figure 9.2: Template matching computation duration; 10% template window shift. (Source: author's work.)

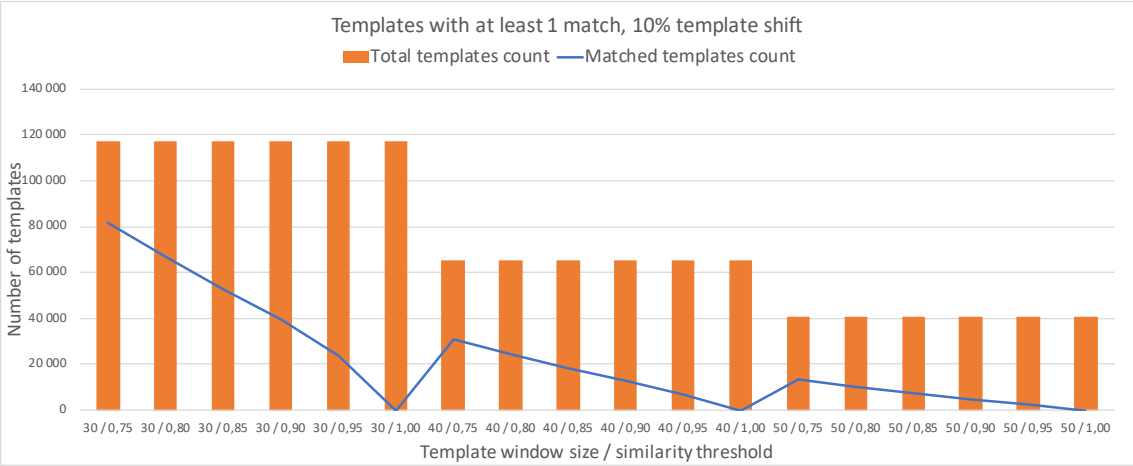


Figure 9.3: Total number of templates with highlighted portion of templates with a match; 10% template window shift. (Source: author's work.)

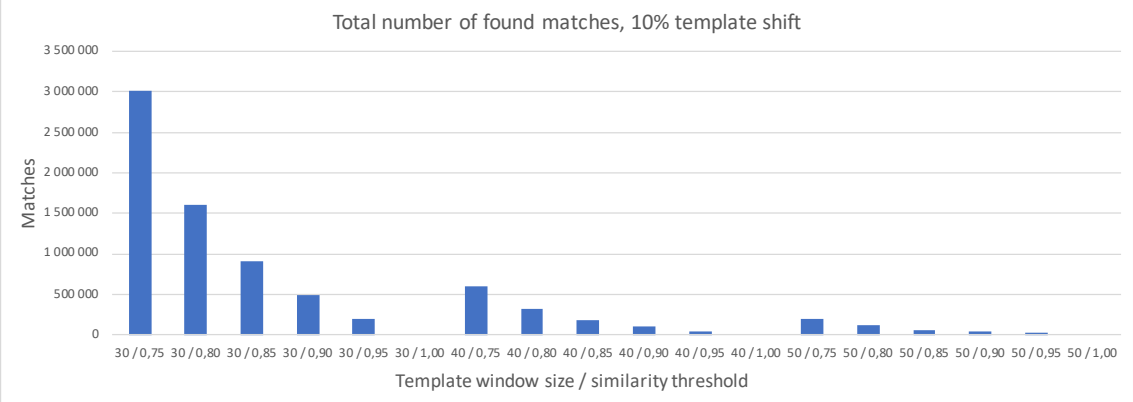


Figure 9.4: Total number of found matches; 10% template window shift. (Source: author's work.)

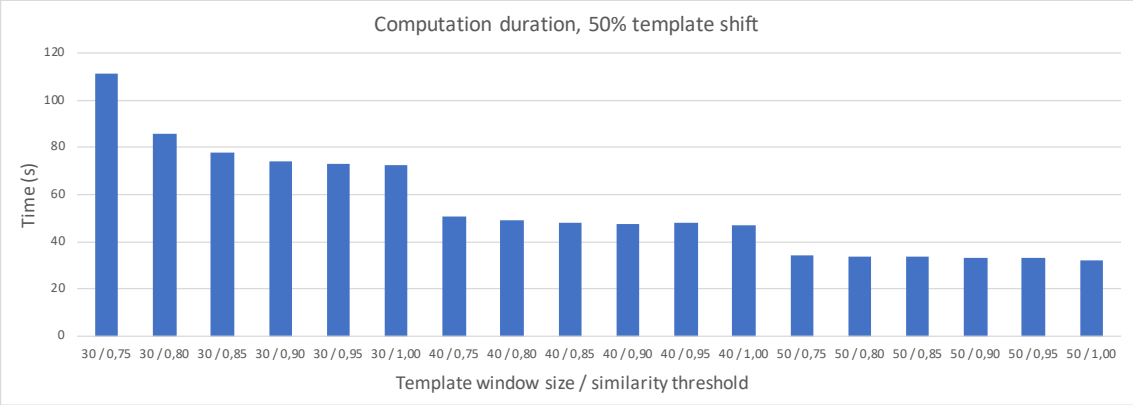


Figure 9.5: Template matching computation duration; 50% template window shift. (Source: author's work.)

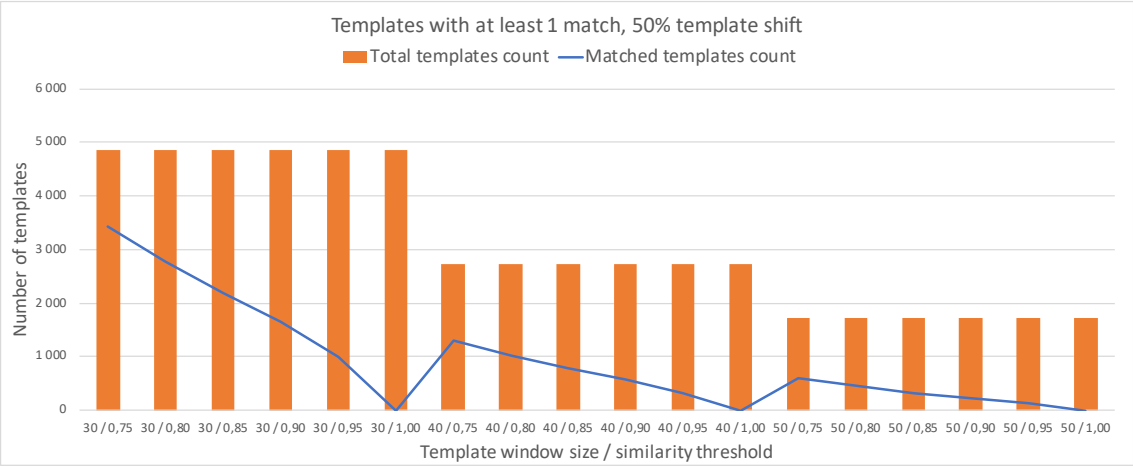


Figure 9.6: Total number of templates with highlighted portion of templates with a match; 50% template window shift. (Source: author's work.)

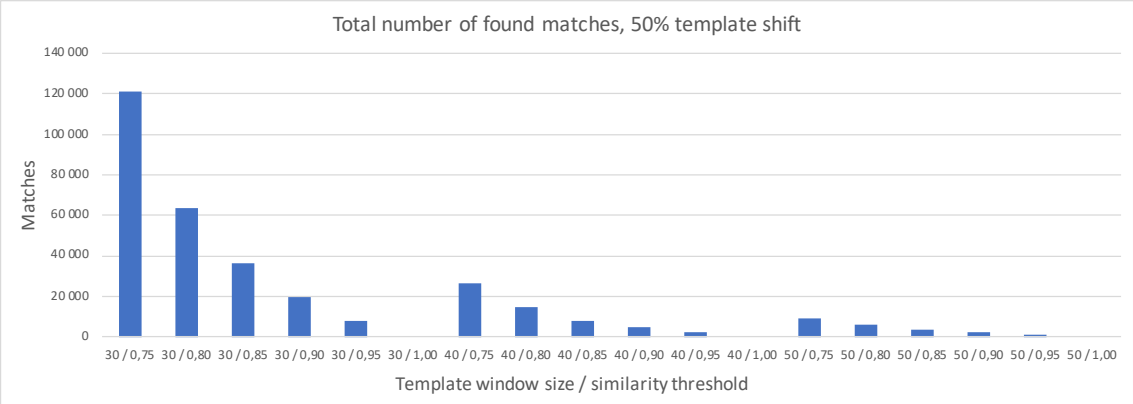


Figure 9.7: Total number of found matches; 50% template window shift. (Source: author's work.)

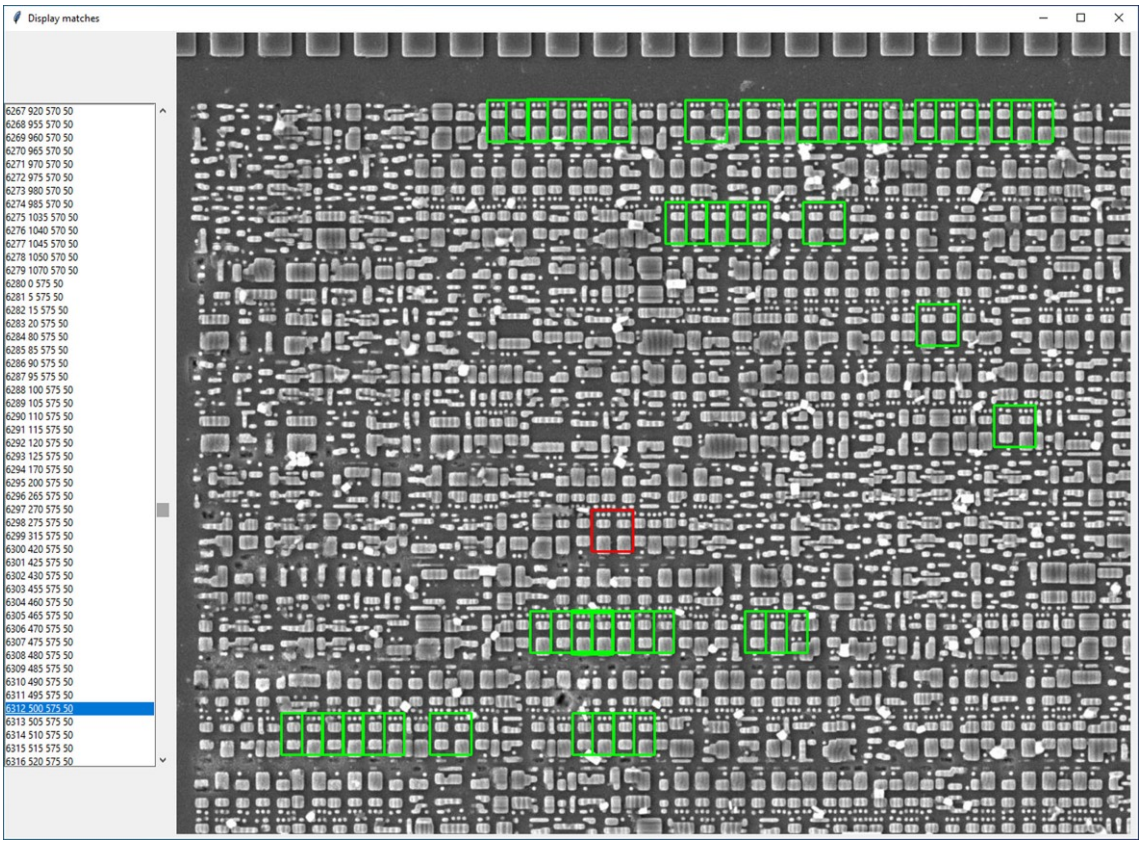


Figure 9.8: The application for browsing results from template matching. (Source: author's work.)

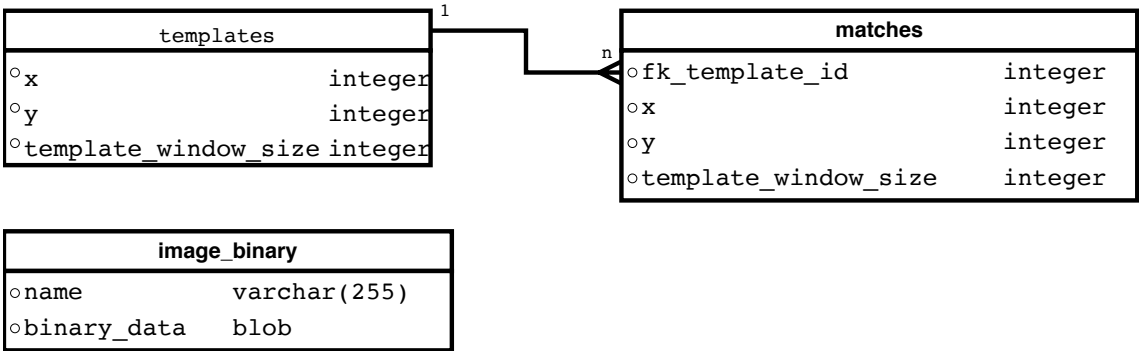


Figure 9.9: DB layout of the exchange format used for the template matching experiments. (Source: author's work.)

irrelevant. Let alone running the same computation on the whole 32000×32079 pixel image without the employment of CUDA would be very impractical.

Applicability of results gained with the use of template matching is suitable for further manual analysis, where the researcher is supported with marked similar parts of the chip. We created a simple matches browser for better evaluation of template matching results, see Fig. 9.8. A similar application for evaluation of the results is used also in the following sections. We also created a simple exchange file format with use of SQLite[8] (chosen due to portability) in order to be able to compute the template matching once and then to be able to filter results without the need of the re-computation. This database file holds image data, all templates that have at least one match and all matches mapped to the templates (see Fig. 9.9). So, it is possible to just exchange these db files with all needed information. Interpretation of the results or further filtering of these is then up to the researcher.

9.2.1 Template Matching Summary

Use of template matching approach is convenient especially for its simplicity, depending on the particular use-case, there are very few variables (size and shape of the template, template shift, similarity threshold and mode of matching) that can be easily tuned in order to get results suitable just for the use-case. On the other hand, the output is usable rather for support of manual research activities, as there is no other output than coordinates of the matched windows.

There are several things to be dealt with when employing this template matching approach. First of all, there are many mutual matches that can be merged together—i.e., there are 50 similar parts in an image, when moving the template window to each of them, the others are matched. Second, depending on the set threshold, the are several matches around the really matched part of the image. Let us assume 50 pixels width template window and threshold 0.7 that means 70% of the images have to be similar. Thus, a similar area is not matched just once with the highest possible similarity level, but all slightly moved windows around this area that are still within the 70% threshold are matched as well. This can be either influenced by setting higher threshold value or setting up a filter that always takes only one match from the area and ignores the others.

From the results presented in this chapter, we do not recommend using 10% template shift unless it is truly demanded for spotting very irregular elements. Increased computation

[8] https://www.sqlite.org

demands simply do not payback. 50% template shift will be enough for many use-cases with regular chip structures and as shown in Fig. 9.3, Fig. 9.4, Fig. 9.6, Fig. 9.7; the amount of matching results for further manual processing is anyway in thousands or tens of thousands. Subjectively, the matching results are of good quality, and similar parts are marked correctly. For each data set, it is necessary to find the correct template size and the threshold (i.e., because of possible noise in the image that can influence the matching algorithm). The template size can be determined with the use of a conventional graphics editor, examining the input image data and size of elements we want to pair. Chips are usually designed in the regular orthogonal manner and so finding correct size fitting the imaginary rows and columns is not difficult.

9.3 Feature Descriptors

The next idea in image processing or rather in similarities recognition was to use some kind of image descriptors and to investigate whether these can be suitable for cross-matching similarities within a single SEM image depicting microchip silicon layer.

First, we wanted to examine feature-descriptors Scale Invariant Feature Transform (SIFT), Speed up Robust Feature (SURF) and Oriented FAST and Rotated BRIEF (ORB, where FAST means Features from Accelerated Segment Test and BRIEF means Binary Robust Independent Elementary Features). These methods for features detection are known in image matching and computer vision and are also available in openCV library. [155], [2], [81], [19]

Our idea was to investigate particular key-points and descriptors that are ordinarily used for matching a template image against an another image containing the same object in it; usually in a different scene (these methods are invariant to various transformations). The matching in this standard use case is based on finding key-points in both of the images and trying to correlate them mutually. One could propose to follow a similar process as presented in template matching—taking one template after another from the original image and compare against the rest of the image. In such use case, the template matching could be used directly instead. Our aim was to investigate detected features (key-points and descriptors), within a single image and see whether there are any similarities among them. Especially among the ones located in elements that are visually similar.

In the beginning, we ruled out ORB method—this method is very fast and was actually the only one able to calculate key-points of the whole silicon layer image (32000×32079 pixels) using our testing hardware; moreover, the computation took only 25 seconds. That was a significant difference compared to the two other methods that both ended with an error caused by not enough RAM. The only drawback that predetermines the ORB method for the original purpose of two images mutual matching is the number of key-points it selects. The ORB method chooses only N key-points [81] that are the best for the image characterization, that is why it is so fast compared to the other methods. However, this is not usable for our purpose, where we need ideally a very dense net of key-points. From a comparison of extracted key-points from the original image without preprocessing (see in Fig. 9.10; Although it is not so clear from the provided visualization, the densest net of key-points is calculated by the SIFT method, followed by SURF calculation. To be concrete, the results for our model image are as follows:

- SIFT—25200 key-points.
- SURF—16407 key-points.

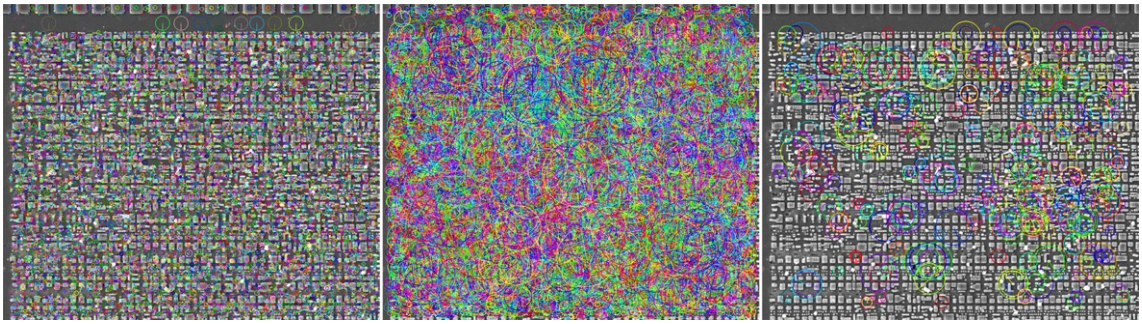


Figure 9.10: Visualization of computed key-points. Left: SIFT. Middle: SURF. Right: ORB. (Source: author's work.)

- ORB—500 key-points.

The algorithms calculate a set of descriptors for each of the key-points. Instead of the standard:

`radiusMatch(queryDescriptors, trainDescriptors, maxDistance)`,

we called the function in this way, comparing the key-points against themselves:

`radiusMatch(queryDescriptors, queryDescriptors, maxDistance)`.

The maxDistance was determined experimentally, for SIFT calculations, the optimal value was around 150 and for SURF 0.15 (for more details see Table 9.1).

Initially, we played also with the classification of the key-points just based on their size and response with neglecting the descriptors. Based on the results, we do not recommend using this method as the first step of the classification. However, it provides useful information that might help in later stages of the classification for fine-tuning of the results gained from the descriptors matching stage. The same as with the descriptors matching, there has to be a tolerance set by a threshold (for both the size and the response). For example, for the SIFT method, we experimentally determined 0.008 threshold for the key-points response and 0.6 threshold for the key-points size. All key-points within the thresholds fall into the same group. During these experiments, we investigated the influence of the angle of the key-points on the classification. The angle is always misleading and should be neglected.

Another attempt of classification was to use clustering on the key-points (based on their location) as the first step of the classification with neglecting the descriptors. We experimented with DBSCAN (Density-Based Spatial Clustering of Applications with Noise) and with K-Means clustering. This approach did not lead to any usable result, because the key-points net is too dense, and thus we got either too many or too little clusters. Nevertheless, this method may be usable in the fine-tuning of the results classified by the descriptors.

Although we also experimented with FLANN based matcher (Fast Library for Approximate Nearest Neighbors), we stayed with the standard Brute-Force matcher with default settings at the end as we did not find any better settings of the FLANN matcher.

Table 9.1: The total number of key-points in the testing image, and the number of key-points matched with another key-point. Tests were performed on the testing data described at the beginning of this chapter; the size of the testing image: 1150×966.

	Total key-points detected	Key-points with at least 1 match with a different key-point
SIFT	25200	7363 (max. vector distance 150)
SURF	16407	4277 (max. vector distance 0,15)

Table 9.2: Comparison of computation length and RAM demands—SIFT and SURF are measured including the classification part. SIFT and SURF have the used max. distance between vectors in brackets. Tests were performed on cut-out of a transistor field from the MIFARE Classis chip; the size of the testing image: 1049×1203.

	Template Matching 50×50 template 10 % template shift	Template Matching 50×50 template 50 % template shift	SIFT (150)	SURF (0.15)
Time	1864 s	143 s	13 s	7 s
RAM	200 MB	200 MB	2200 MB	1800 MB

We experimented with placing the mask over the image to compute key-points only in the areas where we expected the elements and also with a complement of that mask, thus calculating key-point only in the spaces between the elements. However, any restriction of the key-point base did not lead to better results. In fact, SIFT key-points calculated in between the particular elements can map their respective positions, and so it will help in later stages when we plan to perform clustering into bigger logic blocks. Thus, the recommendation is to calculate the key-points on the whole image in order to get as dense key-point net as possible.

For easier evaluation of the results, we decided to create an application for visualization of the matched key-points. This application allows for changing parameters from the command line as follows:

```
python feature-radius.py -image test.jpg -threshold 150 -method sift.
```

We also performed experiments regarding the performance of the methods, but because these experiments are already published in various papers ([155], [2], [81]), let us only confirm the known fact that ORB is the fastest out of the tested methods; followed by SURF; and the slowest one is SIFT.

9.3.1 Feature Descriptors Summary

Subjectively, the best matching was performed with SIFT. What we liked about the descriptors was the fact that this method is capable of marking rotated elements. Moreover, it was able to cope with various artifacts and still match the point with a similar one. On the other hand, it is clear that this crude mutual matching is not absolutely sufficient as there

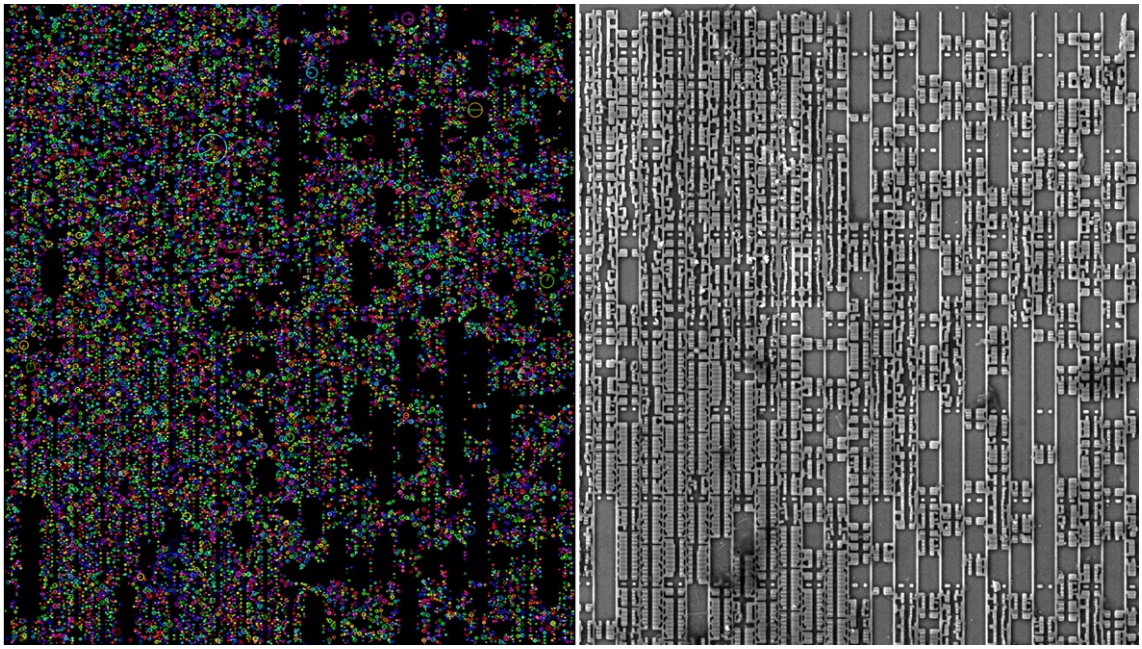


Figure 9.11: Visualization of computed key-points (SIFT) on a different data set—cut-out of a transistor field from MIFARE Classic chip; the size 1049×1203. Left: Key-points displayed on a black background. Right: The input image. (Source: author's work.)

are still many false positives and also many points are not matched at all. Nevertheless, it is clear that this approach is definitely worth further examination.

The disadvantage is basically the same as with the simple template matching method presented in the previous section—so far, we have just single points and highlighted areas without deeper semantics. On the other hand, computation length is incomparably shorter than with the template matching (especially when comparing against the 10% template shift). For instance, with different testing data from the MIFARE Classic chip—cut-out of a transistor field; the size 1049×1203 (key-points distribution is displayed in Fig. 9.11). Comparison of the results is presented in Table 9.2.

The future work within this approach will consist of the following steps:

- Investigate the possibility of using size and response of the key-points for the next step of classification.
- Investigate the possibility of using clustering based on the location of the key-points for the next step of classification.
- Investigate advanced image preprocessing methods—increasing contrast of the source image in order to increase key-points quality and quantity; conversion of the source image to black and white with precise preservation of the object shapes.
- Cluster matched key-points from similar areas.
- Test other descriptors, e.g., LBP (Local Binary Patterns), GLCM (Gray-Level Co-Occurrence Matrix), Gabor wavelets, fractal analysis.

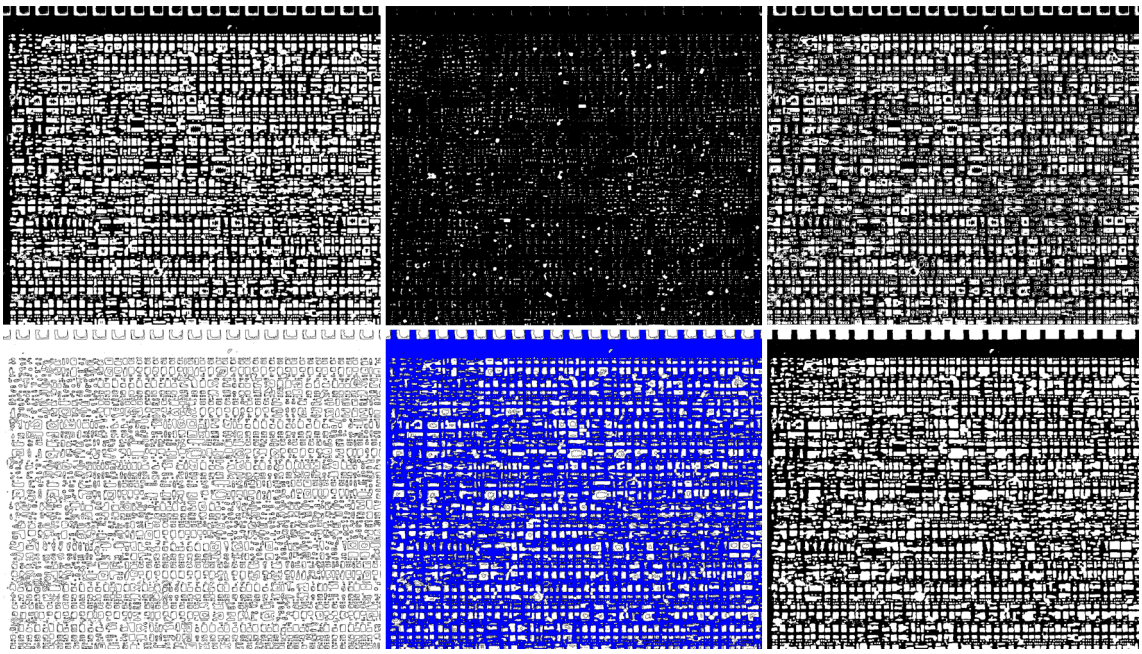


Figure 9.12: Preprocessing of the image. Top left: Step 1—result of the first thresholding. Top middle: Step 2—result of the second thresholding. Top right: Step 3—bitwise XOR of the thresholded images. Bottom left: Step 4—contours found in the image from the previous step. Bottom middle: Step 5—background was flooded with a different color. Bottom right: Step 6—complement of the background was filled with the white color and background color was changed to black. (Source: author's work.)

9.4 Shape Matching

The last approach we wanted to examine in the scope of this thesis was to extract shapes of the objects out of the source image. OpenCV library provides several methods usable for this purpose. After the first experiments, we realized that these methods would not work with our data set just out-of-the-box. The first attempts of detecting shapes directly in the original testing image led to many imprecisely detected objects with variable shapes (also among visually similar objects) and objects with holes inside (area marked with another shape as object in object). Therefore, we experimented with image preprocessing in order to get the whole objects correctly marked. There were many trial and error attempts along the way. Let us present an approach that worked well at least with one of the testing data sets.

First, we calculated two thresholds of the input image:
`_, grey_img = cv2.threshold(img, 110, 255, cv2.THRESH_BINARY)`, and
`_, white_img = cv2.threshold(img, 220, 255, cv2.THRESH_BINARY)`. Then, we applied bitwise exclusive OR on the two thresholded images. This helped us to filter out artifacts present in the source image and thus separate the objects. Further, we searched for contours (with approximation mode `cv2.CHAIN_APPROX_TC89_L1`) in the XORed image. At this point, it was possible to perform the correct background separation. The background was flooded with the blue color because we needed to be able to distinguish among background (blue), object contours (black), and inner parts of objects (white). Finally, we could filter out the contours in order to have clean objects (white), and background (blue). The background was turned from blue to black to respect black and white mode standard. The whole preprocessing procedure is displayed step-by-step in Fig. 9.12.

Subsequently, we decided to apply the Watershed algorithm for the pre-processed image segmentation; into separate objects. This algorithm is available in the OpenCV library (`cv::watershed`). The result of the image segmentation is presented in Fig. 9.13. Please note that the preprocessing procedure is essential for successful image segmentation.

Approximation of the shapes is essential in the whole process. Our final approximation is expressed with circles in the image (see Fig. 9.13). Each object is represented by a circle with a radius calculated as the minimal enclosing circle for the object. This allows us to simplify the object representation for the consequent classification. Moreover, we introduced even more tolerance with truncation of the radius value from float to integer.

A simple classification was based on the radius value, all objects with the same radius belong to the same class. To be completely correct, we merged all odd classes with their even predecessors—thus we got only half of the classes, each with more elements. Although it is obvious that there will be the need for further classification steps in order to distinguish among all the shapes of similarly-sized objects, this first classification step based on the presented approximation provides a very good promise for the further research (for illustration, see Fig. 9.14).

9.4.1 Shape Matching Summary

We were able to segment the input image and perform classification based on approximation of the object shapes with relatively good results—very good classification of the small and big objects; the mid-sized objects still need further step(s) in classification.

On the other hand, it has to be noted that this approach failed completely with our second data set (this data set is displayed in the right part of Fig. 9.11). It was not possible

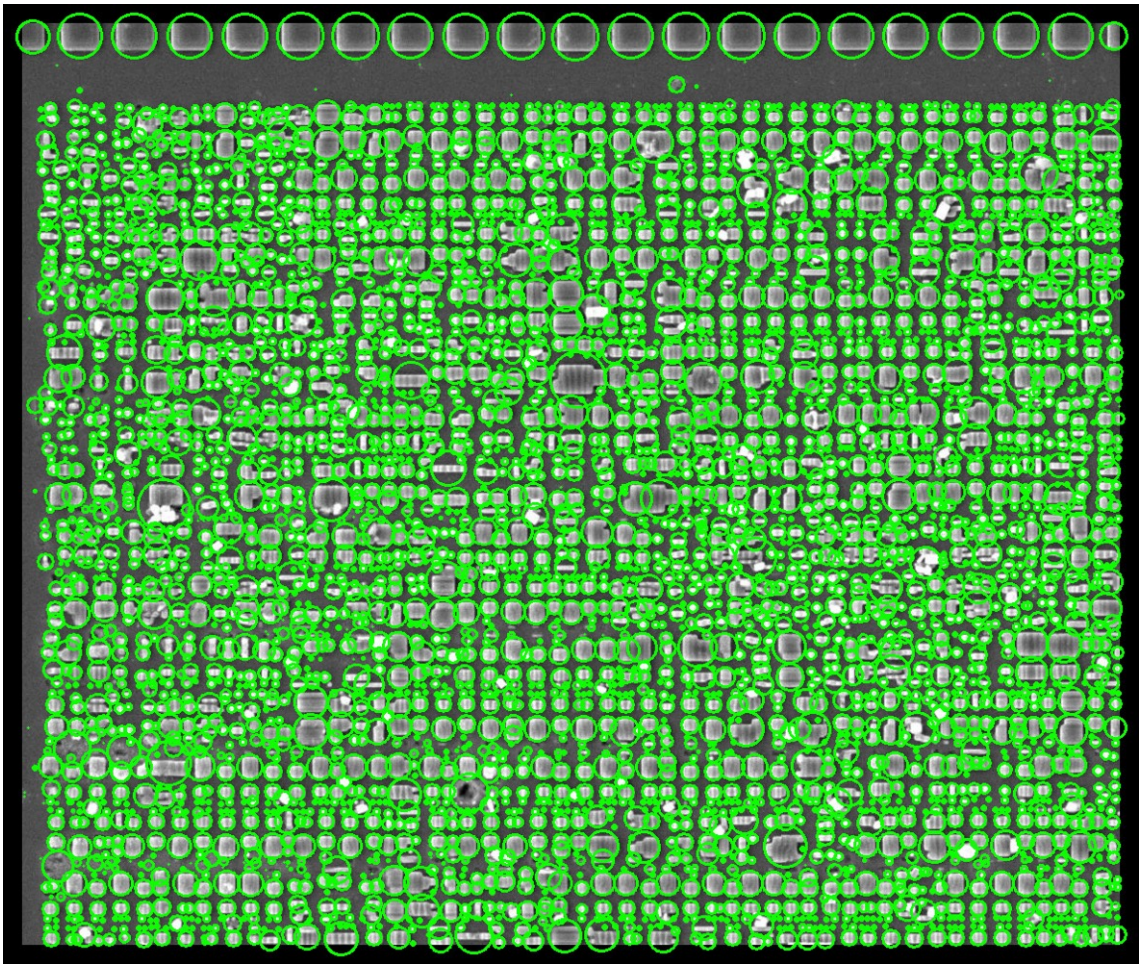


Figure 9.13: The source image was successfully segmented into separate objects. (Source: author's work.)

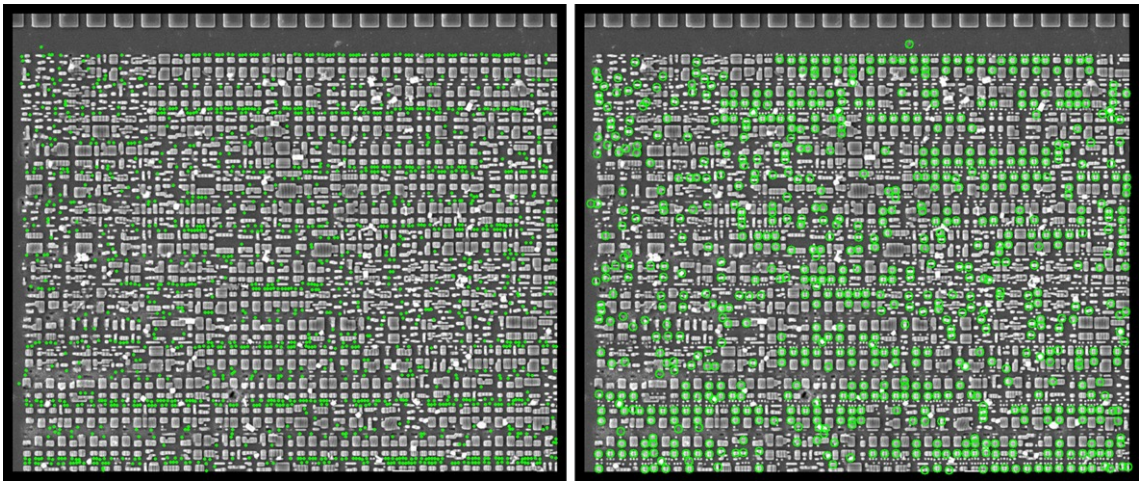


Figure 9.14: Simple classification of the objects detected in the image. Left: Very good results were achieved with the smallest objects (and also with the biggest objects—not displayed). Right: Classification of the mid-sized objects will need further classification steps. (Source: author's work.)

to segment the image in the way we presented. Most probably due to very different production technology—the elements are interconnected, and thus segmentation is not possible in the presented way.

We believe that this approach will be usable with SEM images, where the separate elements represent the transistor field.

The future work within this approach will consist of the following steps:

- Investigate the possibility of various image preprocessing.
- Investigate the possibility of using Hu Moments for shape matching.
- Investigate approximation of the object with the use of squares (or rectangles in general).
- Introduce advanced classification; employing more features of the object in order to distinguish among similar-sized objects.

9.5 Summary and Future Work

Let us present a final summary of the examined techniques—the performance and RAM demands are presented in Table 9.3. It can be concluded out of the results, that the classic template matching is not convenient neither from performance perspective nor from the gained results. The template matching can be relatively precise with high threshold settings; however, we are missing artifact tolerance then. Let alone the detection of rotated elements.

Image descriptors perform very well. Nevertheless, quality of the output still needs to be moved forward—namely, better classification of the key-points and descriptors, and potential future clustering of similar areas. We would like to continue with experiments and further examination of this approach. The same as with the shape-based matching, where

Table 9.3: Comparison of computation length and RAM demands—SIFT, SURF and Shape Matching are measured including the classification part. SIFT and SURF have the used max. distance between vectors in brackets. Tests were performed on the testing data described at the beginning of this chapter; the size of the testing image: 1150×966.

	Template Matching 50x50 template; 50% shift	Template Matching 50x50 template; 10% shift	SIFT (150)	SURF (0.15)	Shape Matching
Time	534 s	34 s	14 s	4 s	16 s
RAM	200 MB	200 MB	2400 MB	1000 MB	200 MB

we showed that it was possible to segment the image and prepare the objects for advanced classification. We will also try to combine these two approaches.

One pre-condition for success in future research will be to collect more data sets for development and testing, and thus to deprocess and scan more chips.

Chapter 10

Conclusion

This thesis primarily tackles microscopic analysis of chips with respect to their security. Consequently, tasks affiliated to this topic, e.g., decapsulation, deprocessing, image acquisition and processing, had to be managed as well. The research topic was more applied research than a theoretical one. Despite being at an IT faculty, we had to dive deeply into chemistry and physics in order to be able to proceed with the experiments. Moreover, due to the fact that we were performing all steps in a low-cost regime, we had to deal with all stages primarily on our own. Last but not least, we had to perform the experiments and research basically barehanded compared to the top-notch laboratories that are specialized in this sphere.

Because the efforts in the academic sphere regarding decapsulation and deprocessing were rather scattered and disconnected from each other, we prepared a detailed overview of the decapsulation and deprocessing techniques. Moreover, all processes that were mapped in detail were also verified several times during the experiments in the chemical laboratory. We were able to polish details of the procedures based on our practical experience. During the experiments, we significantly improved the process of obtaining the chips from thermoplastic compounds resulting in approximately 10× speedup. These compounds are used for the production of the plastic cards holding the smartcard chips. With respect to the number of smartcards used all over the globe, we expect this improvement to be practically usable on a daily basis.

The decapsulation techniques that were presented in this thesis are feasible in low-cost, and thus, anybody should be able to carry out this part of the analysis with the use of the provided detailed description. We must also add that observing safety rules is a must, and we do not advise to perform any experiments without proper laboratory equipment and without proper training.

The deprocessing techniques were prepared mainly for the low-cost scenario we worked in. However, as we were getting to the more recent chips, it was clear that obtaining good quality results at a low-cost (without advanced equipment) is based rather on luck than on reproducible processes. Therefore, we also mapped possibilities that are used in professional laboratories. Nevertheless, we cannot consider this low-cost anymore because of the costs of the needed equipment. The older chips can be deprocessed in a chemical laboratory with acceptable effort (after some practice) and still in low-cost—using inexpensive chemicals and an ordinary optical microscope for inspections. Nevertheless, when getting to more recent chips with technological node below 200 nm, proper cross-section analysis has to be done. Afterward, the precise work based on the results of the cross-section is needed—a cross-section analysis is also covered in the thesis. Such work cannot

be performed just with beakers and some chemicals. There is the need for an advanced plasmatic etching station and ideally also a parallel polishing station. Furthermore, optical microscopes are not sufficient for observations of such devices; thus SEM, CLSM, or similar microscopy is required. The detailed description of the procedures was provided to shorten the learning curve of other researchers. The presented procedures were practically verified with experiments—we successfully decapsulated and deprocessed several chips, e.g., simple RFID tags, MIFARE Ultralight C, MIFARE Classic 1 kB; partially also MIFARE DESfire EV1 and the SmartMX chip NXP P5CD080 V0B.

Further, we managed the acquisition process of large dies with SEM microscopy and subsequent post-processing of this data. A comparison of publicly available tools for processing of this type of data is presented, we focused mainly on the stitching of the SEM microscopy outcomes—separated tiles representing the overall chip layer image. We compared the tools with respect to computing time, RAM demands, and the quality of output perspectives.

Finally, we tackled the actively used chip from SmartMX family—NXP P5CD080 V0B, that can be found in the Czech biometric passport implementation. As we found out later, the same chip was also used in the German electronic ID card (Personalausweis) implementation. Compared to the other smartcard chips, this chip was significantly larger in area. We were able to perform decapsulation of 15 samples followed by a detailed cross-section analysis, which revealed the chip's structure, used materials, thickness of layers and vias, and the layout of the chip. Based on this analysis, we were able to create the exact recipe for the chip deprocessing. However, performing the whole process was not possible due to a missing ultra-high frequency plasmatic etcher. It was shown that barrier layers protecting each metal layer are removable with this procedure. Nevertheless, we were able to obtain at least two layers—the silicon level layer and the M5 metal layer. Consequent analysis of this limited data set allowed us to collect a decent portion of information though. The whole analysis process and its results are presented in this doctoral thesis.

An evaluation of various potentially viable methods for (semi-)automatic logic element recognition is also presented. We focused on methods other than the employment of machine-learning. We decided to explore this approach because the common first choice would be to employ learning mechanisms such as, for example, neural networks. The second reason for taking this direction was the limited testing data set availability. It was shown that there were non-learning-based methods usable for aiding microscopic analysis—particularly, image segmentation and object classification suitable for silicon layer image processing. We also outlined the future plan for further research in this field.

An overview of attack classes is given in the thesis, followed by a discussion regarding particular attacks and proposals of possible countermeasures. These countermeasures are supposed to hinder primarily the microscopic analysis. We presented several scenarios based on recently emerging 3D integration, battery-backed memory protection, active tamper detection system, FPGA employment with memory protection mechanisms to shield the bitstream, etc. Finally, we also discussed a possible scenario for the next generation of e-Passports—when the RFID passport chips will be implanted into human bodies.

10.1 Future Work

We would like to establish strong partnerships with other faculties and facilities at Brno University of Technology in order to perform the chemical-related parts and scanning part of the microscopic analysis in professional environments. We have created a very good starting

point for this cooperation with the detailed description of the whole process. Furthermore, we would like to focus purely on understanding the data and creating software aiding reverse engineering of the chips.

Based on the thesis outcomes, there are several targets emerging for future research work. First of all, we would like to finalize the complete decomposition of the NXP P5CD080 V0B chip. Afterward, the missing layers can be scanned (LIL, M1, M2, M3, M4), stitched, and aligned to provide the complete chip structure representation.

Deeper investigation can begin, for example, with a proper investigation of the ROM memory structure. We would like to see whether there is any observable border or splitting between boot ROM, test ROM, and application ROM parts. The test ROM should not be accessible after delivery to the customer; however, it can definitely provide a lot of valuable information to the analysts. The next point of interest will be the mechanism of disabling access to the test ROM after production and testing phases. Another target will be reading the ROM content and examination of the obtained data. The data could be encrypted; the potential data encryption would be another target.

The next focus will be directed towards the transistor field. With the information from LIL, M1, and M2 layers, we should be able to distinguish separate blocks (based on the locality of the interconnections). This should allow us to distinguish among components on the chip—CPU, co-processors, etc. It is known that the CPU works in several modes with restricted or unrestricted memory accesses; this can lead to another target in the investigation. Last but not least, the chip supports the running of MIFARE Operating System. This deserves a proper investigation as well.

To conclude, there are numerous other microchips worth investigating, which is the overall motivation for our future work.

Bibliography

[1] Abraham, D. G.; Dolan, G. M.; Double, G. P.; et al.: Transaction Security System. vol. 30. Feb 1991: pp. 206–229. doi:10.1147/sj.302.0206.

[2] Ahmed, M.; Shaukat, A.; Akram, M. U.: Comparative analysis of texture descriptors for classification. In *2016 IEEE International Conference on Imaging Systems and Techniques (IST)*. Oct 2016. pp. 24–29. doi:10.1109/IST.2016.7738192.

[3] Ali, I.; Khir, M. H. M.; Baharudin, Z.; et al.: CMOS-MEMS multiple resonant vibration energy harvester for wireless sensor network. In *2015 IEEE Regional Symposium on Micro and Nanoelectronics (RSM)*. Aug 2015. pp. 1–4. doi:10.1109/RSM.2015.7354963.

[4] Ambrose, J.: *Power Analysis Side Channel Attacks: The Processor Design-level Context*. PhD. Thesis. UNSW Sydney. 2018.

[5] Anthony, S.: Zoom into a computer chip: Watch this video to fully appreciate just how magical modern microchips are. Online. 2014. [accessed 03-Feb-2019]. Retrieved from: https://www.extremetech.com/extreme/191996-zoom-into-a-computer-chip-watch-this-video-to-fully-appreciate-just-how-magical-modern-microchips-are

[6] ASM International: *ISTFA 2018: Proceedings from the 44th International Symposium for Testing and Failure Analysis*. A S M International. 2018. ISBN 9781627080996.

[7] ASM International and Electronic Device Failure Analysis Society: *ISTFA 2016: Conference Proceedings from the 42nd International Symposium for Testing and Failure Analysis : November 6-10, 2016, Fprt Worth Convention Center, Fort Worth, Texas, USA*. ASM International. 2016. ISBN 9781627081351.

[8] Atmel Smart Card ICs and Hitachi Europe Ltd. and Infineon Technologies AG and Philips Semiconductors: *Smartcard IC Platform Protection Profile*. Jul 2001. version 1.0.
Retrieved from:
https://www.commoncriteriaportal.org/files/ppfiles/ssvgpp01.pdf

[9] Azar, K. Z.; Kamali, H. M.; Homayoun, H.; et al.: SMT Attack: Next Generation Attack on Obfuscated Circuits with Capabilities and Performance Beyond the SAT Attacks. *IACR Transactions on Cryptographic Hardware and Embedded Systems*. vol. 2019, no. 1. 2019: pp. 97–122. ISSN 2569-2925. doi:10.13154/tches.v2019.i1.97-122.

[10] Bajura, M.; Boverman, G.; Tan, J.; et al.: Imaging Integrated Circuits with X-ray Microscopy. In *36th GOMACTech Conference*. Apr 2011.

[11] Barr, M.: *Programming Embedded Systems in C and C++*. Sebastopol, CA, USA: O'Reilly & Associates, Inc.. first edition. 1998. ISBN 1565923545.

[12] Bi, Y.; Shamsi, K.; Yuan, J.-S.; et al.: Emerging Technology-Based Design of Primitives for Hardware Security. *J. Emerg. Technol. Comput. Syst.*. vol. 13, no. 1. Apr 2016: pp. 3:1–3:19. ISSN 1550-4832. doi:10.1145/2816818.
Retrieved from: http://doi.acm.org/10.1145/2816818

[13] Blackwell, G. R.: *The Electronic Packaging Handbook*. CRC Press. 1999. ISBN 978-0849385919. 640 p.

[14] Bogari, E. A.; Zavarsky, P.; Lindskog, D.; et al.: An investigative analysis of the security weaknesses in the evolution of RFID enabled passport. *International Journal of Internet Technology and Secured Transactions*. vol. 4, no. 4. 2012: pp. 290–311. doi:10.1504/IJITST.2012.054060.
Retrieved from:
https://www.inderscienceonline.com/doi/abs/10.1504/IJITST.2012.054060

[15] Boit, C.; Schlangen, R.; Glowacki, A.; et al.: Physical IC debug - Backside approach and nanoscale challenge. *Advances in Radio Science - Kleinheubacher Berichte*. vol. 6. May 2008. doi:10.5194/ars-6-265-2008.

[16] Cai, F.; Bai, G.; Liu, H.; et al.: Optical fault injection attacks for flash memory of smartcards. In *2016 6th International Conference on Electronics Information and Emergency Communication (ICEIEC)*. Jun 2016. pp. 46–50. doi:10.1109/ICEIEC.2016.7589684.

[17] Carmo, J. P.; Rocha, R. P.; Silva, A. F.; et al.: Integrated thin-film rechargeable battery in a thermoelectric scavenging microsystem. In *2009 International Conference on Power Engineering, Energy and Electrical Drives*. Mar 2009. ISSN 2155-5516. pp. 359–362. doi:10.1109/POWERENG.2009.4915179.

[18] Carvalho, C.; Paulino, N.: CMOS Indoor Light Energy Harvesting System for Wireless Sensing Applications: An Overview. In *Technological Innovation for Cyber-Physical Systems*, edited by L. M. Camarinha-Matos; A. J. Falcão; N. Vafaei; S. Najdi. Cham: Springer International Publishing. 2016. ISBN 978-3-319-31165-4. pp. 178–194.

[19] Cavalin, P.; Oliveira, L. S.: A Review of Texture Classification Methods and Databases. In *2017 30th SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T)*. Oct 2017. ISSN 2474-0705. pp. 1–8. doi:10.1109/SIBGRAPI-T.2017.10.

[20] Chalfoun, J.; Majurski, M.; Blattner, T.; et al.: MIST: Accurate and Scalable Microscopy Image Stitching Tool with Stage Modeling and Error Minimization. *Scientific Reports*. vol. 7, no. 1. 2017. ISSN 2045-2322. doi:10.1038/s41598-017-04567-y.
Retrieved from: https://doi.org/10.1038/s41598-017-04567-y

[21] Chen, S.; Chen, J.; Forte, D.; et al.: Chip-level anti-reverse engineering using transformable interconnects. In *2015 IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFTS)*. Oct 2015. ISSN 1550-5774. pp. 109–114. doi:10.1109/DFT.2015.7315145.

[22] Chen, S.; Shinseki, B.; Barutha, C.; et al.: Infrared imaging and backside failure analysis techniques on multilayer CMOS technology. In *Proceedings of the 1997 6th International Symposium on the Physical and Failure Analysis of Integrated Circuits*. Jul 1997. pp. 17–20. doi:10.1109/IPFA.1997.638066.

[23] Chen, Y.-C.: Enhancements to SAT Attack: Speedup and Breaking Cyclic Logic Encryption. *ACM Trans. Des. Autom. Electron. Syst.*. vol. 23, no. 4. May 2018: pp. 52:1–52:25. ISSN 1084-4309. doi:10.1145/3190853.
Retrieved from: http://doi.acm.org/10.1145/3190853

[24] Chernyy, N.: HOW TO: write an IC Friday post. Online. [accessed 01-Oct-2014]. Retrieved from: https://web.archive.org/web/20110710033130/http://microblog.routed.net/2008/07/15/how-to-write-an-ic-friday-post/

[25] Chung, D.: *Materials for electronic packaging*. Butterworth-Heinemann. 1995. ISBN 978-0750693141.

[26] Cocchi, R. P.; Baukus, J. P.; Chow, L. W.; et al.: Circuit camouflage integration for hardware IP protection. In *2014 51st ACM/EDAC/IEEE Design Automation Conference (DAC)*. Jun 2014. ISSN 0738-100X. pp. 1–5. doi:10.1145/2593069.2602554.

[27] Cole, E. I.: Non-destructive IC defect localization using optical beam-based imaging. In *2008 IEEE Custom Integrated Circuits Conference*. Sep 2008. ISSN 0886-5930. pp. 53–56. doi:10.1109/CICC.2008.4672018.

[28] Cole, E. I.; Soden, J. M.; Rife, J. L.; et al.: Novel failure analysis techniques using photon probing with a scanning optical microscope. In *Proceedings of 1994 IEEE International Reliability Physics Symposium*. Apr 1994. pp. 388–398. doi:10.1109/RELPHY.1994.307808.

[29] Cottone, F.; Basset, P.; Guillemet, R.; et al.: Non-linear MEMS electrostatic kinetic energy harvester with a tunable multistable potential for stochastic vibrations. In *2013 Transducers Eurosensors XXVII: The 17th International Conference on Solid-State Sensors, Actuators and Microsystems (TRANSDUCERS EUROSENSORS XXVII)*. Jun 2013. ISSN 2159-547X. pp. 1336–1339. doi:10.1109/Transducers.2013.6627024.

[30] Courbon, F.: Practical Partial Hardware Reverse Engineering Analysis. *Journal of Hardware and Systems Security*. Apr 2019. ISSN 2509-3436. doi:10.1007/s41635-019-00068-8.
Retrieved from: https://doi.org/10.1007/s41635-019-00068-8

[31] Courbon, F.; Skorobogatov, S.: Direct charge measurement in Floating Gate transistors of Flash EEPROM using Scanning Electron Microscopy. In *In Proceedings of the 42nd International Symposium for Testing and Failure Analysis (ISTFA)*. 2016.

[32] Courbon, F.; Skorobogatov, S.; Woods, C.: Reverse Engineering Flash EEPROM Memories Using Scanning Electron Microscopy. In *Smart Card Research and Advanced Applications*. Mar 2017. ISBN 978-3-319-54668-1. pp. 57–72.

[33] Courtland, R.: 3D X-ray Tech for Easy Reverse Engineering of ICs, IEEE Spectrum. Online. 2017. [accessed 23-Jul-2018].
Retrieved from: https://spectrum.ieee.org/semiconductors/processors/3d-xray-tech-for-easy-reverse-engineering-of-ics

[34] Courtland, R.: X-rays map the 3D interior of integrated circuits. IEEE Spectrum. Mar 2017.

[35] D. Halliday, J. W., R. Resnick: *Fyzika - 4. část: Elektromagnetické vlny – Optika – Relativita*. VUTIUM, PROMETHEUS. 2006. ISBN 80-214-1868-0.

[36] Davis, S.: Intel's e-DRAM Shows Up In The Wild. Online. 2014. [accessed 03-Feb-2019].
Retrieved from: https://www.semiconductor-digest.com/2014/02/07/intels-e-dram-shows-up-in-the-wild-2/

[37] Davis, W. R.; Wilson, J.; Mick, S.; et al.: Demystifying 3D ICs: the pros and cons of going vertical. *IEEE Design Test of Computers*. vol. 22, no. 6. Nov 2005: pp. 498–510. ISSN 0740-7475. doi:10.1109/MDT.2005.136.

[38] Dini, M.: *Nano-Power Integrated Circuits for Energy Harvesting*. PhD. Thesis. alma. Maggio 2015.
Retrieved from: http://amsdottorato.unibo.it/6947/

[39] Dofe, J.; Gu, P.; Stow, D.; et al.: Security Threats and Countermeasures in Three-Dimensional Integrated Circuits. In *Proceedings of the on Great Lakes Symposium on VLSI 2017*. GLSVLSI '17. New York, NY, USA: ACM. 2017. ISBN 978-1-4503-4972-7. pp. 321–326. doi:10.1145/3060403.3060500.
Retrieved from: http://doi.acm.org/10.1145/3060403.3060500

[40] Dofe, J.; Yu, Q.; Wang, H.; et al.: Hardware security threats and potential countermeasures in emerging 3D ICs. In *2016 International Great Lakes Symposium on VLSI (GLSVLSI)*. May 2016. pp. 69–74. doi:10.1145/2902961.2903014.

[41] Drahanský, M.: *Fingerprint Recognition Technology - Related Topics*. Lambert Academic Publishing. 2011. ISBN 978-3-8443-3007-6. 172 p.
Retrieved from: http://www.fit.vutbr.cz/research/view_pub.php?id=9636

[42] Drahanský, M.: *Liveness Detection in Biometrics*. InTech - Open Access Publisher. 2011. ISBN 978-953-307-487-0. pp. 179–198. doi:10.5772/17205.

[43] Drahanský, M.; Březinová, E.; Orság, F.; et al.: Classification of Skin Diseases and Their Impact on Fingerprint Recognition. In *Proceedings of the Special Interest Group on Biometrics and Electronic Signatures*. Lecture Notes in Informatics. Society for Informatics. 2009. ISBN 978-3-88579-249-9. pp. 173–176.
Retrieved from: http://www.fit.vutbr.cz/research/view_pub.php?id=9048

[44] Drahanský, M.; Doležel, M.; Urbánek, J.; et al.: Influence of Skin Diseases on Fingerprint Recognition. *Journal of Biomedicine and Biotechnology*. vol. 2012, no. 4. 2012: pp. 1–14. ISSN 1110-7243. doi:10.1155/2012/626148.
Retrieved from: http://www.fit.vutbr.cz/research/view_pub.php?id=9900

[45] Drahanský, M.; Kanich, O.: *Influence of Skin Diseases on Fingerprints*. Springer Nature Singapore. 2019. ISBN 978-981-1311-43-7. pp. 1–39. doi:10.1007/978-981-13-1144-4.
Retrieved from: http://www.fit.vutbr.cz/research/view_pub.php?id=11703

[46] Drahanský, M.; Lodrová, D.: Liveness Detection for Biometric Systems Based on Papillary Lines. *International Journal of Security and Its Applications*. vol. 2, no. 4. 2008: pp. 29–37. ISSN 1738-9976.
Retrieved from: http://www.fit.vutbr.cz/research/view_pub.php?id=8792

[47] Drahanský, M.; Orság, F.; Doležel, M.; et al.: *Biometrie (Biometrics)*. Computer Press, s.r.o. first edition. 2011. ISBN 978-80-254-8979-6. 294 p.

[48] Drost, R. J.; Hopkins, R. D.; Ho, R.; et al.: Proximity communication. *IEEE Journal of Solid-State Circuits*. vol. 39, no. 9. Sep 2004: pp. 1529–1535. ISSN 0018-9200. doi:10.1109/JSSC.2004.831448.

[49] Dupuis, S.; Flottes, M.-L.: Logic Locking: A Survey of Proposed Methods and Evaluation Metrics. *Journal of Electronic Testing*. vol. 35, no. 3. Jun 2019: pp. 273–291. ISSN 1573-0727. doi:10.1007/s10836-019-05800-4.
Retrieved from: https://doi.org/10.1007/s10836-019-05800-4

[50] European Commission: COMMISSION DECISION of 4.8.2011, C(2011) 5499 final. Online. 2011. [accessed 06-Dec-2018].
Retrieved from: https://ec.europa.eu/home-affairs/sites/homeaffairs/files/e-library/docs/comm_native_c_2011_5499_f_en.pdf

[51] European Commission: Borders & Visas - Document security. Online. 2012. [accessed 12-Oct-2015].
Retrieved from:
http://ec.europa.eu/home-affairs/doc_centre/borders/borders_doc_en.htm

[52] Fahad, H.; Hasan, M.; Li, G.; et al.: Thermoelectricity from wasted heat of integrated circuits. *Applied Nanoscience*. vol. 3, no. 3. Jun 2013: pp. 175–178. ISSN 2190-5517. doi:10.1007/s13204-012-0128-2.
Retrieved from: https://doi.org/10.1007/s13204-012-0128-2

[53] Falk, R. A.: Near IR Absorption in Heavily Doped Silicon - An Empirical Approach. Jan 2000.

[54] Fievre, A.M.P.; Rogers, A.-A.A.; Bhansali, S.: Integrated circuit security: an overview. *Journal of InstItute of smart structures and systems (ISSS)*. vol. 4, no. 1. Mar-Sep 2015: pp. 18–37. ISSN 2319-6408.

[55] Forte, D.; Bhunia, S.; Tehranipoor, M. M.: *Hardware Protection Through Obfuscation*. Springer Publishing Company, Incorporated. first edition. 2017. ISBN 3319490184, 9783319490182.

[56] Fuechsle, M.; Miwa, J. A.; Mahapatra, S.; et al.: A single-atom transistor. Nature Nanotechnology. Macmillan Publishers Limited, part of Springer Nature. 2012. ISSN 1748-3387. pp. 242–246.

[57] Fyrbiak, M.; Strauß, S.; Kison, C.; et al.: Hardware reverse engineering: Overview and open challenges. In *2017 IEEE 2nd International Verification and Security Workshop (IVSW)*. Jul 2017. pp. 88–94. doi:10.1109/IVSW.2017.8031550.

[58] Gu, P.; Li, S.; Stow, D.; et al.: Leveraging 3D technologies for hardware security: Opportunities and challenges. In *2016 International Great Lakes Symposium on VLSI (GLSVLSI)*. May 2016. pp. 347–352. doi:10.1145/2902961.2903512.

[59] Gueron, S.: Attacks on Encrypted Memory and Constructions for Memory Protection. In *2016 Workshop on Fault Diagnosis and Tolerance in Cryptography (FDTC)*. Aug 2016. pp. 1–3. doi:10.1109/FDTC.2016.20.

[60] Guin, U.; DiMase, D.; Tehranipoor, M.: Counterfeit Integrated Circuits: Detection, Avoidance, and the Challenges Ahead. *J. Electron. Test.*. vol. 30, no. 1. Feb 2014: pp. 9–23. ISSN 0923-8174. doi:10.1007/s10836-013-5430-8.
Retrieved from: http://dx.doi.org/10.1007/s10836-013-5430-8

[61] Guin, U.; Huang, K.; DiMase, D.; et al.: Counterfeit Integrated Circuits: A Rising Threat in the Global Semiconductor Supply Chain. *Proceedings of the IEEE*. vol. 102, no. 8. Aug 2014: pp. 1207–1228. ISSN 0018-9219. doi:10.1109/JPROC.2014.2332291.

[62] Guin, U.; Zhang, X.; Forte, D.; et al.: Low-cost on-chip structures for combating die and IC recycling. In *2014 51st ACM/EDAC/IEEE Design Automation Conference (DAC)*. Jun 2014. ISSN 0738-100X. pp. 1–6. doi:10.1145/2593069.2593157.

[63] Holenda, T.: Odborná konference Quality & Security *(Conference Quality & Security)*. Online. 2007. [accessed 15-Jul-2017].
Retrieved from:
http://www.telematika.cz/tp/semin/pdf/qas_07/4_Holenda_Mayer.pdf

[64] Holenda, T.: Prezentace projektu ePas pro odbornou konferenci iDEME (Presentation of the project ePassport for conference iDEME). Online. 2008. [accessed 15-Jul-2017].
Retrieved from:
http://ideme.net/wp-content/uploads/sites/2/23_Holenda_MV_CR_2008.pdf

[65] Horowitz, P.; Hill, W.: *The Art of Electronics / Third Edition*. Cambridge University Press. 2015. ISBN 978-0521809269. 1224 p.

[66] Houghton Mifflin Company: The American Heritage Dictionary of the English Language, Fourth Edition. Online. [accessed 10-Apr-2019].
Retrieved from: https://www.thefreedictionary.com/analysis

[67] Hutter, M.; Schmidt, J.-M.: The Temperature Side-Channel and Heating Fault Attacks. In *Proceedings of the International Conference on Smart Card Research and Advanced Applications*, vol. 8419. Nov 2013. pp. 219–235. doi:10.1007/978-3-319-08302-5_15.

[68] ICAO: *The International Civil Aviation Organisation: Machine Readable Travel Documents (Part 1, Volume 2)*. International Civil Aviation Organization. first edition. 2006. ISBN 92-9194-757-1.

[69] ICAO: *Machine Readable Passports; Passports with Machine Readable Data Stored in Optical Character Recognition Format (Part 1, Volume 1)*. vol. 1. International Civil Aviation Organization. sixth edition. 2006. ISBN 978-92-9231-139-1.

[70] ICAO: *Machine Readable Official Travel Documents; MRtds with Machine Readable Data Stored in Optical Character Recognition Format (Part 3, Volume 1)*. vol. 1. International Civil Aviation Organization. third edition. 2008. ISBN 978-92-9231-139-1.

[71] ICAO: Supplemental Access Control forMachine Readable Travel Documents, Version 1.1. Technical report. Apr 2014.

[72] ICAO: *Part 10: Logical Data Structure (LDS) for Storage of Biometrics and Other Data in the Contactless Integrated Circuit (IC)*. International Civil Aviation Organization. seventh edition. 2015. ISBN 978-92-9249-798-9.

[73] ICAO: *Part 11: Security Mechanisms for MRTDs*. International Civil Aviation Organization. seventh edition. 2015. ISBN 978-92-9249-799-6.

[74] ICAO: *Part 9: Deployment of Biometric Identification and Electronic Storage of Data in MRTDs*. International Civil Aviation Organization. seventh edition. 2015. ISBN 978-92-9249-797-2.

[75] Imeson, F.; Emtenan, A.; Garg, S.; et al.: Securing Computer Hardware Using 3D Integrated Circuit (IC) Technology and Split Manufacturing for Obfuscation. In *Presented as part of the 22nd USENIX Security Symposium (USENIX Security 13)*. Washington, D.C.: USENIX. 2013. ISBN 978-1-931971-03-4. pp. 495–510. Retrieved from: https://www.usenix.org/conference/usenixsecurity13/technical-sessions/presentation/imeson

[76] Jones, M. H.: *A Practical Introduction to Electronic Circuits, Third edition*. Cambridge University Press. 1996. ISBN 978-0-52-147286-9. 544 p.

[77] Juretus, K.; Savidis, I.: Importance of Multi-parameter SAT Attack Exploration for Integrated Circuit Security. In *2018 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS)*. Oct 2018. pp. 366–369. doi:10.1109/APCCAS.2018.8605696.

[78] Kahng, A.; Lienig, J.; Markov, I.; et al.: *VLSI Physical Design: From Graph Partitioning to Timing Closure*. Jan 2011. ISBN 978-90-481-9590-9. doi:10.1007/978-90-481-9591-6.

[79] Kanda, K.: 1.27Gb/s/pin 3[mgr]W/pin Wireless Superconnect (WSC) Interface Scheme. In *2003 IEEE International Solid-State Circuits Conference, 2003. Digest of Technical Papers. ISSCC.*. Feb 2003. ISSN 0193-6530. pp. 186–187. doi:10.1109/ISSCC.2003.1234193.

[80] Kanich, O.; Drahanský, M.: *State of the art in fingerprint recognition*. IET Book Series on Advances in Biometrics. The Institution of Engineering and Technology.

2018. ISBN 978-1-78561-224-4. pp. 83–110.
Retrieved from: http://www.fit.vutbr.cz/research/view_pub.php?id=11692

[81] Karami, E.; Prasad, S.; Shehata, M. S.: Image Matching Using SIFT, SURF, BRIEF and ORB: Performance Comparison for Distorted Images. *CoRR*. vol. abs/1710.02726. 2017. 1710.02726.
Retrieved from: http://arxiv.org/abs/1710.02726

[82] Karmakar, R.; Kumar, H.; Chattopadhyay, S.: On Finding Suitable Key-Gate Locations In Logic Encryption. In *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*. May 2018. ISSN 2379-447X. pp. 1–5. doi:10.1109/ISCAS.2018.8351235.

[83] Ke, S.; Teng-Sing, W.; Yeop, A. B.; et al.: 3D Printing of Interdigitated Li-Ion Microbattery Architectures. *Advanced Materials*. vol. 25, no. 33. 2013: pp. 4539–4543. doi:10.1002/adma.201301036.
Retrieved from:
https://onlinelibrary.wiley.com/doi/abs/10.1002/adma.201301036

[84] Kerckhoffs, A.: La cryptographie militaire. *Journal des Sciences Militaires*. vol. IX. 1883: pp. 161–191.

[85] Kocher, P. C.: Timing Attacks on Implementations of Diffie-Hellman, RSA, DSS, and Other Systems. In *Proceedings of the 16th Annual International Cryptology Conference on Advances in Cryptology*. CRYPTO '96. London, UK, UK: Springer-Verlag. 1996. ISBN 3-540-61512-1. pp. 104–113.
Retrieved from: http://dl.acm.org/citation.cfm?id=646761.706156

[86] Kömmerling, O.; Kuhn, M. G.: Design Principles for Tamper-resistant Smartcard Processors. In *Proceedings of the USENIX Workshop on Smartcard Technology on USENIX Workshop on Smartcard Technology*. WOST'99. Berkeley, CA, USA: USENIX Association. 1999. pp. 2–2.

[87] Kryszczuk, K.; Richiardi, J.: *Encyclopedia of Cryptography and Security*. Springer, Boston, MA. Jan 2011. ISBN 978-1-4419-5905-8.

[88] Kumagai, J.: Chip detectives [reverse engineering]. *Spectrum, IEEE*. vol. 37, no. 11. Nov 2000: pp. 43–48. ISSN 0018-9235. doi:10.1109/6.880953.

[89] Kumar, A.: SRAM cell design with minimum number of transistor. In *2014 Recent Advances in Engineering and Computational Sciences (RAECS)*. Mar 2014. pp. 1–3. doi:10.1109/RAECS.2014.6799510.

[90] Kumar, D.; Ryu, Y.: A Brief Introduction of Biometrics and Fingerprint Payment Technology. In *2008 Second International Conference on Future Generation Communication and Networking Symposia*, vol. 3. Dec 2008. pp. 185–192. doi:10.1109/FGCNS.2008.11.

[91] Kumar, N.: *Comprehensive Physics XII*. Laxmi Publications. 2003. ISBN 978-81-7008-592-8. 1809 p.

[92] Kundra, S.; Dureja, A.; Bhatnagar, R.: The study of recent technologies used in E-passport system. In *2014 IEEE Global Humanitarian Technology Conference - South Asia Satellite (GHTC-SAS)*. Sep. 2014. pp. 141–146. doi:10.1109/GHTC-SAS.2014.6967573.

[93] Kwon, I.: *Integrated Circuit Design for Radiation Sensing and Hardening*. PhD. Thesis. University of Michigan. 2015.

[94] Learning about Electronics: Transistor Schematic Symbols. Online. [accessed 03-Feb-2019].
Retrieved from: http://www.learningaboutelectronics.com/Articles/Transistor-schematic-symbols.php

[95] Lucarelli, N.; Cavone, M.; Muschitiello, M.; et al.: Thermally Induced Voltage Alteration (TIVA) applied to ESD induced failures. *Microelectronics Reliability*. vol. 43. Sep 2003: pp. 1699–1704. doi:10.1016/S0026-2714(03)00337-8.

[96] Majzoobi, M.; Koushanfar, F.; Potkonjak, M.: *Introduction to Hardware Security and Trust*. Jan 2012. ISBN 978-1-4419-8079-3.

[97] Malčík, D.: *Microscopic analysis of chips security*. Master's Thesis. Faculty of Information Technology, Brno University of Technology. 2011.

[98] Malčík, D.; Drahanský, M.: Microscopic Analysis of Chips. In *Security Technology*. Communications in Computer and Information Science. Springer Verlag. 2011. ISBN 978-3-642-27188-5. pp. 113–122. doi:10.1007/978-3-642-27189-2_12.
Retrieved from: http://www.fit.vutbr.cz/research/view_pub.php?id=9848

[99] Malčík, D.; Drahanský, M.: Microscopic Analysis of The Chips: Chips deprocessing. *Advanced Science and Technology Letters*. vol. 2012, no. 7. 2012: pp. 80–85. ISSN 2287-1233.
Retrieved from: http://www.fit.vutbr.cz/research/view_pub.php?id=10041

[100] Malčík, D.; Drahanský, M.: Microscopic Analysis of Chips. *International Journal of Security and Its Applications*. vol. 2016, no. 11. 2016: pp. 47–66. ISSN 1738-9976. doi:10.14257/ijsia.2016.10.11.05.
Retrieved from: http://www.fit.vutbr.cz/research/view_pub.php?id=11107

[101] Maleč, F.: Druhá generace elektronických pasů a nová generace elektronických průkazů o povolení k pobytu *(The second generation of electronic passports and a new generation of electronic documents)*. Online. 2010. [accessed 15-Jul-2017].
Retrieved from:
https://2010.smartcardforum.cz/prezentation/ke_stazeni/malec.pdf

[102] Mangard, S.; Oswald, E.; Popp, T.: *Power Analysis Attacks: Revealing the Secrets of Smart Cards*. Advances in information security. Springer US. 2008. ISBN 9780387381626.
Retrieved from: https://books.google.cz/books?id=YXkASFjeUswC

[103] Massad, M. E.; Garg, S.; Tripunitara, M.: Reverse engineering camouflaged sequential circuits without scan access. In *2017 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*. Nov 2017. ISSN 1558-2434. pp. 33–40. doi:10.1109/ICCAD.2017.8203757.

[104] Mayer, P.: Biometrické pasy v České republice *(Biometric passports in the Czech Republic)*. Online. 2007. [accessed 15-Jul-2017].
Retrieved from:
http://www.telematika.cz/tp/semin/pdf/qas_07/4_Holenda_Mayer.pdf

[105] McMahan, J.; Cui, W.; Xia, L.; et al.: Challenging on-chip SRAM security with boot-state statistics. In *2017 IEEE International Symposium on Hardware Oriented Security and Trust, HOST 2017, McLean, VA, USA, May 1-5, 2017*. 2017. pp. 101–105. doi:10.1109/HST.2017.7951806.
Retrieved from: https://doi.org/10.1109/HST.2017.7951806

[106] mcmaster: mcmaster:st:24c026 [Silicon Pr0n]. Online. 2015. [accessed 19-Jul-2018].
Retrieved from:
https://siliconpr0n.org/archive/doku.php?id=mcmaster:st:24c026

[107] Mick, S.; Wilson, J.; Franzon, P.: 4 Gbps high-density AC coupled interconnection. In *Proceedings of the IEEE 2002 Custom Integrated Circuits Conference (Cat. No.02CH37285)*. 2002. pp. 133–140. doi:10.1109/CICC.2002.1012783.

[108] Moore, G. E.: Cramming More Components onto Integrated Circuits. *Electronics*. vol. 38, no. 8. Apr 1965: pp. 114–117. ISSN 0018-9219.
doi:10.1109/JPROC.1998.658762.

[109] Moradi, A.; Barenghi, A.; Kasper, T.; et al.: On the Vulnerability of FPGA Bitstream Encryption Against Power Analysis Attacks: Extracting Keys from Xilinx Virtex-II FPGAs. In *Proceedings of the 18th ACM Conference on Computer and Communications Security*. CCS '11. New York, NY, USA: ACM. 2011. ISBN 978-1-4503-0948-6. pp. 111–124. doi:10.1145/2046707.2046722.
Retrieved from: http://doi.acm.org/10.1145/2046707.2046722

[110] Morris, J.: Materials Science PART III: ELECTROMAGNETIC PROPERTIES. Online. 2008. [accessed 07-May-2017].
Retrieved from:
http://www.mse.berkeley.edu/groups/morris/MSE200/III-Electromagnetic.pdf

[111] Morton, D.; Gabriel, J.: *Electronics: The Life Story of a Technology*. The Johns Hopkins University Press. 2004. ISBN 978-0801887734. 216 p.

[112] National Institute of Standards and Technology: FIPS 140-2: Security requirements for cryptographic modules. Information Technology Laboratory. 2001.

[113] Ni, Y.; Devos, F.: A 4-transistor static memory cell design with a standard CMOS process. In *IEEE International Symposium on Circuits and Systems,*. May 1989. pp. 162–166 vol.1. doi:10.1109/ISCAS.1989.100317.

[114] Ning, H.; H Pikul, J.; Zhang, R.; et al.: Holographic patterning of high-performance on-chip 3D lithium-ion microbatteries. *Proceedings of the National Academy of Sciences of the United States of America*. vol. 112. May 2015.

[115] Nohl, K.; Evans, D.; Starbug, S.; et al.: Reverse-engineering a Cryptographic RFID Tag. In *Proceedings of the 17th Conference on Security Symposium*. SS'08. Berkeley, CA, USA: USENIX Association. 2008. pp. 185–193.

[116] Novotný, R.; Kadlec, J.; Kuchta, R.: NAND Flash Memory Organization and Operations. In *Journal of Information Technology & Software Engineering*, vol. 5. Brno, CZ: Longdom Publishing. 2015. ISSN 2165-7866. doi:10.4172/2165-7866.1000139. 8 p.
Retrieved from: https://www.longdom.org/open-access/nand-flash-memory-organization-and-operations-2165-7866-1000139.pdf

[117] NXP Semiconductors: *P5CD080/P5CN080/P5CC080/P5CC073V0B*. May 2007. rev. 1.1.
Retrieved from:
https://www.commoncriteriaportal.org/files/epfiles/0410_ma1b.pdf

[118] Odstrcil, M.; Holler, M.; Raabe, J.; et al.: High resolution 3D imaging of integrated circuits by x-ray ptychography. 2018. pp. 10656–10658. doi:10.1117/12.2304835.
Retrieved from: https://doi.org/10.1117/12.2304835

[119] Orland, K.: MAME devs are cracking open arcade chips to get around DRM. Online. 2017. [accessed 19-Jul-2018].
Retrieved from: https://arstechnica.com/gaming/2017/07/mame-devs-are-cracking-open-arcade-chips-to-get-around-drm/

[120] Oswald, D.; Paar, C.: Breaking Mifare DESFire MF3ICD40: Power Analysis and Templates in the Real World. In *Cryptographic Hardware and Embedded Systems – CHES 2011*, edited by B. Preneel; T. Takagi. Berlin, Heidelberg: Springer Berlin Heidelberg. 2011. ISBN 978-3-642-23951-9. pp. 207–222.

[121] Pecht, M.; Tiku, S.: Bogus: electronic manufacturing and consumers confront a rising tide of counterfeit electronics. *IEEE Spectrum*. vol. 43, no. 5. May 2006: pp. 37–46. ISSN 0018-9235. doi:10.1109/MSPEC.2006.1628506.

[122] Perakslis, C.; Michael, K.; Michael, M. G.; et al.: Perceived barriers for implanting microchips in humans: A transnational study. In *2014 IEEE Conference on Norbert Wiener in the 21st Century (21CW)*. Jun 2014. pp. 1–8. doi:10.1109/NORBERT.2014.6893929.

[123] Pikhay, E.; Roizin, Y.; Nemirovsky, Y.: Ultra-Low Power Consuming Direct Radiation Sensors Based on Floating Gate Structures. *Journal of Low Power Electronics & Applications*. vol. 7, no. 3. 2017: pp. 20–22. doi:10.3390/jlpea7030020.

[124] Pisarski, A.: An Exploration into the World of Microprocessors. Online. [accessed 03-Feb-2019].
Retrieved from:
http://www2.optics.rochester.edu/workgroups/cml/opt307/spr06/alex/

[125] Powell, D.: Finding Solutions to China's E-waste Problem. Online. 2013. [accessed 19-Jul-2018].
Retrieved from: https://ourworld.unu.edu/en/assessing-and-improving-the-e-waste-problem-in-china

[126] Quadir, S. E.; Chen, J.; Forte, D.; et al.: A Survey on Chip to System Reverse Engineering. *J. Emerg. Technol. Comput. Syst.*. vol. 13, no. 1. Apr 2016: pp.

6:1–6:34. ISSN 1550-4832. doi:10.1145/2755563.
Retrieved from: http://doi.acm.org/10.1145/2755563

[127] Rahman, M. T.; Shi, Q.; Tajik, S.; et al.: Physical Inspection Attacks: New Frontier in Hardware Security. In *2018 IEEE 3rd International Verification and Security Workshop (IVSW)*. Jul 2018. pp. 93–102. doi:10.1109/IVSW.2018.8494856.

[128] Rajendran, J.; Sam, M.; Sinanoglu, O.; et al.: Security Analysis of Integrated Circuit Camouflaging. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security*. CCS '13. New York, NY, USA: ACM. 2013. ISBN 978-1-4503-2477-9. pp. 709–720. doi:10.1145/2508859.2516656.
Retrieved from: http://doi.acm.org/10.1145/2508859.2516656

[129] Rajendran, J.; Sinanoglu, O.; Karri, R.: Is split manufacturing secure? In *2013 Design, Automation Test in Europe Conference Exhibition (DATE)*. Mar 2013. ISSN 1530-1591. pp. 1259–1264. doi:10.7873/DATE.2013.261.

[130] Rašek, L.: Elektronické pasy – jak fungují (Electronic passports – how it works). In *Proceedings of the 29th conference EurOpen.CZ*. Mikulov, CZ: EurOpen.CZ. 2006. ISBN 80-86583-11-2. pp. 15–46.
Retrieved from: https://europen.cz/Anot/29/HLAVNI.pdf

[131] Roshanisefat, S.; Mardani Kamali, H.; Sasan, A.: SRCLock: SAT-Resistant Cyclic Logic Locking for Protecting the Hardware. In *Proceedings of the 2018 on Great Lakes Symposium on VLSI*. GLSVLSI '18. New York, NY, USA: ACM. 2018. ISBN 978-1-4503-5724-1. pp. 153–158. doi:10.1145/3194554.3194596.
Retrieved from: http://doi.acm.org/10.1145/3194554.3194596

[132] Sabarillo, R. M.; Mocorro, C. O.: Indoor light energy harvesting system for battery recharging and wireless sensor networks implemented in 90nm CMOS technology. In *2015 International Conference on Humanoid, Nanotechnology, Information Technology,Communication and Control, Environment and Management (HNICEM)*. Dec 2015. pp. 1–5. doi:10.1109/HNICEM.2015.7393174.

[133] Samyde, D.; Skorobogatov, S.; Anderson, R.; et al.: On a new way to read data from memory. In *First International IEEE Security in Storage Workshop, 2002. Proceedings.*. Dec 2002. pp. 65–69. doi:10.1109/SISW.2002.1183512.

[134] Saraswat, K.: Interconnect Scaling. Online. [accessed 03-Feb-2019].
Retrieved from: https://web.stanford.edu/class/ee311/NOTES/InterconnectScalingSlides.pdf

[135] Scherz, P.; Monk, S.: *Practical electronics for inventors, Fourth edition*. McGraw Hill Professional. 2016. ISBN 978-1-25-958755-9. 1072 p.

[136] Schlangen, R.; Leihkauf, R.; Kerst, U.; et al.: Functional IC analysis through chip backside with nano scale resolution - E-beam probing in FIB trenches to STI level. In *2007 14th International Symposium on the Physical and Failure Analysis of Integrated Circuits*. Jul 2007. ISSN 1946-1542. pp. 35–38. doi:10.1109/IPFA.2007.4378053.

[137] Schobert, M.: All Chips Reversed. *Die Datenschleuder*. vol. 94. 2010: pp. 17–36. ISSN 0930-1054.

[138] Schor, D.: IEDM 2017 + ISSCC 2018: Intel's 10nm, switching to cobalt interconnects. Online. 2018. [accessed 03-Feb-2019].
Retrieved from: https://fuse.wikichip.org/news/525/iedm-2017-isscc-2018-intels-10nm-switching-to-cobalt-interconnects/2/

[139] Schwartz, G. C.; Srikrishnan, K. V.: *Handbook of Semiconductor Interconnection Technology, Second Edition*. Boca Raton, FL, USA: CRC Press, Inc.. 2006. ISBN 1574446746.

[140] Sengupta, A.; Mazumdar, B.; Yasin, M.; et al.: Logic Locking with Provable Security Against Power Analysis Attacks. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*. 2019: pp. 1–1. ISSN 0278-0070. doi:10.1109/TCAD.2019.2897699.

[141] Shakya, B.; Asadizanjani, N.; Forte, D.; et al.: Chip editor: Leveraging circuit edit for logic obfuscation and trusted fabrication. In *2016 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*. Nov 2016. pp. 1–8. doi:10.1145/2966986.2967014.

[142] Shakya, B.; Tehranipoor, M. M.; Bhunia, S.; et al.: *Introduction to Hardware Obfuscation: Motivation, Methods and Evaluation*. Cham: Springer International Publishing. 2017. ISBN 978-3-319-49019-9. pp. 3–32. doi:10.1007/978-3-319-49019-9_1.
Retrieved from: https://doi.org/10.1007/978-3-319-49019-9_1

[143] Shamsi, K.; Li, M.; Meade, T.; et al.: AppSAT: Approximately deobfuscating integrated circuits. In *2017 IEEE International Symposium on Hardware Oriented Security and Trust (HOST)*. May 2017. pp. 95–100. doi:10.1109/HST.2017.7951805.

[144] Sheetal, S.: Technical analysis of security mechanisms used in RFID E-passport, related threats, security and privacy issues. Viterbi School of Engineering, University of Southern California. Jul 2006. California, Los Angeles, CA, USA.

[145] Skorobogatov, S.: Flash Memory 'Bumping' Attacks. In *Proceedings of the 12th International Conference on Cryptographic Hardware and Embedded Systems*. CHES'10. Berlin, Heidelberg: Springer-Verlag. 2010. ISBN 3-642-15030-6, 978-3-642-15030-2. pp. 158–172.
Retrieved from: http://dl.acm.org/citation.cfm?id=1881511.1881526

[146] Skorobogatov, S.: How Microprobing Can Attack Encrypted Memory. In *2017 Euromicro Conference on Digital System Design (DSD)*. Aug 2017. pp. 244–251. doi:10.1109/DSD.2017.69.

[147] Skorobogatov, S.; Woods, C.: Breakthrough Silicon Scanning Discovers Backdoor in Military Chip. In *Proceedings of the 14th International Conference on Cryptographic Hardware and Embedded Systems*. CHES'12. Berlin, Heidelberg: Springer-Verlag. 2012. ISBN 978-3-642-33026-1. pp. 23–40. doi:10.1007/978-3-642-33027-8_2.
Retrieved from: http://dx.doi.org/10.1007/978-3-642-33027-8_2

[148] Skorobogatov, S. P.: Semi-invasive attacks - A new approach to hardware security analysis. Technical report. Computer Laboratory, University of Cambridge. 2005.

[149] Skorobogatov, S. P.; Anderson, R. J.: Optical Fault Induction Attacks. In *Revised Papers from the 4th International Workshop on Cryptographic Hardware and Embedded Systems*. CHES '02. London, UK, UK: Springer-Verlag. 2003. ISBN 3-540-00409-2. pp. 2–12.

[150] Skvortsov, D.; Yin Shyang Ng, M.; Lundquist, T.; et al.: Laser Voltage Imaging: A New Perspective of Laser Voltage Probing. Nov 2010.

[151] Steehler, J.: Chemistry: The Central Science, 7th ed. (Brown, Theodore L.; LeMay, H. Eugene, Jr.; Bursten, Bruce E.) Chemistry and Chemical Reactivity, 3rd ed. (Kotz, John C.; Treichel, Paul, Jr.). *Journal of Chemical Education*. vol. 74, no. 4. 1997. doi:10.1021/ed074p378. 378 p.
Retrieved from: https://doi.org/10.1021/ed074p378

[152] Szendiuch, I.: *Základy technologie mikroelektronických obvodů a systémů*. VUTIUM. 2006. ISBN 80-214-3292-6.

[153] Tajik, S.; Nedospasov, D.; Helfmeier, C.; et al.: Emission Analysis of Hardware Implementations. In *Proceedings - 2014 17th Euromicro Conference on Digital System Design, DSD 2014*. Aug 2014. pp. 528–534.

[154] Tan, B.; Lewicke, A.; Schuckers, S.: Novel methods for fingerprints image analysis detect fake fingers. *Spie Newsroom*. Jan 2008. doi:10.1117/2.1200705.1171.

[155] Tareen, S. A. K.; Saleem, Z.: A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK. In *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*. Mar 2018. pp. 1–10. doi:10.1109/ICOMET.2018.8346440.

[156] Teardown, V.: Motorola MC6847P Video Display Generator. Online. 2018. [accessed 19-Jul-2018].
Retrieved from: https://vintageteardown.com/die-photo/motorola-mc6847p-video-display-generator/

[157] Thorne, S.; Ippolito, S.; Eraslan, M.; et al.: High resolution backside thermography using a numerical aperture increasing lens. Jan 2003.

[158] Tooley, M.: *Electronic Circuits: Fundamentals and Applications*. Elsevier Ltd. 2007. ISBN 978-0-75-066923-8. 440 p.

[159] Torrance, R.; James, D.: Reverse Engineering in the Semiconductor Industry. In *Custom Integrated Circuits Conference, 2007. CICC '07. IEEE*. Sep 2007. ISSN 0886-5930. pp. 429–436. doi:10.1109/CICC.2007.4405767.

[160] Torrance, R.; James, D.: The state-of-the-art in semiconductor reverse engineering. In *2011 48th ACM/EDAC/IEEE Design Automation Conference (DAC)*. Jun 2011. ISSN 0738-100x. pp. 333–338.

[161] Tsutomu Matsumoto, K. Y., Hiroyuki Matsumoto; Hoshino, S.: Impact of artificial „gummy“ fingers on fingerprint systems. In *Proceedings of the Optical Security and Counterfeit Deterrence Techniques IV*, vol. 4677. 2002. doi:10.1117/12.462719. Retrieved from: https://doi.org/10.1117/12.462719

[162] Tummala, R.: *Fundamentals of Microsystems Packaging*. McGraw Hill Professional. 2001. ISBN 978-0071371698. 967 p.

[163] Udalagama, C. J.: Electrical energy generation from body heat. In *2010 IEEE International Conference on Sustainable Energy Technologies (ICSET)*. Dec 2010. ISSN 2165-4387. pp. 1–5. doi:10.1109/ICSET.2010.5684932.

[164] Van Tho, L.; Baeg, K.-J.; Noh, Y.-Y.: Organic nano-floating-gate transistor memory with metal nanoparticles. *Nano Convergence*. vol. 3, no. 1. Apr 2016. ISSN 2196-5404. doi:10.1186/s40580-016-0069-7. 10 p. Retrieved from: https://doi.org/10.1186/s40580-016-0069-7

[165] Vigil, K.; Lu, Y.; Yurt, A.; et al.: Integrated circuit super-resolution failure analysis with solid immersion lenses. *Electronic Device Failure Analysis*. vol. 16. Jan 2014: pp. 26–32.

[166] Vijayakumar, A.; Patil, V. C.; Holcomb, D. E.; et al.: Physical Design Obfuscation of Hardware: A Comprehensive Investigation of Device and Logic-Level Techniques. *IEEE Transactions on Information Forensics and Security*. vol. 12, no. 1. Jan 2017: pp. 64–77. ISSN 1556-6013. doi:10.1109/TIFS.2016.2601067.

[167] Villasenor, J.; Tehranipoor, M.: The Hidden Dangers of Chop-Shop Electronics. Online. 2013. [accessed 19-Jul-2018]. Retrieved from: https://spectrum.ieee.org/semiconductors/processors/the-hidden-dangers-of-chopshop-electronics

[168] Wang, X.; Gao, M.; Zhou, Q.; et al.: *Gate Camouflaging-Based Obfuscation*. Cham: Springer International Publishing. 2017. ISBN 978-3-319-49019-9. pp. 89–102. doi:10.1007/978-3-319-49019-9_4. Retrieved from: https://doi.org/10.1007/978-3-319-49019-9_4

[169] Watson, I.: China: The electronic wastebasket of the world. Online. 2013. [accessed 19-Jul-2018]. Retrieved from: https://edition.cnn.com/2013/05/30/world/asia/china-electronic-waste-e-waste/index.html

[170] Weste, N.; Harris, D.: *CMOS VLSI Design: A Circuits and Systems Perspective, Fourth Edition*. Addison-Wesley Longman, Inc.. 2011. ISBN 978-0321547743. 864 p.

[171] Wikipedia contributors: File:Logic-gate-index.png. Online. [accessed 03-Feb-2019]. Retrieved from: https://commons.wikimedia.org/wiki/File:Logic-gate-index.png

[172] Wikipedia contributors: Microchip implant (human). Online. 2018. [accessed 23-Jul-2018]. Retrieved from: https://en.wikipedia.org/wiki/Microchip_implant_(human)

[173] Wikipedia contributors: Static random-access memory — Wikipedia, The Free Encyclopedia. Online. 2019. [accessed 03-Feb-2019].
Retrieved from: https://en.wikipedia.org/w/index.php?title=Static_random-access_memory&oldid=909028588

[174] www.microchemicals.eu: Wet-Chemical Etchingof Silicon. Online. 2012. [accessed 31-Apr-2019].
Retrieved from:
https://www.seas.upenn.edu/~nanosop/documents/silicon_etching.pdf

[175] Xie, Y.; Bao, C.; Serafy, C.; et al.: Security and Vulnerability Implications of 3D ICs. *IEEE Transactions on Multi-Scale Computing Systems*. vol. 2, no. 2. Apr 2016: pp. 108–122. doi:10.1109/TMSCS.2016.2550460.

[176] Xie, Y.; Srivastava, A.: Anti-SAT: Mitigating SAT Attack on Logic Locking. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*. vol. 38, no. 2. Feb 2019: pp. 199–207. ISSN 0278-0070. doi:10.1109/TCAD.2018.2801220.

[177] Yasin, M.; Mazumdar, B.; Rajendran, J. J. V.; et al.: SARLock: SAT attack resistant logic locking. In *2016 IEEE International Symposium on Hardware Oriented Security and Trust (HOST)*. May 2016. pp. 236–241. doi:10.1109/HST.2016.7495588.

[178] Yasin, M.; Mazumdar, B.; Sinanoglu, O.; et al.: Removal Attacks on Logic Locking and Camouflaging Techniques. *IEEE Transactions on Emerging Topics in Computing*. 2017: pp. 1–1. ISSN 2168-6750. doi:10.1109/TETC.2017.2740364.

[179] Yasin, M.; Mazumdar, B.; Sinanoglu, O.; et al.: Security analysis of Anti-SAT. In *2017 22nd Asia and South Pacific Design Automation Conference (ASP-DAC)*. Jan 2017. ISSN 2153-697X. pp. 342–347. doi:10.1109/ASPDAC.2017.7858346.

[180] Yasin, M.; Sinanoglu, O.: Evolution of logic locking. In *2017 IFIP/IEEE International Conference on Very Large Scale Integration (VLSI-SoC)*. Oct 2017. ISSN 2324-8440. pp. 1–6. doi:10.1109/VLSI-SoC.2017.8203496.

[181] Yu, F.-X.; Jia-Rui, L.; Zheng-Liang, H.; et al.: Overview of Radiation Hardening Techniques for IC Design. *Information Technology Journal*. vol. 9. Jun 2010. doi:10.3923/itj.2010.1068.1080.

[182] Zhang, X.; Tehranipoor, M.: Design of On-Chip Lightweight Sensors for Effective Detection of Recycled ICs. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*. vol. 22, no. 5. May 2014: pp. 1016–1029. ISSN 1063-8210. doi:10.1109/TVLSI.2013.2264063.

[183] Zhang, X.; Tuzzio, N.; Tehranipoor, M.: Identification of Recovered ICs Using Fingerprints from a Light-weight On-chip Sensor. In *Proceedings of the 49th Annual Design Automation Conference*. DAC '12. New York, NY, USA: ACM. 2012. ISBN 978-1-4503-1199-1. pp. 703–708. doi:10.1145/2228360.2228486.
Retrieved from: http://doi.acm.org/10.1145/2228360.2228486

Appendix A

FIB cross-section details

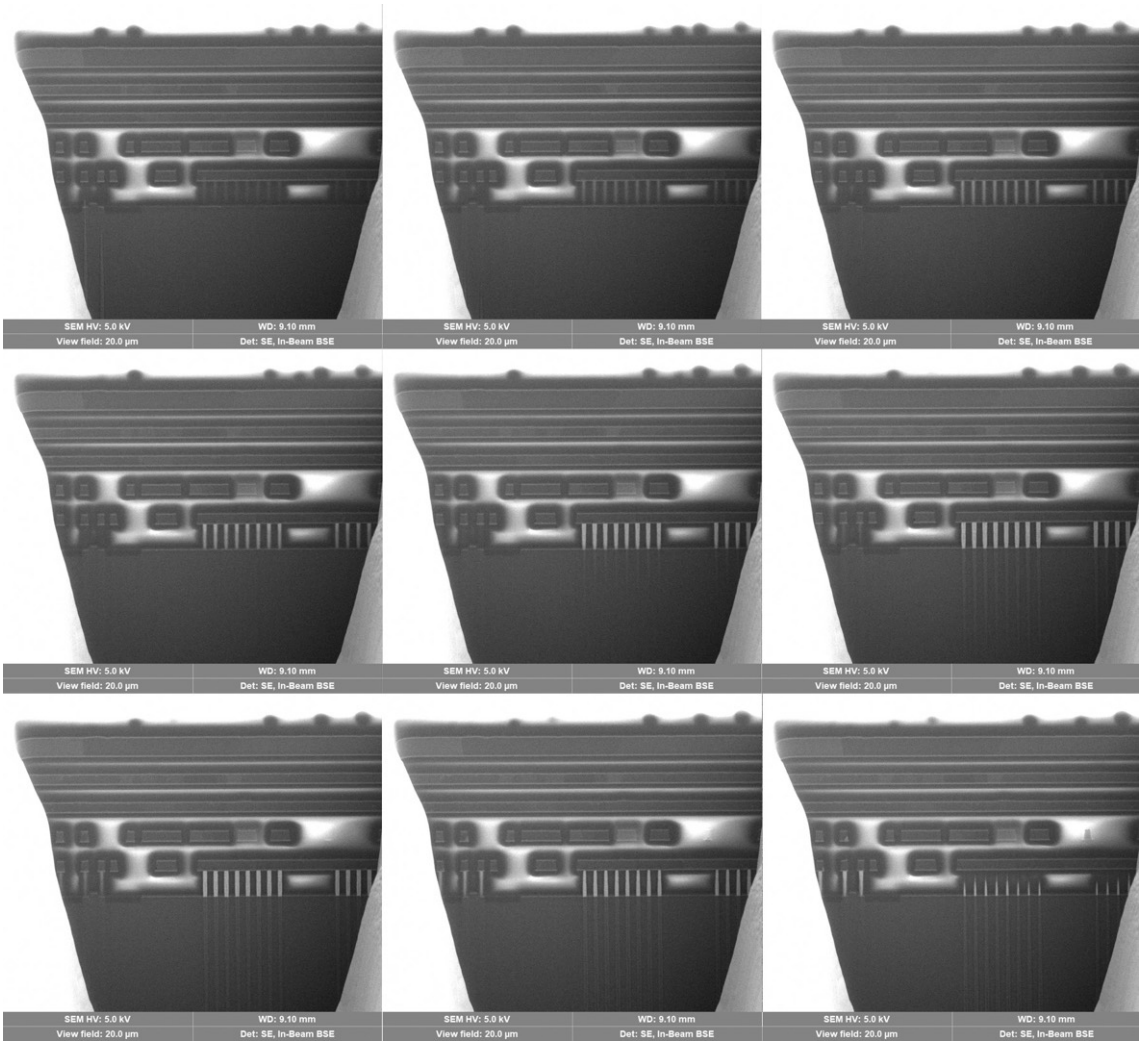


Figure A.1: FIB milling performed on the NXP P5CD080 V0B; cross-section images are scanned after 50 nm milling step. (Source: author's work.)

Appendix B

MIFARE Classic layers

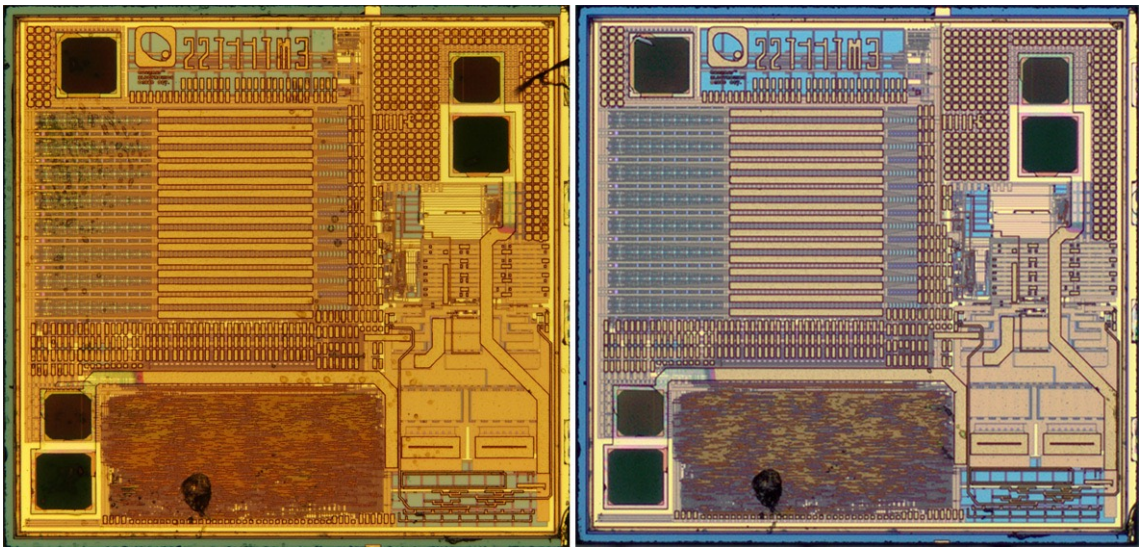


Figure B.1: Left: MIFARE Classic before deprocessing. Right: MIFARE Classic after the first step of deprocessing. (Source: author's work.)

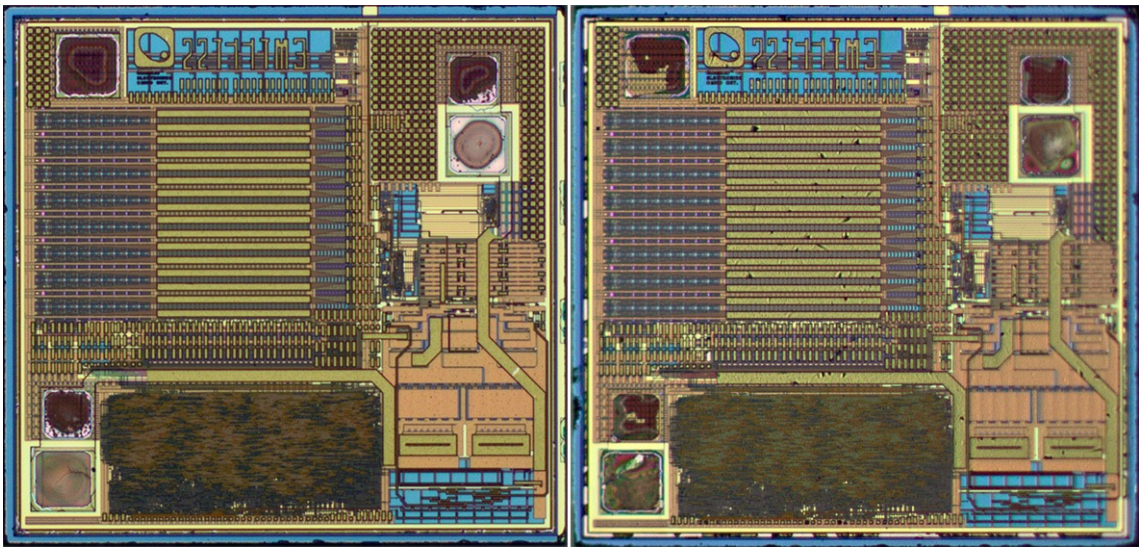


Figure B.2: Left: MIFARE Classic after the second step of deprocessing.. Right: MIFARE Classic after the third step of deprocessing. (Source: author's work.)

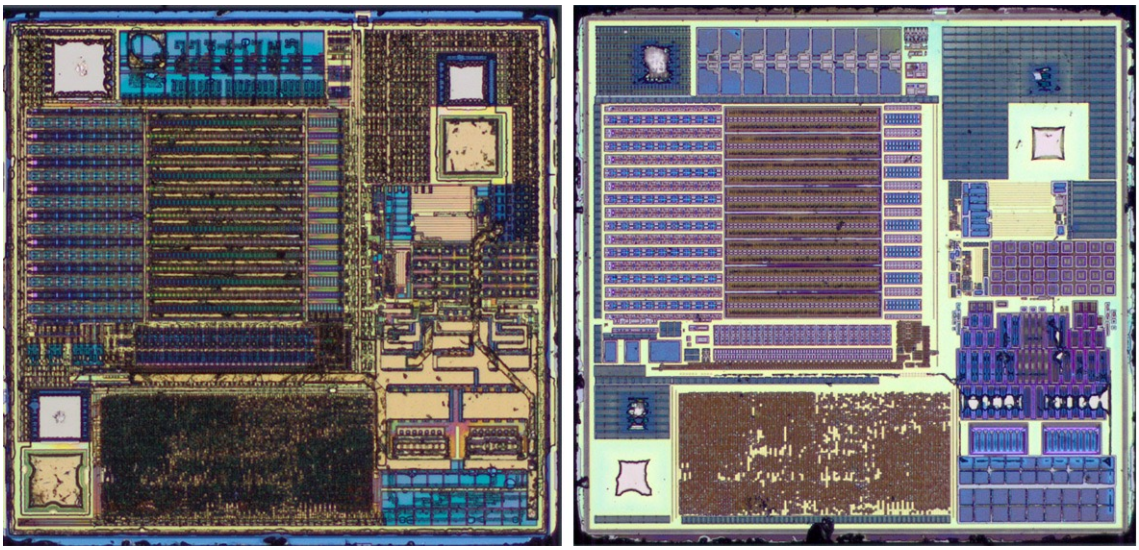


Figure B.3: Left: MIFARE Classic after the fourth step of deprocessing. Right: MIFARE Classic after the fifth step of deprocessing. (Source: author's work.)

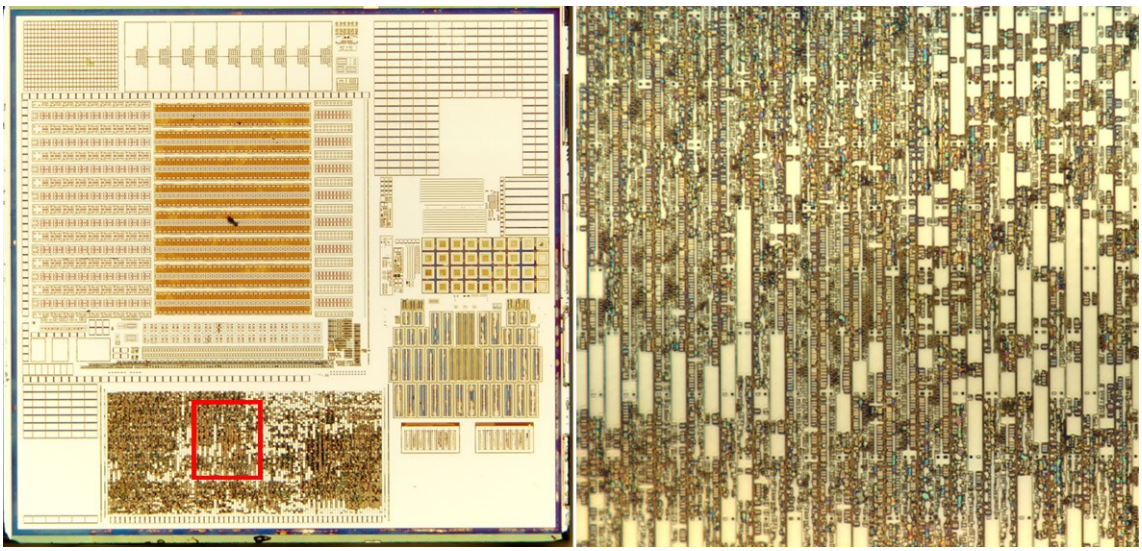


Figure B.4: Left: MIFARE Classic after the sixth step of deprocessing—silicon layer. Right: Cut-out of the transistor field of the MIFARE Classic. (Source: author's work.)

Appendix C

Setup of MIRA3 system

While scanning the whole silicon layer of the chip NXP P5CD080 V0B, each tile is accompanied with log information holding settings of the scanning setup and other relevant information. The first tile was stored with the following information:

```
[MAIN]
AccFrames=1
Date=2018-10-19
Device=MIRA3 XMU
FullUserName=
ImageStripSize=0
Magnification=4.1520e3
Note=TopOverlaping=76; LeftOverlaping=76
PixelSizeX=65.104e-9
PixelSizeY=65.104e-9
SerialNumber=117-0157
TagRevision=2
Time=16:51:18
UserName=
Version=4.2.26.0

[SEM]
3DBeamTiltX=0.0
3DBeamTiltY=0.0
BeamIntensityIndex=10.000
ChamberPressure=87.825e-3
Detector=SE
DwellTime=10.0000e-6
EmissionCurrent=145.64e-6
Gun=Schottky
GunShiftX=0.0
GunShiftY=0.0
GunTiltX=-760.00e-3
GunTiltY=-5.2700
HV=10.0000e3
IMLCenteringX=5.5341
```

```
IMLCenteringY=-5.2783
ImageShiftX=0.0
ImageShiftY=0.0
InBeamExtractor=0.0
InBeamScintillator=0.0
LUTGamma=1.0000
LUTMaximum=255
LUTMinimum=0
MixingMode=0
OBJCenteringX=-3.8300
OBJCenteringY=13.710
OBJPreCenteringX=-2.1100
OBJPreCenteringY=7.6000
PredictedBeamCurrent=263.454700999e-12
PrimaryDetectorGain=39.290
PrimaryDetectorOffset=89.000
ScanMode=FIELD
ScanRotation=0.0
ScanSpeed=5
SpecimenCurrent=167.765567766e-12
SpotSize=49.8568023860e-9
StageRotation=356.49
StageTilt=0.0
StageX=1.1017e-3
StageY=-1.9044e-3
StageZ=19.373e-3
StigmatorX=2.8300
StigmatorY=-2.1900
SystemPressure=233.76e-6
TiltCorrection=0.0
WD=19.356e-3
```

The last tile was then stored with this information:

```
[MAIN]
AccFrames=1
Date=2018-10-20
Device=MIRA3 XMU
FullUserName=
ImageStripSize=0
Magnification=4.1520e3
Note=TopOverlaping=76; LeftOverlaping=76
PixelSizeX=65.104e-9
PixelSizeY=65.104e-9
SerialNumber=117-0157
TagRevision=2
Time=08:02:55
UserName=
Version=4.2.26.0
```

```
[SEM]
3DBeamTiltX=0.0
3DBeamTiltY=0.0
BeamIntensityIndex=10.000
ChamberPressure=87.887e-3
Detector=SE
DwellTime=10.0000e-6
EmissionCurrent=145.49e-6
Gun=Schottky
GunShiftX=0.0
GunShiftY=0.0
GunTiltX=-760.00e-3
GunTiltY=-5.2700
HV=10.0000e3
IMLCenteringX=5.5341
IMLCenteringY=-5.2783
ImageShiftX=0.0
ImageShiftY=0.0
InBeamExtractor=0.0
InBeamScintillator=0.0
LUTGamma=1.0000
LUTMaximum=255
LUTMinimum=0
MixingMode=0
OBJCenteringX=-3.8300
OBJCenteringY=13.710
OBJPreCenteringX=-2.1100
OBJPreCenteringY=7.6000
PredictedBeamCurrent=263.454700999e-12
PrimaryDetectorGain=39.290
PrimaryDetectorOffset=89.000
ScanMode=FIELD
ScanRotation=0.0
ScanSpeed=5
SpecimenCurrent=243.223443223e-12
SpotSize=49.8568023860e-9
StageRotation=356.49
StageTilt=0.0
StageX=-1.4633e-3
StageY=705.63e-6
StageZ=19.373e-3
StigmatorX=2.8300
StigmatorY=-2.1900
SystemPressure=46.664e-6
TiltCorrection=0.0
WD=19.356e-3
```

Appendix D

Publications and Activities

D.1 Publications

D.1.1 Conferences

1. Malčík, D.: Mikroskopická analýza čipů *(Microscopic Analysis of Chips)*, In: Proceedings of the 17th Conference STUDENT EEICT 2011, Brno, CZ, FIT VUT, 2011, p. 306–308. ISBN 978-80-214-4272-6.

2. Malčík, D., Drahanský, M.: Microscopic Analysis of Chips, In: Security Technology, Jeju, Jeju Island, KR, Springer, 2011, p. 113–122. ISBN 978-3-642-27188-5.

3. Malčík, D., Drahanský, M.: Anatomy of Biometric Passports, In: Information Science and Industrial Applications 2012, Cebu, PH, Springer, 2012, p. 258–263. ISSN 2287-1233.

4. Malčík, D., Drahanský, M.: Microscopic Analysis of The Chips: Chips deprocessing, In: The Third International Conference Ubiquitous Computing and Multimedia Applications 2012, Bali, ID, Springer, 2012, p. 80–85. ISSN 2287-1233.

D.1.2 Journals

1. Malčík, D., Drahanský, M.: Anatomy of Biometric Passports, In: Journal of Biomedicine and Biotechnology, Vol. 2012, No. 1, New York, US, p. 1–8. ISSN 1110-7243. *IF: 2,436*

2. Malčík, D., Drahanský, M.: Microscopic Analysis of Chips, In: International Journal of Security and Its Applications, Vol. 2016, No. 11, p. 47–66. ISSN 1738-9976.

3. Malčík, D., Drahanský, M.: Improving The Physical Security Of Microchips Against Side-Channel Attacks, In: International Journal of Advanced Science and Technology, Vol. 2019, No. 127, p. 13–24. ISSN 2207-6360.

D.1.3 Submitted Publications

1. Malčík, D., Drahanský, M.: Improving the Physical Security of Microchips, In: International Journal of Security and its Applications, Vol. 13, Number 2, ISSN 2207-9629. *Accepted; publishing is expected in September 2019.*

D.2 Products

- Heart failure one year patient's mortality model, software, authors: Malčík Dominik, Semerád Lukáš, Drahanský Martin
- Microscopic data analyzer, software, authors: Malčík Dominik, Drahanský Martin

D.3 Projects

- FIT-S-17-4014—Secure and Reliable Computer Systems, BUT
- LD14013—New solutions for multimodal biometrics—enhancement of security and reliability of biometric technologies, COST
- FIT-S-14-2486—Reliability and Security in IT, BUT
- CZ.1.05/1.1.00/02.0123—St. Anne's University Hospital in Brno, International Clinical Research Center
- FR-TI1/195—Research and development of technologies for intelligent optical tracking systems, MPO
- FR1239/2013/Aa—Innovation of laboratory of chips security analysis—SEM Phenom Pure G2 (preparation of the project application in cooperation with prof. Drahanský)
- GD102/09/H083—Information Technology in Biomedical Engineering, GACR

D.4 Teaching

- BIO—Biometric Systems (Biometrické systémy): laboratory exercises
- BIO—Biometric Systems (Biometrické systémy): lectures
- SEN—Intelligent Sensors (Inteligentní senzory): numeric exercises
- SEN—Intelligent Sensors (Inteligentní senzory): laboratory exercises
- SEN—Intelligent Sensors (Inteligentní senzory): lectures

D.4.1 Theses

- Bachelor's Thesis (supervisor)—5 students
- Bachelor's Thesis (consultant)—1 student (FI MUNI)
- Master's Thesis (consultant)—3 students
- Master's Thesis (supervisor)—1 student

D.5 Established Cooperation

- Faculty of Chemistry, Brno University of Technology, CZ
- ONSEMICONDUCTOR, Rožnov pod Radhoštěm, CZ
- TESCAN Brno, s.r.o., Brno, CZ
- Presto Engineering Europe, Caen, FR
- STATE PRINTING WORKS OF SECURITIES, state enterprise *(STÁTNÍ TISKÁRNA CENIN, státní podnik)*, Prague, CZ

D.6 Presentations

- Technische Universität Graz (TU GRAZ), Graz, AT, 2011
- BUSLAB, Brno University Security Laboratory, Brno, CZ, 2012
- Technische Universität Wien (TU WIEN), Vienna, AT, 2012
- FH Campus Wien, Vienna, AT, 2014
- Technische Universität Dresden (TU Dresden), Dresden, DE, 2014
- Fakultät für Mathematik, Informatik und Physik (Faculty of Mathematics, Computer Science and Physics), University of Innsbruck, Innsbruck, AT, 2014

D.7 Others

- Intensive Program on Information Communication Security (IPICS 2011), Corfu, GR, 20. 8. 2011–2. 9. 2011