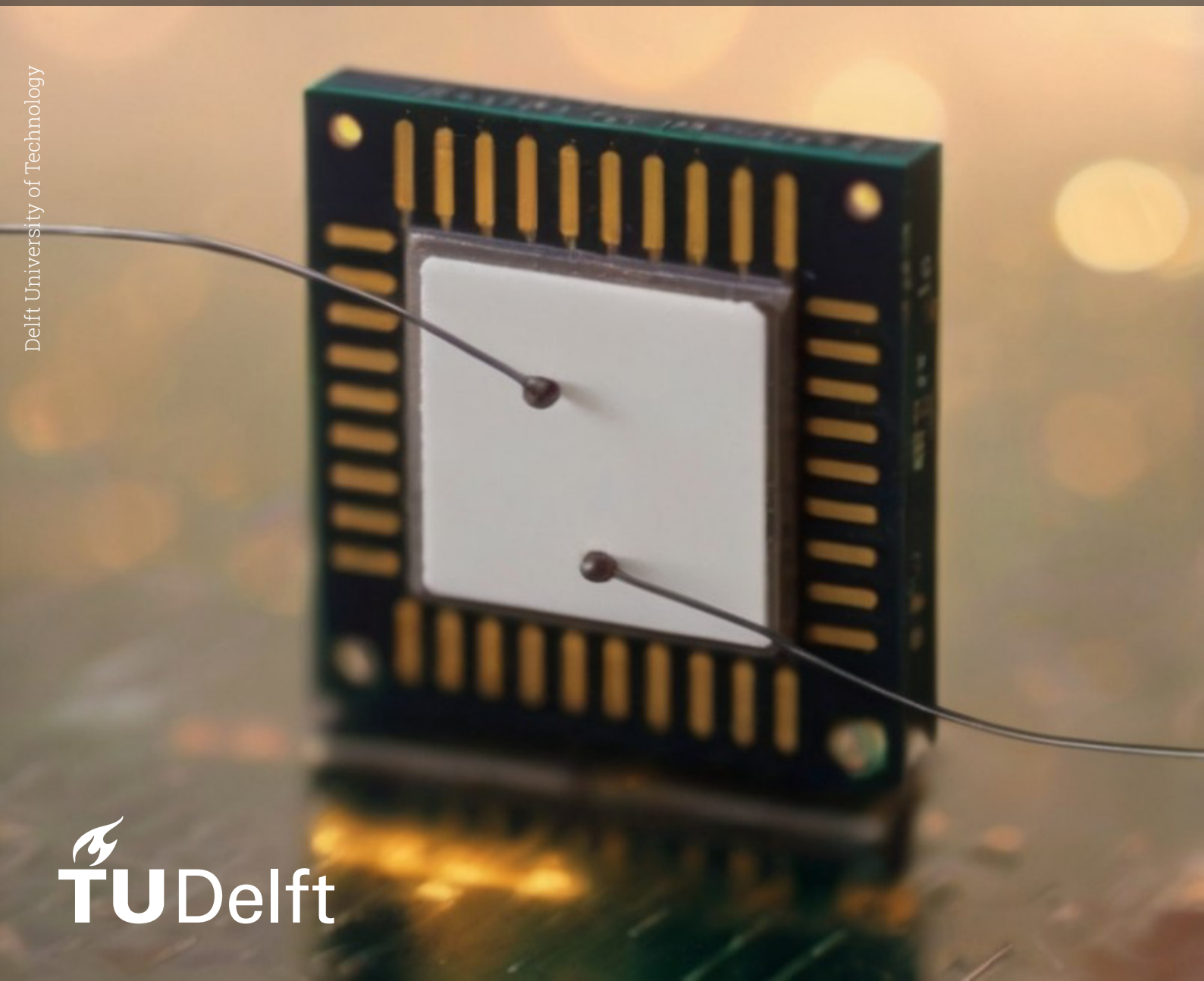
Backside contacting of individual signals in multilayer ICs for chip forensics

Msc Thesis in Microelectronics

Tibbe van der Biezen



Delft University of Technology

TU Delft

Backside contacting of individual signals in multilayer ICs for chip forensics

by

Tibbe van der Biezen

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Thurday July 18, 2024 at 1:00 PM.

Student number:	4686608	
Supervisors:	Dr. ir. S. Vollebregt,	TU Delft
	Dr. ir. J. Romijn	SiTime
	M. Roeloffs MSc	Netherlands Forensic Institute (NFI)
	ir. K. Schot	Netherlands Forensic Institute (NFI)

TU Delft

Abstract

Chip forensics has become an important aspect of law enforcement for retrieving data from data carriers. Improving data encryption techniques, smaller technology nodes and increasing chip design complexity instigate the constant need for new software and hardware hacking methods. The work presented in this thesis investigates the potential of a new hardware hacking method named backside contacting. This method aims to connect a probe to one of the bottom interconnects through the backside to listen to the signals sent over that interconnect.

For this purpose, a recipe has been developed. This recipe is a workflow of numerous processes and steps. Several methods like delayering, cross-sectioning and infra-red (IR) imaging were developed to obtain information about the layout and technology of the chip. Atomic layer deposition (ALD), induction coupled plasma enhanced chemical vapour deposition (ICPECVD) and a focussed ion beam (FIB) were used to create structures on the die, in order to realize a backside interconnect. Finally, various setups were built to test the chip throughout the recipe.

The study shows that each step is possible without losing the data on the chip. The yield of surviving chips after each step, however, should be increased to establish a complete backside contact while keeping the chip alive. The study also indicates that there remains considerable potential for improvement within each step. The results are promising and justify further research into the method of backside contacting.

Acknowledgements

This thesis started as an exciting adventure into the world of secrecy and bureaucracy. A world where the knowledge and skill of reverse engineering were something that I had not encountered before in the courses that I followed during my bachelor's and master's. I had to quickly learn about numerous new techniques and tools, understanding their standard use cases as well as unconventional applications for which they were not originally intended. I thoroughly enjoyed the way of thinking, the practical way of problem-solving and the discussions that were had during the course of this project.

In large part, this was made possible by the amazing group of supervisors who helped me navigate through the many obstacles. First, I would like to thank Sten, who was a never-ending source of information, help and new ideas. His patience and knowledge of the possibilities at the TU Delft labs made this project possible.

Secondly, I would like to thank my supervisors at the NFI. They welcomed me from the start and their immediate trust was inspiring. I am grateful for the level of flexibility they offered and the extensive use of their lab and tools. I would like to especially thank Joost, Mark and Kees; Joost for the warm introduction at the NFI, his critical feedback and enlightening explanations; Mark for his help with the tiniest problems, his enthusiasm and of course his background music; and Kees for his advice, shared knowledge and intriguing stories.

I would also like to thank Folkert, who trained me on most of the tools at the NFI and helped me understand all their intricate parts and functions. Additionally, I would like to thank Bas and Marinus who gave me access to the Kavli Nanolab and trained me on the deposition tools.

Last but certainly not least, I would like to thank my friends and family who have supported me during my whole education. I would like to especially thank my brothers who kept me motivated by always asking intriguing questions and having a listening ear for problems they were not supposed to solve. Special thanks also go out to my parents, who always supported me, whatever decision I made, good or bad and always trusting in a good outcome.

Tibbe van der Biezen
Delft, July 2024

Contents

List of Figures

List of Tables

Acronyms

AES advanced encryption standard 17
AI artificial intelligence 14, 18
ALD atomic layer deposition i, v–vii, 8, 31, 43, 44, 63, 68–71, 74, 75

BPDN backside power distribution network 14, 15, 74

CCPECVD capacitivily coupled plasma enhanced chemical vapour deposition 8
CFET complementary field-effect transistor v, 12, 14
CG control gate 10
CHE channel hot electrons 10
CPA correlation power analysis 16
CPU central processing unit 11
CR cleanroom 23, 43
CVD chemical vapour deposition 8, 9

DES data encryption standard 17
DFA differential fault analysis 16
DRIE deep reactive-ion-etching 7, 8

EBIC electron beam induced current 22
ECC elyptic curve cyphers 17, 18
ECDLP elyptic curve discrete logarithmic problem 18
EDX energy dispersive X-ray 19, 20, 26
EEPROM electronically erasable programmable read-only memory 10
EKL Else Kooi Laboratory 40
ELM electrical linewidth measurement 52, 53
EPROM erasable programmable read-only memory 9, 10
ESD electrostatic discharge 65, 75
ETD Everhart–Thornley detector 41

FG floating gate 10
FIB focussed ion beam i, v, 17, 29–31, 40–42, 45, 48, 60, 66
FN Fowler-Nordheim tunnelling 10
FSM finite state machine 17

GAA-FSFET gate all-around forksheet field-effect transistor v, 12–14
GAA-NSFET gate all-around nanosheet field-effect transistor v, 12–14
GIS gas injection system 41
GPT generative pre-trained transformer 18
GUI graphic user interface 48

HVQFN heat sink very-thin quad flat no lead 4, 48

IC integrated circuit 6, 15
ICPECVD induction coupled plasma enhanced chemical vapour deposition i, v–vii, 8, 31, 43–45, 63, 68–71, 75
IEE insulator enhanced etch 29, 41, 67
IP intellectual property 2
IR infra-red i, iv–vi, 20, 21, 28, 29, 40, 41, 47, 53, 54, 65, 66, 74, 75, 84

LOCOS localized oxidation of silicon 8

MSDS material safety datasheet 23

NFI Netherlands Forensic Institute 1, 25, 35, 45
NIR near infra-red vi, 20, 21, 29, 40, 53, 66
NMOS N-channel metal oxide semiconductor 9, 10, 14

PCB printed circuit board 46, 50, 51
PMOS P-channel metal oxide semiconductor 9, 14
PoP Package-on-Package 5, 15
PVD physical vapour deposition 8, 9

RIE reactive-ion etching 7, 27, 28, 37, 74, 76, 85
RMG replacement metal gate 13

SEM scanning electron microscope v, vi, 28–31, 34, 40, 41, 45, 47–51, 55–58, 60, 61, 66, 74, 75
SIT sidewall image transfer 13
SoC system on a chip 16
SOI silicon on isolator 19
SPA simple power analysis 16
STI shallow trench isolation 8, 19, 57

TSV through-silicon-vias 15
TVT rhin-film transistor 9

VCI voltage contrast imaging 50, 75
VR virtual reality 14

1 Introduction

Trias politica, or separation of powers is a commonly known model to describe the division of power within a government. These branches, often divided into a legislature, an executive, and a judiciary, hold significant power and responsibility. This thesis takes place in the context of the executive and judiciary. These branches strive to execute and interpret the law. In case of a major crime, numerous investigations have to take place to correctly apply this law. This is where the Netherlands Forensic Institute (NFI) plays its part.

When a crime of a certain degree has taken place, evidence is gathered by the police and associated parties. This is then brought to the NFI where it is analysed and prepared for the public prosecutor. To guarantee the maximum amount of information is extracted from a piece of evidence, research is done by multiple disciplines. One of these disciplines is device forensics which consists of both hardware and software analysis. Their goal is to extract data from data carriers, like computers, laptops or phones.

The gathered information may be presented in court and impact the verdict. Therefore, it must be collected using scientifically rigorous methods. These methods, however, need constant updating as technology advances and society finds more secure ways to communicate or store data. This phenomenon occurs not only from a consumer perspective but is also driven by the ongoing cat-and-mouse game between criminals and forensic researchers. One example that illustrates both the necessity for this research as well as its impact was the breakthrough in intercepting crypto communication [1].

1.1. Motivation and problem statement

With the increase in digital devices for both communication and data storage, more and more evidence can be gathered from those devices. Evidence that can point towards a possible perpetrator or clear someone's name. One of the more interesting devices is the mobile phone. It stores important information like communication history, GPS coordinates and photos. However, these devices are developed continuously, implementing better data encryption and more safety features in the hardware. Furthermore, the transistor sizes become smaller and the amount of metal layers keeps increasing. This means new methods need to be developed to access the important information which helps the judiciary.

The first step in accessing the data is using software hacking. Chapter 2 will go into more detail about these methods. It shows that these methods are limited to the external connection pads. Furthermore, the signals and memory data that can be recovered are often encrypted if they can be recovered at all. That is where hardware hacking might be the solution. By being able to access these intermediate signals and possibly the memory directly, new software attacks are possible and encryption can be circumvented.

This thesis will analyse a proposed method called *backside contacting*. The main idea behind this method is that instead of editing the front side of the chip, which is done nowadays, a hole is created through the backside to read the information travelling over the bottom interconnects. This should bring two major advantages. First, by going from the backside it should be easier to keep the chip alive by not breaking the higher interconnects. Secondly, the lower interconnects generally contain more raw data since they are directly on top of the transistors. The encryption of this data is therefore limited,

increasing the possibility to decrypt and use this information.

Keeping the chip alive is important in forensic research because the data on the chip should still be intact and accessible. This means that the main functionalities of the chip should still be performed like power-up and flash reads. This requirement does not apply to most reverse engineering techniques and processes which are mainly used for intellectual property (IP) protection and analyzing state-of-the-art chip manufacturing technology. That means, either these processes have to be adapted or new processes need to be designed.

1.2. Thesis objectives

The main objective of this thesis is to validate the proof of principle of being able to connect to the bottom metal interconnects while keeping the chip alive. This was divided into three practical objectives; getting to know the available techniques and technology for chip forensics; getting to know the chip fabrication technology and layout; and finally, defining process steps and exploring the best options per step. As a result, a recipe should be designed containing the whole process of creating a backside contact.

1.3. Thesis outline

The thesis consists of six chapters:

2. Background information: a chapter that provides background knowledge, concepts, and the latest research outcomes, regarding chip manufacturing, reverse engineering and contamination risks;

3. Recipe design: a chapter that explains the processes that were used and the considerations that were taken into account while designing the recipe. It concludes with the recipe concept;

4. Experiments: a chapter that describes all the tools, processes and steps that are used in the recipe. Additionally, it describes the various test setups that were designed.

5. Results and evaluation: a chapter that summarizes all the results. Ranging from the chip layout analysis to fault analysis, fabrication results and many intermediate test results. Additionally, it evaluates these results and offers explanations where possible.

6. Conclusion and future work: a chapter that reflects on the thesis objectives, describes some concluding arguments and provides ideas for further research.

2

Background information

As was briefly mentioned in the introduction, the transistor sizes become smaller as are the lower interconnect layers. To accommodate the smaller transistors and increase the computational capacity, the amount of interconnect layers steadily increases. On top of that, the encryption of our data becomes more and more complex and robust[2]. In some areas of research, these layers and data need to be accessed, for example in reverse engineering, testing of chip functionality or when it comes to forensic research. Ample of research is already been done when it comes to reverse engineering and testing of chips[3]. Most if not all of the proposed methods have one thing in common, they damage the chip core functionalities meaning the chip partially or completely stops working. However, in the case of forensic research, it is vital that the chip stays alive in order to be able to get information out of it.

This chapter will be the foundation to determine if it is possible to access the bottom interconnects using existing processes and machines while keeping the chip alive. First of all, background knowledge on packaging, die manufacturing and transistors should be established. This will help to understand better where the interconnects are situated, what they are used for and why it is useful to gain access to them. Next, an overview of the different methods and technologies for the fabrication of chips can be put together. This overview is used to design a recipe while taking into account what is available and which processes are compatible with each other. In chapter 3 a closer look is taken at how to keep the chip alive. It is not only important to carefully select and tune the fabrication process but to also look at contamination and potential future developments.

This chapter is structured in the following way. Section 2.1 will discuss the environment within and the fundamentals on which chips are built. This will include the way the sample chips are provided, different technologies and where this project lies with respect to other fields of research. Section 2.2 will go into detail about analysing the chip and where to find potential areas for holes. Finally, section 2.3 will be about the contamination risk when using different machines and moving samples within and between labs.

2.1. Basics of hardware hacking

To understand why hardware hacking is becoming more important nowadays a brief introduction to hacking in general is needed. First, the definition of hacking varies from field to field but generally involves identifying and exploiting weaknesses of a device or network to gain unauthorized access to its data. As an illustration, a spectrum can be constructed where hardware hacking sits on the right-hand side of hacking techniques, see figure 2.1. Hardware hacking is very destructive and always involves physically altering the device. Examples of hardware hacking are the implementation of counterfeit elements or Trojan horses. The latter means that either during the design or fabrication of the chip extra connections or transistors are made or connections are altered in order to change the functionality of the chip[4]. The left side of the spectrum would then be software hacking, in which the device has no physical or destructive alteration. Examples of these techniques are computer viruses, DDoS attacks or zero-day exploits. In between are techniques which combine both extremes. For example, physically changing a bit in memory can cause some security features to malfunction, letting you access data without a password. Other techniques involve side-channel or fault injection attacks[5].

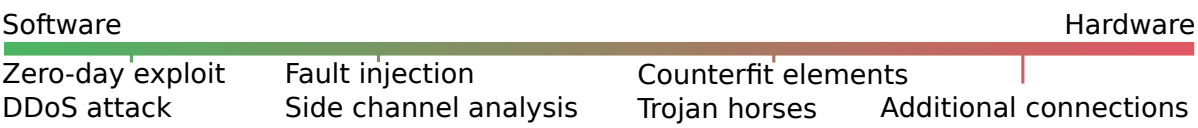


Figure 2.1: Spectrum from software to hardware hacking

This project focuses purely on hardware hacking, where the proposed hardware hacking technique is backside contacting. This is done after the chip is fabricated to extract digital information stored on the chip or to combine it with fault injection or side-channel attacks. Most importantly, this means that during all processing steps, the chip must stay alive and the most important modules must remain functional. To better understand how hardware hacks are designed, an overview of the fundamental processes and physics regarding packaging and fabrication processes is given in the next section. Furthermore, this section aims to indicate the possibilities of tuning these processes in favour of this project. After the basics are covered, a short discussion about the history and future of transistors will be given to get more insight into the future limitations of this attack. The final section is about getting information out of the chip and how this project can help achieve higher yields in successful hacking.

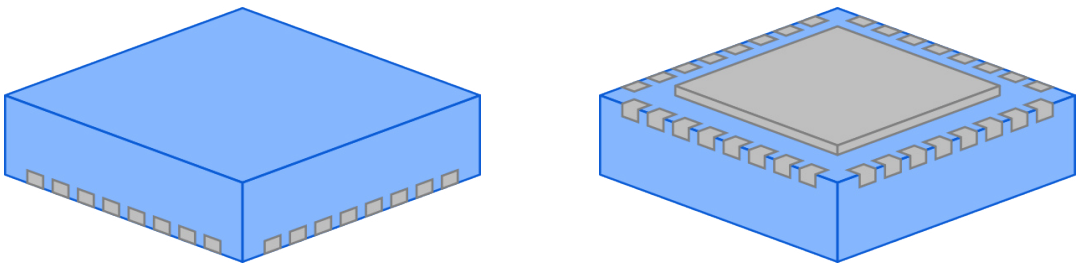
2.1.1. Device physics

Packaging

Because this project is about breaking into a chip, it is only logical to explain the chip fabrication process in reverse. The last step of the fabrication process is the packaging. This step is about encapsulating the die to protect it against the environment.

Outside First, figure 2.2 shows how the chip looks like once it comes out of the factory. Here, the package is shown from the top and bottom side with in blue the package material and in grey the in and output pads. Most of the time these pads are made of aluminium but other materials like gold and nickel are also used. The packaging material can be made of many types of materials like metal or ceramic but is generally a form of plastic. The most common plastics used are epoxies, silicones, polyurethanes or phenolics [6]. A package needs to fulfil many constraints regarding mechanical stresses, heat transfer, thermal expansion, electrical properties and production cost. Additives are used to tailor the plastic to its specific needs.

On the edge of the package, the in- and output pins are clearly visible. The big square in the centre is a heat sink. A setup like this is called a heat sink very-thin quad flat no lead (HVQFN)[7], which is one of many ways to package a chip.



(a) Top side view **(b)** Bottom side view

Figure 2.2: Packaging of the chip

Cross-section Figure 2.3 shows the cross-section of the chip in the package. Now, the actual die is visible as shown in green. During packaging, the chip is mounted on the heat sink using a thermal conductive glue. In the chips used for this project, the heat sink is made of copper with an aluminium coating. This combination gives good thermal conductivity because of the copper centre, while the aluminium protects the copper against corrosion and makes it easier to connect to. Other combinations are also possible like silver or copper and aluminium alloys. More optimal shapes with fins have also been proposed. Additionally, the wire bonds are visible. These ensure that the die is connected to the pads on the outside. Wire bonds can be categorized into two aspects; the wire material and bonding

technique. Popular materials for wires are gold and copper but other materials like silver or aluminium are also used.

Regarding bonding techniques, there are three main techniques. The first and most common is ball bonding which makes use of a needle that at both points welds the wire to the bond pads. The first weld starts by heating the tip of the wire with an induced current. This forms a ball which is placed and welded onto a bond pad by applying heat, pressure and ultrasonic vibrations. The other end of the wire is welded using the second technique which is called wedge bonding. It uses the same principles as ball bonding except the creation of a ball is skipped over.

The third technique is compliant bonding. This technique, also referred to as solid-state bonding, mostly makes use of applying pressure with a small amount of heat. To prevent the sample and wire from too much pressure a compliant medium is used in between the bonding probe and the lower surface. This medium deforms slightly, thereby distributing the forces as required. After the wire bonding, a fluid packaging material is placed which hardens over time resulting in a well-protected chip.

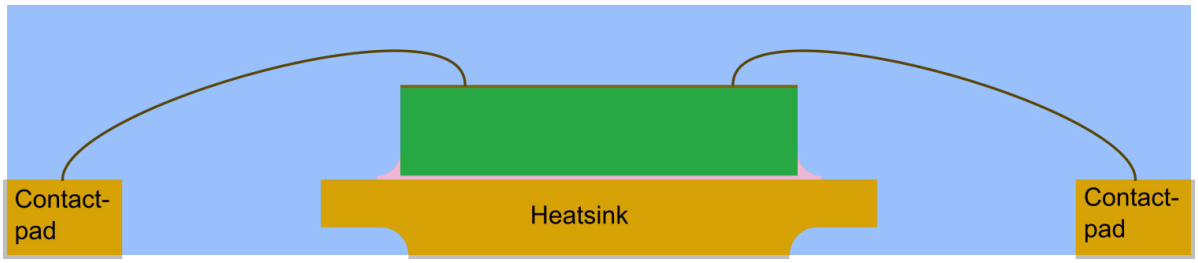
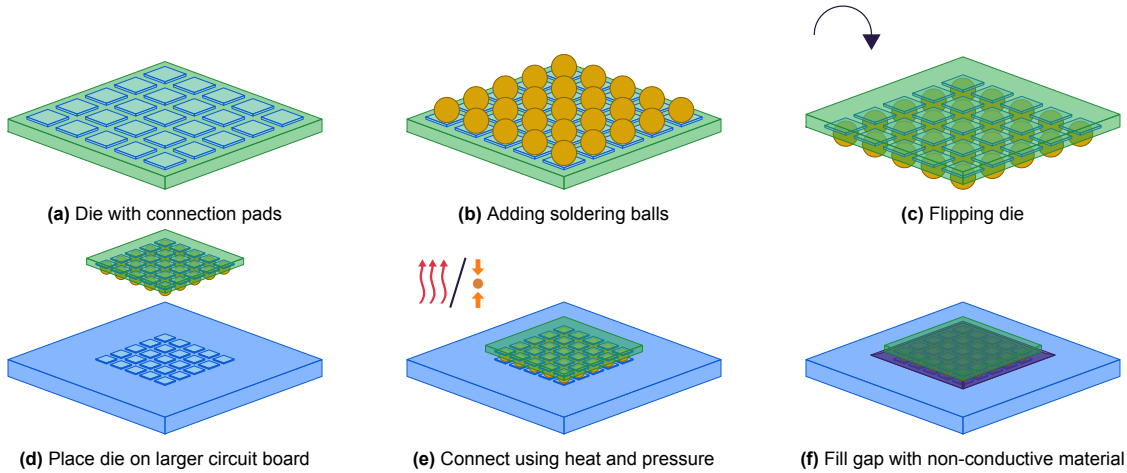


Figure 2.3: Wirebonded cross-section

There exists a second method besides wire bonding which is called flip-chip. The flip-chip technique is used to create shorter wires, flatter packaging and is also useful in case of many contacts. This technique is technologically more difficult because of the constraints on stress and the flatness of both surfaces is much higher. That being said, the principle concept is quite straightforward to explain.

Figure 2.4 shows the basic steps. To start, the die is manufactured in the same way with the connection pads on top. Some steps are undertaken to improve the soldering and adhesion capabilities after which the soldering balls are applied. Then the chip is flipped over, aligned and placed on the larger circuit board. Using pressure and heat the balls are then soldered in place. Finally underfill is applied to help with heat conduction and distributing potential stress differences between the die and circuit board. This technique can also be used to stack multiple packages on top of each other which leads to Package-on-Package (PoP) chips, see section 2.1.2.



(a) Die with connection pads

(b) Adding soldering balls

(c) Flipping die

(d) Place die on larger circuit board

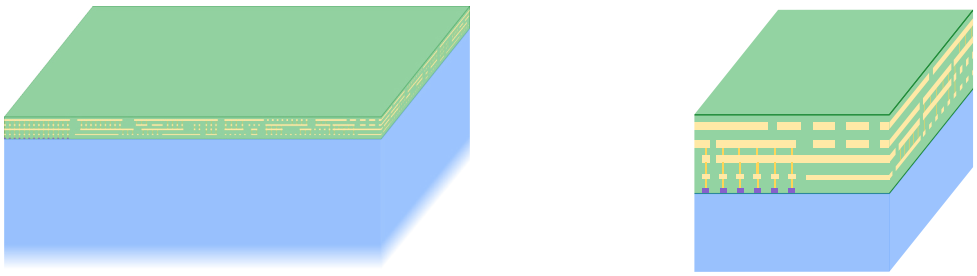
(e) Connect using heat and pressure

(f) Fill gap with non-conductive material

Figure 2.4: Cross-sections of a chip in package

Die

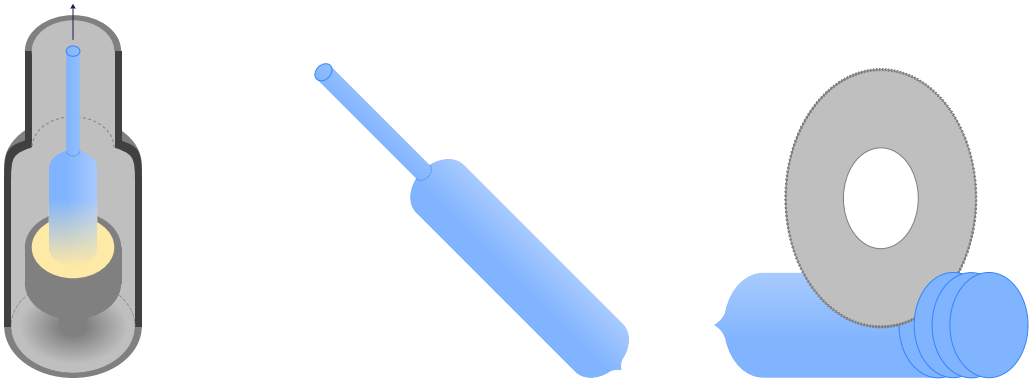
The die is by far the most complex structure inside the whole package and can be described as a small city with roads connecting all the houses. In this case, the roads would be the interconnects and the houses the transistors. Figure 2.5a shows a cut-out version of what a chip could look like with in yellow the interconnects and a little bit darker the vias (vertical interconnects). The chip from the figure shows four metal layers. However, modern chips can have more than 12 layers[8]. The purple blocks indicate a connection to the bulk (in blue) meaning those metal lines would either connect to ground or to any structure present in the silicon. The transistor would be located in the bulk touching the green area. To build this small city there is an almost standardized method consisting of a few general steps which will be discussed briefly below.



(a) Larger overview cut-out **(b)** Zoomed in cut-out

Figure 2.5: Cut-outs of the top of a chip

Wafer A die is a rectangular but mostly square integrated circuit (IC) and starts off as being part of a wafer. Figure 2.6 shows the first few steps. Wafers are made from a silicon ingot. The ingots are grown from a silicon seed also known as the Czochralski method[9]. This is a precise and controlled process which results in a large column of monocrystalline, very pure silicon. The process ensures that the silicon crystal structure is aligned and connected perfectly everywhere. Wafers have to be made so carefully to guarantee uniform electrical properties. In figure 2.6b such a column is shown. This column is then sliced into very thin disks called wafers as depicted in figure 2.6c. The wafers are still very rough after the slicing step which is not preferable. To get the smooth shining wafers many polishing and sometimes even etching steps are done. With newer technologies wafers have been increasing in diameter and thickness. Large diameters mean that there fit more dies on a single wafer or that the dies themselves can be larger. This means that the next processing steps will be more efficient. However, larger sizes will increase the risk of impurities and contamination and since the wafer also acts as a structural basis it should also be made thicker to prevent it from breaking. On the wafer, multiple dies are made in parallel in a grid pattern and are eventually cut out.



(a) Silicon crystal furnace **(b)** Silicon ingot **(c)** Slicing

Figure 2.6: First steps of creating a wafer

Doping The die itself is built up in layers. The bottom layer will be the wafer, which because of the silicon structure will be great to build on and gives a mechanical strong foundation. The next step

is called doping where parts of the wafer are altered to create an excess (n-type) or a lack (p-type) of electrons. These doped parts are called negatively and positively doped wells because they exist at the surface and only reach a certain depth. They are made with processes like diffusion or ion implantation. With ion implantation, ions are shot into the silicon, altering the structure. Depending on the ion type, the acceleration given to the ion and the duration of the process the concentration and depth of the well can be controlled quite accurately. The most common ions for CMOS technology are Boron for p-type and Phosphorus and Arsenic for n-type wells. To make sure only certain parts are doped a protective layer shields the other areas. This layer is applied using lithography which will be discussed in the next section. Nowadays, often the silicon wafer itself is already doped during the growth of the silicon to reduce the time spent on p-type doping and guarantee a uniformly doped bulk, meaning the underside of the chip. After the doping, the surface is still flat but now the silicon is prepared for the transistors to be built.

Lithography Lithography itself consists of three steps. First, a thin film of photoresist is put on the mask using spin coating. The photoresist is a liquid containing three components; a resin which is the base, a sensitizer which reacts to light and a solvent to keep it a liquid. During spin coating, the photoresist is applied while the wafer is spinning. This results in an evenly spread thin layer without any bumps. Finally, the wafer is baked making the photoresist solid. The second step is to expose designated areas to light. For this step, a photomask is needed. This is a thin piece of glass made of nearly pure silicon dioxide therefore ensuring well-known, uniform optics. On top of this glass plate, a pattern is made out of a thin film of chromium. This mask is placed in between the light source and the wafer thereby blocking the light according to the pattern. The parts which are exposed to the light will be chemically changed. Finally, a developer is applied. At this point, it depends on whether a positive or negative photoresist is used. Positive photoresist which has been exposed will dissolve into the developer. With negative photoresist, the areas which were not exposed will dissolve. Now a pattern of protected and exposed areas is created on the wafer which can be used by for example etching or depositing. After these processes are done the remaining photoresist can easily be removed using, for example, a resist stripper which removes the adhesion effect of the photoresist.

Etching and depositing In an iterative process where lithography is combined with etching and depositing, removing and adding materials, the transistors are built up. Build on top of those transistors are the metal layers also called interconnects, with higher layers containing stronger signals and being therefore also physically larger. Once a layer is complete it is almost impossible to change anything below this layer. Therefore it is very important to consider all the materials and for example their melting point during the making of the fabrication process. For this project, a lot of these techniques will be considered and tested. Therefore the next paragraphs will briefly summarize the most common methods.

Etching is the removal of material from the waver or device. This can be done in two ways, namely wet or dry etching.

Wet etching or chemical etching is a method that mostly relies on the chemical reaction between the material that needs to be etched and the etchant, a liquid chemical. For wet etching, the sample is partially or fully submerged in an etchant. This etchant will then chemically and most of the time selectively react to one or more materials on the sample. This etching process is mainly determined by firstly the material that needs to be etched, secondly the etchant, thirdly the temperature and finally the time. Interesting structures and shapes can be made by careful selection of these variables like caves or wells with certain gradients for the walls.

Dry etching makes use of a plasma which releases neutrals and ions. Depending on the pressure in the chamber and the voltages of the magnetic fields applied to the plasma roughly three techniques can be described. At lower pressures, ion milling or sputter etching is done where ionized atoms are shot towards the surface to knock off atoms. In the middle sits reactive-ion etching (RIE) which makes use of large positive ions released by a plasma. These ions are attracted by a negatively charged wafer which results in the surface atoms getting knocked off because of momentum transfer or a chemical reaction. This process is much more selective than the previous technique. Finally, there is deep reactive-ion-etching (DRIE), a more refined implementation of RIE. DRIE exists of multiple short RIE etches with between each etch a deposition of a passivation layer to reduce the amount of sideways

etching. All dry etching techniques are well suited for high aspect ratio etching where DRIE can give the most vertical walls.

There are also less precise ways to remove material. For example, milling or polishing where the chip is ground down or flattened using high-quality sandpaper and felt. This is often done with reverse engineering when a complete layer or large part of a chip needs to be removed.

Depositing is the addition of material onto the wafer or device. In this field of study, this mostly means adding a thin film of only a few atoms till a few micrometres thick onto the entire wafer. This film needs to be as uniform as possible and is made from a large range of materials. To fulfil all the needs of the industry there are roughly four techniques in use.

Thermal oxidation makes use of the very fundamental principle of oxidation. This technique is used to create a thin layer of oxide by exposing the layer to an oxidising agent under very high temperatures. Very pure steam is commonly used as an oxidising agent with a temperature ranging from 800 to 1200$^\circ C$ [10], [11]. Under these circumstances, for example, the silicon bonds with the oxygen in the water, creating a thin layer of silicon dioxide. It could be argued that this is not a deposition method as it consumes part of the silicon in the process however since this method is often part of constructing CMOS transistors it is placed in this section. In the CMOS transistor, the gate is insulated from the silicon with a thin silicon dioxide layer, hence this technique's potential use. To reduce cross-talk between the transistors more localized oxidation of silicon (LOCOS) can be done. Finally, nowadays trenches are made beforehand which then also get filled with the dielectric silicon dioxide called shallow trench isolation (STI) [12].

chemical vapour deposition (CVD) is a widely used method to deposit material on a wafer. The basic principle is a vacuum chamber in which the sample is placed together with one or more volatile precursors. The precursors are turned into vapours which in turn react with the surface of the wafer or decompose altogether to create a thin layer on the wafer. Mainly, four variables can change the process. First of all, adaptations can be made to accommodate solid or liquid precursor materials. Next, the pressure inside the chamber can be changed, resulting in better uniformity or higher aspect ratios [13]. Then there is the option to heat the chamber which is not only required for most recipes but is also a useful parameter to tune the growth rates. Last but not least a plasma can be created to increase the reactivity of the precursor. This will result in much lower operating temperatures. Depending on the plasma this can even be down to room temperature. Apart from these four options, there are plenty of other alterations which use different energy sources like UV light, combustion or lasers to let the chemicals react to the wafer.

As a result of the high demand for CVD, many varieties and adaptations were developed to fit all use cases. Three adaptations are worth mentioning in the context of this project, namely ALD, ICPECVD and capacitivily coupled plasma enhanced chemical vapour deposition (CCPECVD). ALD is a specialised form of CVD where the focus is to create one atom thick, uniform layers. To create a thin film, multiple ALD-cycles are done, each cycle consisting of multiple steps. Each step is carried out sequentially, contrary to CVD. The purpose of a step can vary from deposition, a chemical reaction or etching. The precursor used in each step is self-limiting which results in one atom thick layers. To ensure the steps happen sequentially, careful purging, cleaning and evacuation steps are performed in between.

ICPECVD and CCPECVD are two types of CVD where plasma is used to enhance the process. The addition of using a plasma means that processes which would otherwise require high chamber temperatures can now be done at very low temperatures like room temperature or around $100 - 300^\circ C$. Both induction and capacitive coupled plasmas have their benefits like low contamination risk for induction or lower deposition rates for capacitive coupled plasmas.

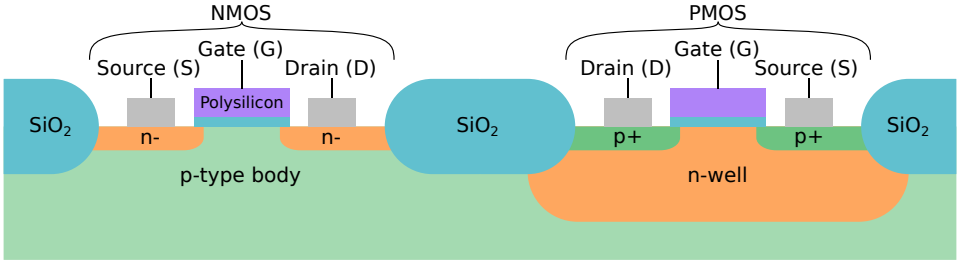
physical vapour deposition (PVD) most common implementations are sputtering or evaporation. Both make use of a vacuum chamber with a source which provides the material for the thin film. Often the source is charged negatively and the wafer positively to guide the source atoms. With sputtering the vacuum chamber is filled with an energized inert gas whose atoms are similar in mass to the source atoms. When the gas atoms collide with the source, source atoms are knocked off and momentum is kept due to similar mass. When the source atoms reach the wafer they stick to the surface thereby creating a thin layer. The source atoms can also stick to the walls of the chamber however they can be reused by bombarding them with ions or atoms. Like other techniques, this process is highly dependent on pressure, temperature and materials used.

Evaporation, as the name already indicates, starts by heating the source till the point where it starts evaporating. The high vacuum causes the evaporated atoms to fly almost directly towards the wafer, without interacting with anything else. Once they collide with the wafer, due to the colder temperature of the wafer they immediately turn back to a solid state. These thin uniform layers are used in a wide variety of applications like rhin-film transistors (TVTs) [14].

Epitaxy is another form of deposition and relies on the fact that crystals can grow. Epitaxy is a refined process where the growth of the crystal is very well defined. It is a similar process with which wafers are made, see section 2.1.1. A crystal structure can have different orientations and the process can be tailored to make use of these characteristics. If tuned correctly, features can be grown on specific orientations while not being able to grow on other orientations. This is called grain-to-grain epitaxy. Epitaxy does not only grow the already present crystal but with the right choice of materials is also able to grow new materials on top. This is respectively called homo and hetero epitaxy. Finally, the growth can be limited to only one direction or in all directions. The latter is then called opotaxy. The method for actually growing the crystal can differ a lot. Under the right circumstances CVD or PVD can be used but also dipping the seed in a melt or covering it with a solid and heating it is being used [15].

Transistor

For the final step, a look at the transistor level is needed. This project aims to make a hole between the transistors to access one of the lower metal interconnects. The most standard representation of a CMOS transistor is given in figure 2.7a. In CMOS technology there are two types of transistors depending on the doping beneath the gate. N-channel metal oxide semiconductor (NMOS) and P-channel metal oxide semiconductor (PMOS) behave the same but inversely. Because of a difference in the mobility of electrons compared to holes (the lack of an electron), careful sizing needs to be applied to ensure equal driving strength. By applying a voltage onto the gate, negative for NMOS and positive for PMOS, a channel is created in which respectively electrons or holes can travel from one side (source) to the other side (drain). The channel can take on different forms as indicated in figure 2.7b. The relation between gate-voltage and conductance is non-linear which is utilized to satisfy many different applications. There are no moving parts, meaning this is an electrically controlled solid state switch. These switches are connected to create an intricate circuit. With enough transistors, these circuits can perform the most complex functions. The connections between the transistors are called the interconnect and exist in multiple layers. Now it becomes clear why the bottom interconnects are of most interest. Because these interconnects are directly on top of the transistors, all the information will need to pass through them first before being processed.



(a) Transistors in off state

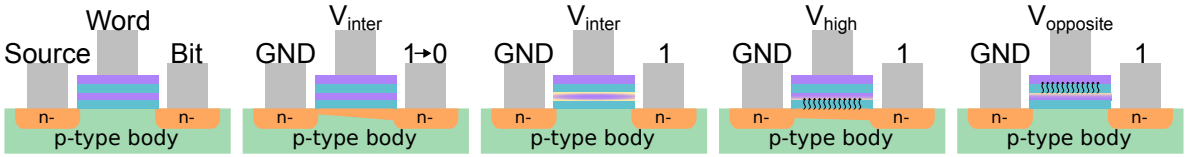
(b) Transistors in different on states

Figure 2.7: Two basic CMOS transitors

One point of interest for device forensics is the memory inside a chip. More specifically the non-volatile memory, meaning that the data will be kept intact when there is no external power supply. Within this category of memory are generally three options. Erasable programmable read-only memory (EPROM) is programmable using an external tool. Once programmed it retains its data until it is

removed using for example UV-light. Since at this point the chip is already integrated into a PCB it is very hard to connect an external tool again to reprogram the EPROM. To solve this the electronically erasable programmable read-only memory (EEPROM) was invented. This device can be reprogrammed on chip. Although this device has great flexibility in being able to program individual bits it is not very information dense. Every bit needs two transistors which with increasing demand for data storage is not efficient enough. Over the years this amount of transistors per bit has decreased resulting in Flash memory. Nowadays, Flash memory is able to store multiple bits in one transistor. One drawback of Flash is that instead of being able to program single bits it is only possible to program in blocks of for example kilobytes. As of now, all non-volatile memory is built up using either floating gate transistors or charge traps. To get a better understanding of where the interesting data traces lie, let's have a closer look at how these transistors look and function.

First of all, they are based on the NMOS transistor as shown in figure 2.7. To lock the transistor into a state, thereby saving it as a zero or one, an extra gate is built on top of the previous one. Only this top gate, called the control gate (CG), is connected to the outside and the gate in the middle is insulated. This insulated gate is called the floating gate (FG) and acts as a charge storage. Using different physical phenomena like channel hot electrons (CHE) and Fowler-Nordheim tunnelling (FN) [16], [17] the floating gate can be charged or discharged. When the FG is charged it dampens the electric field from the CG which means that the threshold voltage of the transistor will be higher. To determine whether charge is stored on the FG a voltage in between the low and high threshold voltage is applied to the CG. If the transistor turns on, it means no charge was present which is interpreted as a zero (or one). Besides reading there are two more functionalities required from a memory cell. Namely writing the bit to a one and the possibility to erase a one. Figures 2.8 indicate how these functionalities take place. Very briefly, figure 2.8a shows the default of a memory cell. The two states in which the cell can be read are indicated in figures 2.8b and 2.8c. Figure 2.8d shows a high voltage, typically over $5V$, being applied on the CG. These high voltages are not immediately available on the chip as they are higher than the supply voltage. Therefore charge pumps are created near Flash memory. By accumulating charge on capacitors they are able to provide the high voltages. The high voltage on the CG creates a large channel which in turn forms channel hot electrons. These electrons charge the FG and write an one into the memory cell. To remove the one, an opposite polarity voltage is applied to the CG thereby creating the FN effect and pulling out the trapped charges. This is indicated in figure 2.8e.



(a) Default memory cell **(b)** Reading a zero **(c)** Reading a one **(d)** Writing a one **(e)** Erasing the one

Figure 2.8: Different operating states of a memory cell

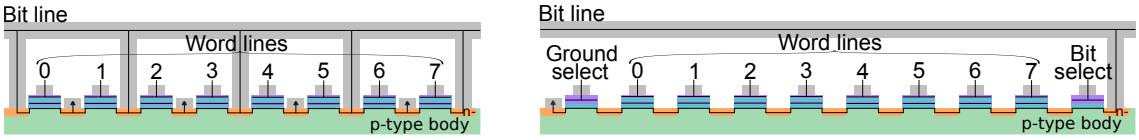
Charge traps work very similar to floating gate transistors with the exception that an insulator is used in the floating gate instead of a conductor. The charge is still injected using CHE however to remove the charge the FN tunnelling needs to be adapted. By changing the current to an induced field, holes are injected into the trap thereby removing the charge. Using charge traps has many benefits for the industry mainly in reduced cost, smaller feature sizes and higher reliability.

A memory block consists of many transistors and over the years many layouts of this principle have been developed but the most popular are the common ground NOR and the NAND configurations [18]. The schematics of these configurations can be found in figure 2.9. It immediately shows that there is a big difference in the way they are wired. Let's first focus on the NOR type Flash. This structure easily follows from putting many single memory cells in parallel. Expanding this topology with many bit lines in parallel will result in words, blocks and pages. By setting one of the word lines to high, this word can be read out on the bit lines. Because all transistors can be controlled directly, it has faster reading speeds compared to a NAND topology. As a result, it is often used to store execution code.

The NAND topology makes use of two extra transistors, one for ground select and one for bit select. Furthermore, the memory cells are placed in series. To read a word not only needs that specific line

to be set at the intermediate voltage but the other word lines will need a voltage above the second threshold to make sure they conduct. Although this limits the reading speed it is well worth it. Because the transistors are in series it is now possible to add two extra transistors, one at each end. They can disconnect or connect a single bit line to the rest of the circuit. Thereby making it possible to address single bits. Furthermore, these transistors reduce a lot of wiring in the ground and bit line circuitry. In practice, this saves up so much space that a higher transistor density can be achieved using a NAND topology. NAND type Flash is mostly used for file storage.

To conclude, a memory block contains a lot of information that is always available to read. Getting information directly out of the Flash without using the controlling logic behind it won't be useful since there is no indication of what kind of information is acquired. Ideally, eavesdropping a bit line, word line or both will yield the most useful information. Combining this data with the code that is being executed could result in important data about its security like where certain encryption keys or bits are stored or what they are.



(a) NOR Flash structure

(b) NAND Flash structure

Figure 2.9: Different Flash architectures

Memory is a very important topic for new research. Together with a central processing unit (CPU), it is a fundamental part of a computer and memory performance is still lacking compared to the processing units. To improve the capabilities of memory without developing entire new processing techniques a new method was invented called package-on-package. It removes the memory from the main die and combines multiple memory dies to create one large memory chip. How technology moves forward is discussed in the next section with more details about package-on-package in section 2.1.2.

2.1.2. Technological advancements

Although there might be earlier versions made, it is generally accepted that the first working transistor was demonstrated by Walter Brattain and John Bardeen to their colleagues at Bell Labs[19]. This was in the afternoon of 23 December 1947 and since then the developments regarding this device have taken off. What started as a point contact transistor, became a BJT, was slowly replaced by MOSFET and is nowadays more and more finFET. Only reducing the transistor size will not be enough as the industry still strives to uphold Moore's law[20] by trying to double the density every two years. More fundamental changes regarding chip design and fabrication will have to be made. The next sections will discuss what these new-generation transistors and chips will look like. The final section is a short discussion on how these advancements might impact the feasibility of backside contacting in the future.

Roadmap

A great source to get an idea of the future of transistors is IMEC, the Interuniversity Microelectronics Centre. This institute provides a basis for all leading micro-electronics companies to work together towards new and improved technologies and their roadmap is therefore a good representation of what is currently in development. Their latest roadmap shows a GAA-NSFET, GAA-FSFET and CFET [21]. The roadmap depicted in figure 2.10 also shows that they expect that in the coming years, the technology will advance to below 10 Ångström ($10\mathring{A} = 1nm$). These numbers are used to refer to the gate length of a transistor plus a few other features like the metal pitch. However, since the $45nm$ node the gate length reached its minimum without introducing many second-order effects[22]. From then on, the naming was used more symbolically to indicate the transistor density. The next paragraphs will discuss the new transistors in more detail.

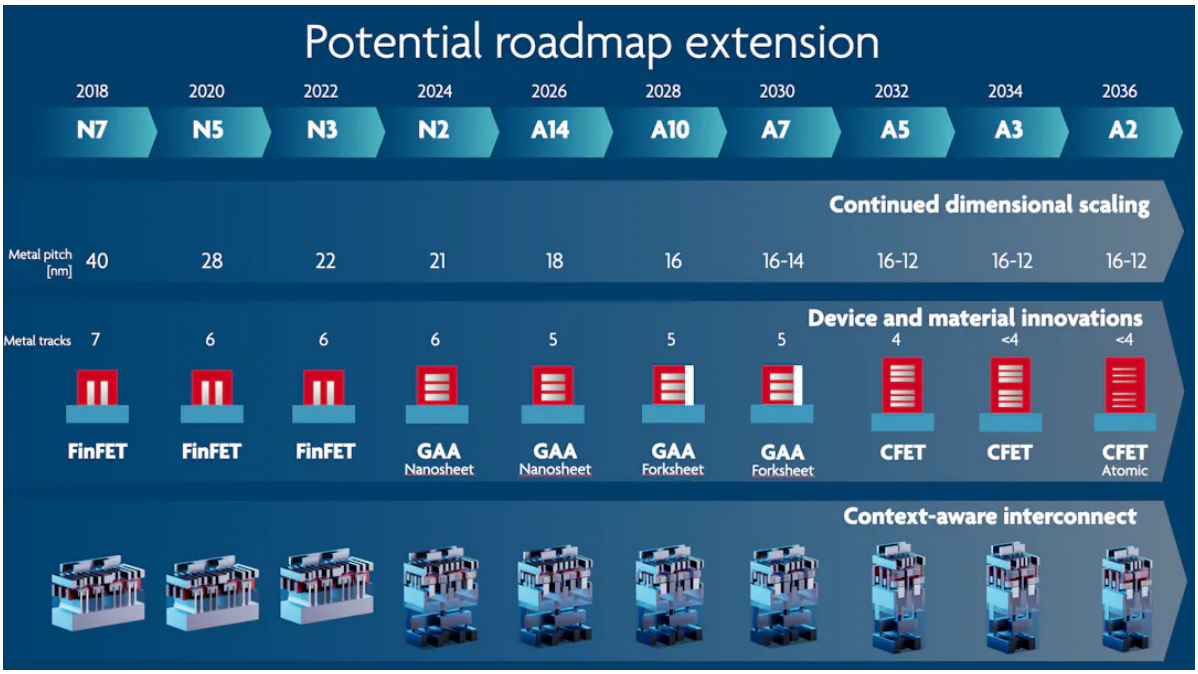
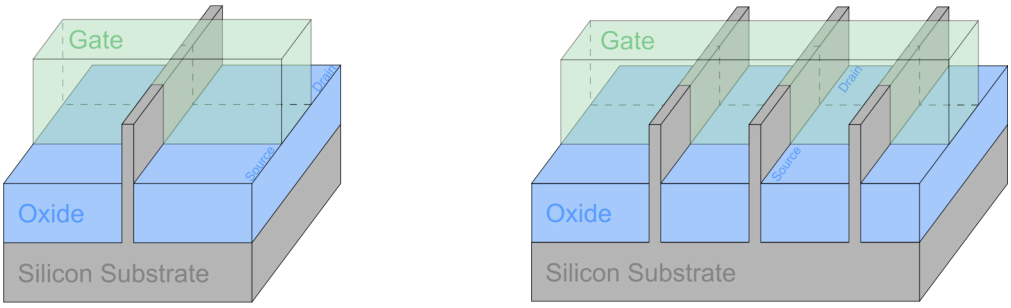


Figure 2.10: Transistor roadmap presented by IMEC [21]

FinFET Since the beginning of the 2010s, FinFET transistors started being used in mass production [23], [24]. What sets FinFETs apart from previously used transistors is the three-dimensional structure. Figure 2.11a shows how the gate now wraps around the channel. This means that the same drain current can now be created with lower voltages, thereby reducing the power. Furthermore, this layout ensures faster switching speeds and a smaller footprint. In practice, CMOS transistors are often put in parallel to increase their driving capability. The same can be done with FinFET as figure 2.11b shows.

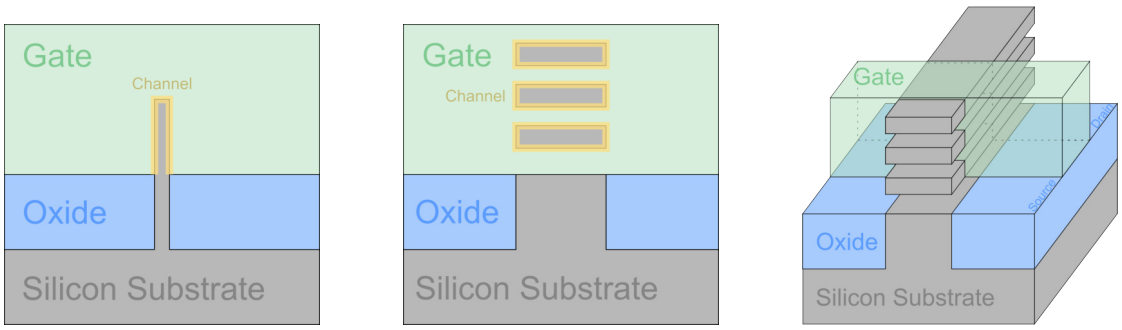


(a) FinFET **(b)** Tri-gate transistor

Figure 2.11: Schematic representations of FinFETs

GAA-FET In this section, the two variations of Gate-All-Around field-effect transistors will be discussed. These transistors are predicted to be the next breakthrough in transistor technology and are the industry's hope to keep scaling according to Moore's law. The new design comes with many benefits, as is indicated in the next few paragraphs, but also comes with significant challenges mainly in the form of higher fabrication complexity. New technologies like sidewall image transfer (SIT) for better resolution below $10nm$, replacement metal gate (RMG) for better temperature budgets and internal spacers for more mechanical strength and reduced parasitic capacitance's [12], [25] are making the step to mass production of GAA-FET's more and more feasible. The development of these transistors is an ongoing field of research to achieve a fabrication technique which ensures high reliability and low impact of process variations.

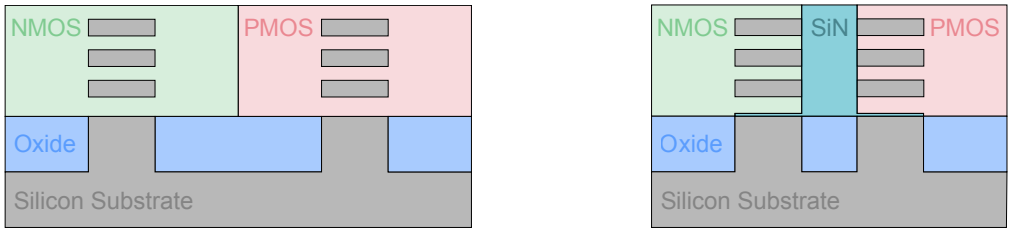
Nanosheet The next iteration of transistors that is currently being worked on are the GAA-NSFETs. In this design, the gate consists of a very thin layer of for example silicon, called a nano-sheet [26]. Figure 2.12b shows how the GAA-NSFET is an improvement on the FinFET. The gate now wraps around thereby creating more channel area which in turn reduces the short channel effects and results in better gate control. Instead of having multiple gates in parallel to improve driving performance, gates can now be stacked on top of each other. This saves valuable area and potentially allows for higher transistor density. Although IMEC suggests this layout because of better design control over the width and height of the sheets, other options with vertical sheets or nano-wires have also been proposed [27].



(a) FinFET cross-section **(b)** GAA-NSFET cross-section **(c)** GAA-NSFET

Figure 2.12: Schematic representation from FinFET to GAA-NSFET

Forksheet A logical improvement after the GAA-NSFET would be the implementation of GAA-FSFET. These structures are very similar with the exception that between the N-type and P-type doped regions a dielectric is placed. This means that the channels can be closer together, saving on precious surface area[28]. Furthermore, this layer shields not only the channels from each other but also the substrate from the doped regions, in turn saving a complex doping step during fabrication[29]. Figure 2.13 shows how this looks as a cross-section and how it relates to the previous nano-sheet design.



(a) NMOS and PMOS GAA-NSFET together

(b) Cross-section of GAA-FSFET

Figure 2.13: Schematic representation from GAA-NSFET to GAA-FSFET

CFET Finally, as of now, all roadmaps end at the CFET technology. The main innovation compared to its predecessor GAA-FSFET is the fact that the negative and positive doped regions are now stacked. The stacked N-type and P-type regions look a lot like the layout of the complementary circuits that are designed today in CMOS, hence the name CFET. Figure 2.14 shows this structure from which it is visible that this technology has again a lower footprint. Compared to its predecessors, the high aspect ratio that comes from stacking structures on top of each other is an even bigger challenge. Time will tell what the limitations on for example the amount of stacked channels or the channel thickness are. In any case, these developments have the potential to start off the Ångström age.

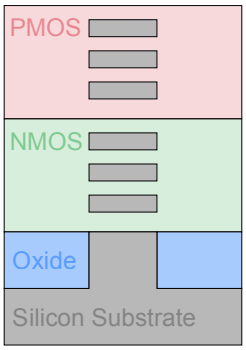


Figure 2.14: Cross-section of a CFET

Additional advancements

Apart from optimizing the transistor density, speed and power consumption it is also necessary to develop the circuitry and interconnects around the transistors. Two of the most well-known obstacles are the memory wall and the power wall. Both limit the performance that can be extracted from the processor and chip. With increasing transistor density these problems get even more important and need a rigorous solution. Finally, there is a great demand from the industry to keep developing within the multi-purpose characteristic of CMOS. For example, high computing power chips for artificial intelligence (AI) or highly efficient chips for virtual reality (VR) headsets. As of now, there are three proposed and partially implemented solutions; backside power distribution network (BPDN), the use of chiplets and chip stacking.

BPDN BPDN has the potential to solve a lot of problems at once. Nowadays, the interconnect is situated on top of the transistors and consists of both the signal network and the power distribution network. By moving the power distribution to the backside of the chip, and burying it inside the silicon bulk, a lot of area is freed to improve the signal network. Its density can be improved, the lines can be shorter and the parasitic capacitance's will be lower meaning that the operating frequency can be higher. The power distribution network also benefits because there is again more space to place the network and it suffers less from leakage and the heat transfer will be easier. To connect the power to the transistors so-called through silicon vias will have to be implemented.

The major foundries have already agreed to research and implement the above-mentioned ideas to be able to connect and extract the maximum performance out of the new smaller and faster transistors discussed in the previous section 2.1.2 and avoid hitting a wall [30].

Chiplets Although the power and memory walls are partially solved by the BPDN there is another solution which can improve the performance of the chip even more. This is the use of chiplets. Chiplets are smaller chips that only contain part of an integrated circuit. By combining multiple different chiplets for example a larger processing unit can be built up. This way the chiplets can be designed independently meaning different technology nodes can be used. Furthermore, the design can become more optimized and better tested before implementation. This means a lower failure rate and because the chip itself is also smaller, fewer defects. On top of that, because the chiplets will contain mostly basic functionality they can be used in multiple systems. Additionally, when a larger integrated circuit breaks down, only the chiplet has to be replaced. Finally, by using chiplets it is possible to create larger chips and systems than would be possible on only one monolithic integrated circuit.

Chip stacking Advancements are also being made on a higher level. Another way to increase memory capacity and processing capabilities is to use more dies. These dies however need to communicate at high speeds meaning they need to be placed together as close as possible. This led to the implementation of 3D ICs. In 3D ICs, two or more dies are stacked on top of each other in one single package. The moment the dies are stacked in the process varies from wafer-to-wafer to die-to-die. It is even possible to have monolithic dies which grow a Si layer on top of which the second 'die' can be built. The stacked dies are connected using through-silicon-vias (TSV) and are mainly used for memory applications. This offers many benefits in decreased cost, area and power consumption and increased bandwidth, security and design options. There are however also some drawbacks like testing individual modules, design complexity and, because the TSVs are relatively big compared to transistors and other components, the overhead of the TSVs.

To tackle those disadvantages a new concept was brought to life, namely package-on-package. With PoP like the name suggests multiple packages are stacked on top of each other. In practice due to limitations in heat dissipation mostly 2 dies are stacked. The stacking happens using the flip chip technology discussed in section 2.1.1. Figure 2.15 shows the cross-section of how such a stack could look like. In this case, both 3D ICs and PoP are combined to create a complete chip. For example, the top package might be two memory dies while the bottom chip could contain the computational logic. There are many advantages to using this technique, two of which stand out. Firstly, the packages can be tested separately. This means that faulty chips can be taken out of assembly earlier on, thereby increasing the yield. The second one is the flexibility of stacking different packages. If designed correctly, combinations can be made and the customer can select which chip combination suits their demands best.

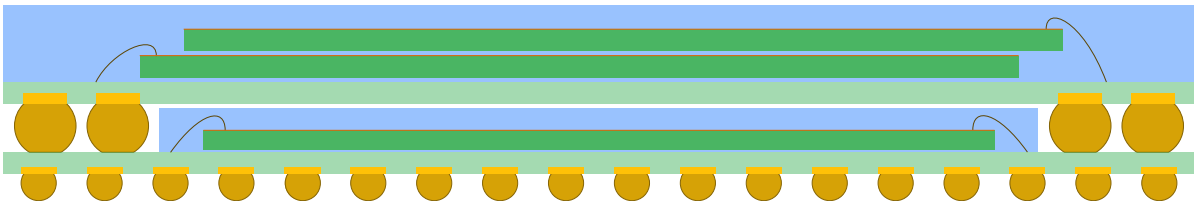


Figure 2.15: Chip stacking

Limitations

Comparing these technological advancements against the goal of this project shows very clearly that accessing the lower interconnects will become more and more difficult. With higher transistor density the free space to create a hole will become smaller and less frequent. Higher aspect ratios in for example FinFETs and stacked transistors show that the the holes also need to be deeper. That is if the backside stays available. With the introduction of the BPDN, thinning the backside will no longer be possible and maybe the frontside will become more attractive again. The same can be said with stacked dies where maybe holes need to be made from the side or chips need to be reconstructed after package removal and drilling.

In short, preferably a scalable technique needs to be found which has to be able to adapt to the challenges of the future. By already striving to create holes and access the bottom and smallest interconnects it can be proven that this new concept will be useful in the future. Luckily, some of these new techniques like internal spacers could also be very helpful to achieve the resolution and aspect ratio

that is required.

Probably one of the biggest challenges will not only be to get a signal from one of the interconnects out of the chip but also to be able to detect it. Since these signal strengths become lower and lower to reduce power consumption it will be harder to read and interpret them. This however is out of the scope of this project and will require further research.

2.1.3. Acquiring data

Important data that is stored on a system on a chip (SoC) is most of the time encrypted. Because of the high throughput of data that is required in modern devices, this encryption is done by designated cryptographic modules. These encryptions are designed to be mathematically very hard to crack so it is easier to extract the encryption key in some other way. As of now, there are mainly two well-known concepts to hack into single chips; side channel analysis and fault injection. After the chip has been processed using the methods developed within this project it will most likely be subject to either one of these methods to get data out of the chip. A little bit of background information about these techniques will therefore be helpful to determine what possible requirements are. Furthermore finding out their limitations could point in a new direction in which this concept can contribute the most.

Side channel analysis

Side channel analysis or SCA for short makes use of the information that is leaked on the side channels like computation time, power consumption, optical observations or electromagnetic radiation. The relation between this information and the activity of the encryption module can help predict encryption keys[31]. Techniques evolved from simple power analysis (SPA) to differential power analysis (DPA) which includes more signals that are analysed and finally correlation power analysis (CPA) which not only uses a modelled power consumption but also takes into account the real power consumption. On top of that electromagnetic sensors are added to detect the EM radiation from the device to get a more complete model.

Although these models get more precise they are not able to accurately represent the real power consumption. One way to solve this mismatch is to use an (almost) identical device to build up a database and thereby an accurate model of its power consumption. Although the ”Hamming distance” or the “Hamming weight” models have dominated the field for a long time a new model, the ”Switching distance” model, changed that. Because this model is not only able to detect a bit switch but also determines if it went from high to low or vice versa, it can more accurately find encryption keys[32]. After creating a model it is then possible to get the encryption keys of the original device with only a few signals to analyse. This methodology is called Template Attack. A variation on this attack is the Stochastic Model Attack which instead of looking at the mean value of a signal applies predefined functions thereby focusing more on improving the model.

Today, however, these techniques are not used as much anymore due to all the countermeasures that the chip manufacturers have implemented. These countermeasures aim to break the correlation between power consumption and intermediate values of the encryption algorithms. This is done in both hardware and software by adding a form of randomness. In software, this can be done by designing the encryption algorithm such that the power consumption characteristic stays the same but the intermediate values are also encrypted. In hardware, random time interrupts or jitter could be used to introduce noise and randomness. Of course, newer, more complex versions like DPA will deal better with these countermeasures so a continuous battle remains between the hacker and chip manufacturer.

Fault injection

Fault injection is an umbrella name for analysing the output response of the cryptographic module after injecting (malicious) faults. One of the more refined methods in this category is called differential fault analysis (DFA) which reduces the number of experiments needed significantly. Injecting the faults happens mostly in the sidebands by exposing the chip to extreme light, EM radiation, low power supply or abnormal clock frequencies.

Applying a low power supply can result in singular bit faults. This can trickle through the computations resulting in large errors. A more sophisticated method is to apply well-timed power spikes or drops. By synchronising such pulses with the clock cycles of the chip, the execution of specific instructions can be bypassed. It has been shown that large integrated circuits can be hacked using this rather cheap and easily implemented method[33]. Similarly, speeding up or slowing down the clock

when bits are written will also lead to bit errors. Extreme light or heat is used more on storage modules and changes or clears the already stored bits. Finally, EM radiation can be used to alter the signals between components. The closer the source is to the chip the better it works although shielding can be applied effectively to protect sensitive areas.

A more invasive but also more expensive way is to use a FIB to repair fuses, connect signal paths or cut certain connects. This way, again, certain functionalities of the chip can be altered or circumvented. All with the purpose of extracting the encryption key or avoiding the encryption of the data altogether.

As there exists a multitude of encryption methods, each method requires its own fault injection technique. Most encryption methods can be divided into two categories, symmetric and asymmetric key ciphers. In the first case the data is encrypted using a key and the same key is used for decryption. This means high-speed encryption and decryption but once the key is known to a third device, this device can listen in and interfere without notice. With asymmetric encryption the encryption key and decryption key are different. This makes it possible to distribute the encryption key without risking other devices to be able to decrypt the message. An additional benefit is that by reversing this method it is possible to make sure a random message is indeed sent by a particular sender because using the private key to encrypt is the only way to create an encrypted message that the public key can decrypt. This is also called signing a message with your signature. Let's discuss some examples of different cyphers and how fault injection can extract the key(s).

The SNOW 3G cypher is an example of a simple symmetric stream cypher. This means it continuously generates a keyword which gets applied to the plain message by an XOR operation. This way fast and power-efficient encryption is applied which is very useful for energy-limited devices like mobile phones. The keyword is generated using amongst other things a finite state machine (FSM) and a pseudo-random shift register. Knowing this opens up the possibility to try and extract the key. By injecting faults into the shift register and measuring the difference it is proven that the key can be extracted [34].

Two other examples of symmetric cyphers are the data encryption standard (DES) and the advanced encryption standard (AES) cyphers. These are both block cyphers meaning they encrypt entire 64-bit data blocks of information. The AES is the newer version of DES and is able to encrypt twice as large blocks using a twice-as-long key. This increases the security quite a bit since getting the key by brute force will be harder and longer blocks are less prone to contain lots of similar data (frequency-based attacks). AES is the standard for network communications and is widely used in commercial products, therefore it has attracted a lot of attention from researchers to see how secure this cypher is. It has proven to be quite complex to extract the encryption key but not impossible. Because this cypher uses multiple rounds of encryption it was first attempted to inject a fault during the final encryption round. After this could be seen at the output and traced back to a specific bit of the encryption key, improvements were made by making it possible to inject a fault in any round. However, this is still a very difficult and timing-sensitive approach. Numerous other attacks have been tried like inserting faults in other parts of the encryption process like wiping the lookup tables proven to either be ineffective or only partially effective[35].

One of the most well-known and widely implemented asymmetric cyphers is RSA, named after its inventors Rivest, Shamir and Adleman [36]. This encryption is mainly based on two prime numbers. The product of the prime numbers (n) in combination with a related relatively prime number (e) results in two keys namely (e, n) and (d, n) where d is the inverse of e times the modulus of n. The security comes from the mathematical very hard and computationally intensive task of factoring n back into its prime numbers. This is however one of the ways in which fault injection can be used to get information to reduce the complexity and time needed for the factorization. A fault needs to be injected during the exponentiation phase of RSA and by using some mathematical tricks it is then possible to deduce the primary numbers. When RSA is used for signing, a fault injection during the signature phase can result in one known bit of the private key. Doing this many times over eventually will result in the complete key. Finally, it is possible to decrypt a message without the need for the private key at all. The basis of this attack is to inject a fault into e and therefore generate two messages with identical contents and product n but different encryption. From this, it is possible to find the two different e_1 and e_2 which in turn leads to a 60% chance to successfully decrypt the message.

The final example of an asymmetric cypher is the newer version of RSA named elyptic curve cyphers (ECC). Instead of prime numbers, this cypher makes use of points on an elliptic curve. To define the curves six parameters need to be agreed upon also called the domain parameters. It is time-consuming

to construct and verify a domain for which all parameters satisfy safety requirements like if the curve is built out of enough points. Therefore publicly available domains exist for different applications. Once a domain is selected, it is possible to generate a key pair based on the curve. This approach is based on the well-known elyptic curve discrete logarithmic problem (ECDLP). Another device can then based on the public key and the agreed-upon curve encrypt a message. This message can then be decrypted by using again the curve and the private key. The major benefit of ECC compared to RSA is that it offers the same amount of security with significantly smaller keys. Although ECCs vulnerabilities are more or less the same as RSA since it uses similar mathematical operations and encryption phases they are less efficient because the operations in ECC are much more complex. Attacks specifically designed for the ECC structure were aimed at reducing the points a curve contains after which brute force can be applied or changing some of the coordinates of the curve to reduce some of the complexity [37].

AI assisted hacking

Over the last few years, enormous advancements have been made in the creation and implementation of AI. A wide variety of models are trained including ones made for hacking. For example, existing models like generative pre-trained transformers (GPTs) are able to help write text or code. These are then used to improve for example phishing emails and malware to acquire data via the user. Other models are used to efficiently filter large amounts of data to find keys or hidden information. Nevertheless, when it comes to hacking into individual chips there are only a few articles published about AI-assisted attacks and remains therefore an open area of research[38].

Limitations

To summarize, both methods need in some way control over the device to force specific output values. Getting to the internal clock line to set up timed attacks or connecting to the internal power lines for DPA has proven to be difficult. Furthermore, these attacks go to great lengths to get intermediate values or signals out of the device. Additionally, directly reading memory blocks cannot be done from the outside. This demonstrates the need for hardware hacking which is aimed at accessing these signals. The current hardware hacking methods are designed to access these signals and values from the frontside or while breaking the chip. With the increase in for example the number of metal layers, this becomes more and more challenging. Backside contacting offers a way around this problem by going via the backside. It will therefore open up new possibilities to retrieve encryption keys and could simplify existing methods.

2.2. Imaging

Part of the process will be to determine where to create a hole. In general, there won't be an overview available of the layout of the die. The first section will discuss how this layout can be obtained. From this, different points of interest for creating a hole can be determined which will be discussed in section 2.2.2. Finally, the machines that can create the holes are different from the ones described in the previous sections. This means the position of the future holes needs to be communicated. Different solutions on how to align the chip again will be discussed in the final section.

2.2.1. Acquiring initial chip layout

To get a first indication of the layout of the chip there are roughly four options available. First of all, a look at the datasheet for the chip's basic functionalities and pin layout can create a first hypothesis on which modules will be present on the inside. The next two options are more destructive beginning with delayering the chip. Here the chip is removed layer by layer exposing all its internals. The other option is to make a cross-section to determine the fabrication node and the amount of metal layers. For a more broad overview, it could be interesting to create either an X-ray image of the package or, as a final option, look through the bulk using infra-red. The next sections will go in order into more detail on these techniques.

Delayering

Delayering can be done from both sides of the chip, each having its benefits and disadvantages. Both techniques are discussed below. The first step however is the same for both sides which is the removal of the packaging. Nitric acid is very effective against packaging material. It will dissolve all packaging while keeping the die intact. By varying the amount of acid from a chip in a full beaker to a droplet on

the chip, a rough selection can be made on where to etch. The die is then further processed on either the front- or backside. After processing the result can be analysed using optical microscopes or for close examination of the transistors a SEM. If the SEM is equipped with energy dispersive X-ray (EDX) spectroscopy this would be a perfect opportunity to get more information on the materials used inside the die.

Frontside When delayering from the front side it is helpful to know which materials are used in which layers because this will impact the process. Although this is normally not disclosed in the datasheet it is very similar across chips and can be predicted quite quickly based on colour and structure. This means that after a few repetitions, a clear view of each interconnect can be obtained.

The chip from the frontside consists of roughly three materials: a protection layer of SiN on top, the metal from the interconnect and the insulation material which is often $SiOx$. Different materials in smaller amounts can for example be found on the interconnect as coating to prevent oxidation or in vias which are often made from a different material. In practice, etching the interconnect and insulation material will cause the other features to be released and therefore removed. This means the complete delayering can be done with two etching steps, simplifying the process without losing too much information.

To etch $SiOx$ an etchant called Alpad is generally used. Alpad is very selective on $SiOx$ and will therefore keep the interconnects intact. Furthermore, Alpad is HF based which means that it will also remove the Ti vias. After a metal layer is fully exposed some cleaning steps should be carried out before moving on to the other etch step. To get rid of the traces Al etchant can be used. The selectivity of Al etchant is the opposite of the previous step meaning that it will mainly etch aluminium. Alternating between these steps will delayer the die slice by slice.

Some remarks can be made which will help to improve this process. First of all, the process heavily depends on the quality of the etchants, temperature and thoroughness of the cleaning in between. To compensate for the lack or abundance of these variables, pictures can be made between the steps in which mainly colour and the state of certain features on the chip give clear indications on what to adjust. Furthermore, there are some side effects to keep in mind. The first one is the existence of coating layers as mentioned before. This can cause slower etching, sudden etch impulses once the coating is dissolved and under etching. Generally, the top layer is coated so especially at the beginning a careful approach is necessary. Finally, some areas etch quicker than others because of the difference in the density of interconnects combined with the fact that etching tends to be faster around edges. A uniform etching process will require multiple iterations and careful tuning of the above-mentioned process dependencies.

That being said, this technique will result in a lot of useful information even if the etching process is not carried out perfectly. Interesting areas can be pointed out and a first impression about different types of materials and fabrication technology can be extracted.

Backside Delayering from the backside in practice means removing the Si bulk to expose the transistors from underneath. A proven method to do this is using choline hydroxide. This has a selectivity on Si versus $SiOx$ in the order of a thousand [39] . Chips fabricated using STI or silicon on isolator (SOI) techniques will show a $SiOx$ pattern which correlates with the transistor layout. The exposed $SiOx$ comes from either the insulation between transistors against cross-talk or the thin insulation layer in the transistor gate. For chips with STI, if the etching is done correctly, the chip might still work and the memory transistors have kept their correct state.

Since the bulk is often around $200\mu m$ thick it can be a very time-consuming process to remove it chemically. An alternative could be to physically mill the bulk away. This can be done till the bulk is around $10 - 5\mu m$ thick. After which a few polishing steps are needed to prepare the surface for further processing. The milling could also be followed up by chemical etching to carefully remove the remaining bulk.

The biggest problem to deal with in backside delayering is the stress that is present inside the die. Normally this is taken care of by the bulk which is the structural backbone of the die. By etching away the bulk the die becomes very fragile. Stresses that have been building up inside the die during fabrication or partially attached wire bonds can easily cause the remaining die to break. It is therefore advisable to attach the front side of the die to a replacement structure. This structure should not only support the chip but should also be able to expand and retract with the die for example when it heats

up or cools down. A small glass plate would therefore not be ideal and rather aluminium structures should be used.

Although this technique only shows the transistor layout and not the interconnect layers, it can be useful to get a quick idea of the layout of the chip, where the dense transistor areas are located and what type of transistors are used. Because this method is a little bit easier to carry out, less time-consuming and less prone to errors like over-etching it can be a good alternative option.

Cross-section

To create a cross-section a die in package is mounted on a holder using for example wax. The holder is then placed on a special kind of sander which holds the sample leveled and rotates the sample to ensure even milling. The sample itself needs to stick out of the holder enough to get to the die inside. Better results are achieved by mounting the sample slightly angled because that way it it is practically ensured that all different components like traces, vias and transistors are exposed somewhere on the cross-section. A grit of 1200 will tear through the sample sufficiently quickly while also keeping the cross-section smooth enough to avoid a final polishing step.

Now the sample can be analysed. As mentioned in the previous section, this can be done using optical microscopes. However, to get accurate measurements for determining the process node and thickness and amounts of interconnect layers it is advised to use a SEM. Again if possible EDX can be applied to get a better understanding of the different materials used.

X-ray

X-ray has the benefit of giving insight into how the die is located within the package. This can be valuable information for example when there are multiple dies in one package or if it is important to know where the wirebonds are positioned or to get a first measurement on the size of the die. This in turn can influence the process for the previously mentioned methods. Like in medical applications, the sample will become see-through, making the positioning inside the package visible albeit in a grey scale. More detailed information about the die is not possible using X-ray because Si does scatter the X-rays making it impossible to see through.

Infra red

Before the IR image can be made, first the backside needs to be exposed. Depending on the packaging numerous methods are possible, one of which is described in the previous section 2.2.1. Once the backside is exposed it is a matter of applying NIR waves ranging from about $350nm - 1700nm$, where higher wavelengths will see through thicker Si but have lower resolution. A special camera needs to be used to detect the reflected waves and create an image. This will be in greyscale but can be used to determine and pinpoint specific modules on the die. Because this is very quick, non-invasive and doesn't involve any chemicals or large machines it is an ideal step to do quickly to check if the chip is still intact and oriented correctly. Furthermore, depending on the thickness of the silicon the absorption and reflection of certain NIR frequencies changes [40]. This can be used to determine the remaining thickness of the silicon bulk.

2.2.2. Finding points of interest

Once the layout of the chip is known and the larger modules have been identified it is possible to determine points of interest on where to make a hole. For this project three areas were indicated to be explored; a connection pad, a remote signal or power line and a flash line. These areas were chosen because of the following reasons.

First of all, they range from easy to access to very hard to access. This is determined by the physical size and area of the metal strip that needs to be accessed and by the amount of structures near the line. For a connection pad, the area is very large and there are no transistors nearby. Furthermore, making the hole too deep is also not a big problem since the multiple metal layers are already connected with vias, see figure 2.5b. For the flash however the interconnect lines are thin, there are a lot of transistors around and the metal layer needs to be hit at exactly the right height.

Secondly, the signal driving strength will most likely decrease further into the chip's intricate parts. For example, the VCC connection pad will have a very high on or off signal, a signal line in the middle of the chip will have less driving strength and a word line in flash even less so. This means not only larger constraints on the connections to the metal line but also for the probe on the outside of the chip.

Thirdly, the signal information density or frequency range increases with each area. The frequency on the connection pad will be DC but further into the chip, this can increase to tens of MHz. This means higher constraints on the probe and the connections. The capacitance and more important inductance of the connections needs to be kept as low as possible to accommodate the data transfer speed. This in turn means good connection between the pillars, the chip and the probe and reducing the distance between the probe and the interconnect on the chip as much as possible.

Finally, these areas create a good roadmap for experimenting. It starts with an easy-to-verify basic principle where the proof of concept is the most important goal. If that is successful, the next steps can be undertaken to optimize the techniques and try to create more precise holes. Ideally, in the end, an attempt to read the most intricate signals can be made. The flash but also other permanent memory modules or clock signals are very useful for device forensics. A positive result for the final area would mean possibilities for direct implementation.

2.2.3. Aligning chip

As mentioned in the introduction of this chapter, the chips have to be re-aligned in the machine that is capable of creating a hole. This becomes significantly more difficult because those machines are often designed to look from the top side which in this case is still covered by the package. Furthermore, they generally look at chips using an electron beam which is not capable of seeing through silicon. Fortunately, there are plenty of solutions available. The next sections will go through the most viable options and discuss some of the advantages and disadvantages.

Near infra red

The main principle of using NIR is discussed in section 2.2.1. This technique becomes very useful in this case because it not only shows the interconnects but if focused differently can also show the top of the bulk layer. This layer will contain traces from removing the package and thinning and polishing the bulk. These traces create unique patterns which can be seen in the other machines as well. By making multiple pictures at different zoom and focus levels the correct location can be found quickly. This is a very quick and robust method. Its main limitation however is precision. This is determined by both the IR setup (camera resolution, IR frequency range and working distance) and the amount and size of the traces.

Flip-chip

Flip-chip technology has been developed as an alternative to wire bonding. Section 2.1.1 briefly explained how flip-chip is used to mount a chip onto a circuit board. The most important message is that this technique flips the chip, remembers its relative position and is able to do this very accurately. This means that the chip can be aligned in the normal way by looking at all the interconnects and features on top. As a result both aligning and making a hole can be done in the same machine thereby simplifying the process, reducing the risk of contamination or breaking the chip and improving the accuracy. However, this also means that the package is completely removed which comes with its challenges like the structural stability once the die comes out of the machine or the lack of wire bonds and therefore connection pads which are needed to activate the chip.

Looking windows and scratching

These techniques are a little less sophisticated but should not be underestimated. As processing the chip consists of many different steps which all contain some variability, each sample will be slightly different. Creating small holes at locations that are known to be redundant not only leads to indicators to align the chip but also gives information on etch speeds and thickness of for example silicon. This information can be used to optimize the steps in locations that are more intricate.

Energy Dispersive X-Ray

Energy dispersive X-ray spectroscopy is a technique that makes use of the unique atomic structure of elements. By exposing a sample to an electron beam or x-ray the electrons of the atoms get excited. When electrons fall back they emit x-rays whose energy level correlates to the unique structures. These X-rays can be detected and used to determine the composition of the sample. The deeper the electrons or x-rays go into the sample, the more they get dispersed. Depending on the sample materials and the energy (mostly expressed in volts) of the source different depths can be reached. Typically for Si at

$30Kv$, materials can be detected till $5\mu m$ under the surface. Applying this technique to the alignment problem in theory means that vias or the bottom metal layer less than $5\mu m$ below the surface would be visible at a resolution good enough to make a hole on top of one of them. However, in practice, the accuracy of detecting different elements at lower depths is unreliable and contains a lot of noise. This is mainly because fewer source electrons or x-rays are able to penetrate the lower regions, meaning fewer excitations happen, resulting in a more vague image.

Electron Beam Induced Current

electron beam induced current (EBIC) among other things is used for fault analysis[41]. The first step of EBIC analysis is to use an electron beam and expose a sample to the electrons. This results in current flows within the sample. Based on the composition of the material and the presence of built-in electric fields electron-hole pairs might be created or recombination could happen. This difference in the flow of minority carriers is being detected and can be combined into an image containing the structural differences within the chip. Like the other techniques, this can then be used to align the sample correctly.

2.3. Contamination

Since this procedure is almost the inverse of making a semiconductor chip an important aspect is contamination. Normally, this is mostly about protecting your sample. However, while this is still true, in this case protecting the environment will be even more important. This short chapter will discuss what to watch out for and which precautions might need to be considered.

2.3.1. Risk analysis

The main risk to be accounted for in this project is contamination. Contamination is generally defined as making something impure by pollution. This goes two ways, the sample or processing step may include some materials that can pollute the environment. On the other side, the environment can contain materials that will disrupt the process or change material properties such that it doesn't function anymore. To take any countermeasures, first information needs to be acquired on what is being used. For this project, this is done using some techniques described in previous chapters like EDX and delayering in combination with optical telescopes. Figure 2.16 shows the result in the form of a rough cross-section with all different materials indicated. Also, some process-defining parameters like etchants, temperature range and aspect ratio are mentioned. This is then used as a baseline to think about all the materials that might get exposed, what their effects are and what measures should be taken.

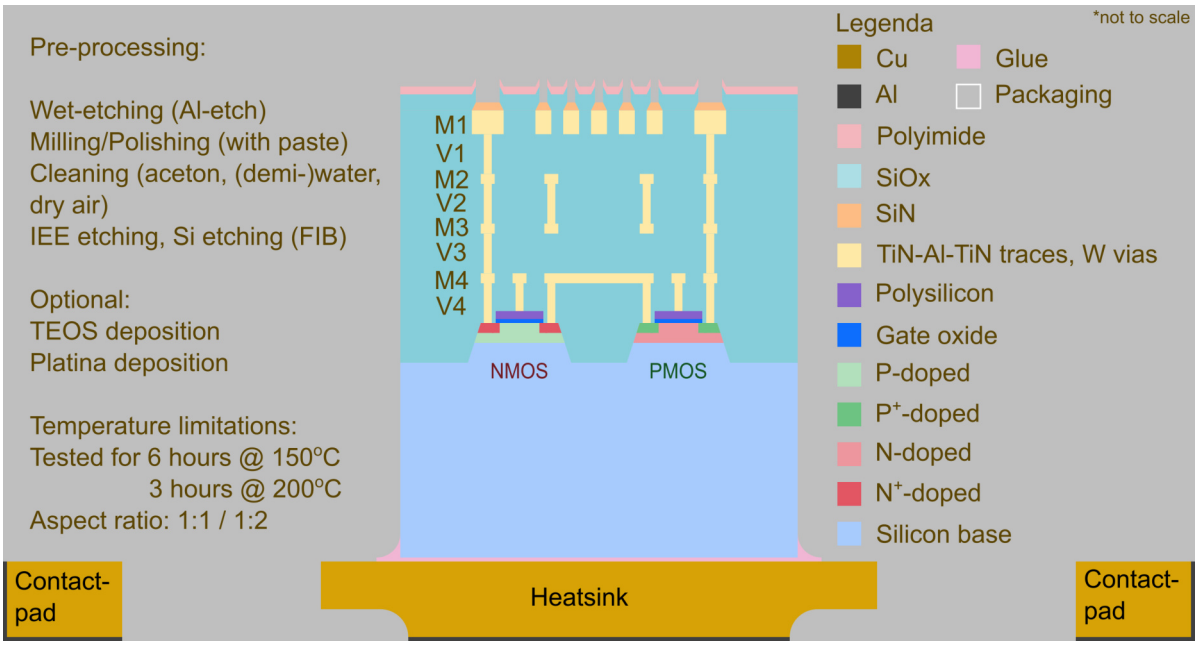


Figure 2.16: Overview materials used inside package

2.3.2. Prevention

As contamination goes in two ways, so does the prevention. Let's start by protecting the sample. For this project, the feature sizes are still relatively big. Moreover, there is no dependency on very pure crystal structures or materials. It is therefore enough to keep the sample away from dust and clean it with for example acetone. The more delicate process steps will take place in high vacuums which already take care of most contamination risks. Between the most critical steps, the sample is kept in a closed package as much as possible and only opened inside cleanrooms (CRs) to reduce the risk of dust even more. CRs are specially conditioned labs in which particles are minimized. They are categorized based on the amount of particles larger than $0.5\mu m$ per ft^3. For this project CRs of class 10.000 and 100 are used.

As mentioned in the introduction, preventing the environment is a much bigger concern during this project. The environment can be roughly split up into three aspects. First of all, it is important that you are well protected. Cleanrooms have many protocols to ensure the safe handling and processing of chemicals. It is nevertheless always important to read the material safety datasheet (MSDS)) of all chemicals that are involved in the process step. On top of that exist the so-called III-V materials. These are chemicals which consist of at least one group III and at least one group V element. The group III elements are B, Al, Ga and In and the group V elements are N, P, As and Sb. It is suspected that these compounds are carcinogenic and should therefore be treated with extra caution.

Secondly, preventing contamination of the tools and machines that are being used. For this, it is very important to know that some materials, mostly metals can be very diffusive at relatively low temperatures. In this case, the main concerns are the copper and aluminium in the contact pads and heatsink. Copper has a thermal diffusivity of $111mm^2/s$ and aluminium around $97mm^2/s$ at $25°C$ [42]. Because samples will be traveling between labs it is also very important to clean them well before transporting them to avoid contamination between labs. Machines are used for many research projects with various degrees of contamination stringency. This means that given the presence of the Al and Cu, it's not allowed to use some machines and the ones that are allowed need to be cleaned thoroughly.

Finally, it is important to consider what to do with the waste after processing. Most labs have equipment and regulations to deal with most substances. In general, it is not allowed to throw away any organic material since this is very harmful to the outside environment and nature. Other chemicals need to be carefully diluted.

3

Recipe design

This chapter discusses all the relevant aspects and possibilities for creating a recipe. The recipe that will be developed for this project is for the connection of backside contacts. Apart from this thesis, this concept has not been executed in any way or form. Therefore, all options had to be considered and plenty of iterations had to be made.

By using the knowledge of chapter 2, a general outline of the main concept can be formulated. This concept was then developed into a recipe during which a practical hierarchical structure emerged. At the top sits the recipe itself. The recipe can be divided into phases which form the second level. These phases are divided into processes which make up the third level. Finally, each process consists of steps resulting in the bottom level. Figure 3.1 further illustrates this structure.

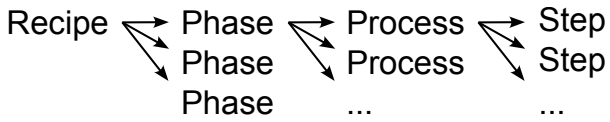


Figure 3.1: Hierarchy as applied to recipe design

This chapter starts by discussing the general outline and some observations that shaped the design process. Next, the different phases and their processes are discussed. The final section combines all the new information which gives an overview of the whole recipe and results in figure 3.4.

3.1. Outline

The first part of this section aims to introduce the main concept for implementing backside contacting. The second part provides some context in which the project was carried out. As there is no reference material for this concept, the development process can grow out of scope quickly. Therefore some thought was given to the approach, expectations and limitations beforehand.

3.1.1. Main concept

The goal is to connect to the bottom interconnects of a chip and the intended approach is to implement backside contacts which already give a rough idea of how the recipe will look. As a start a general concept was created which contains the basic steps and ideas. This concept is discussed with the help of figure 3.2.

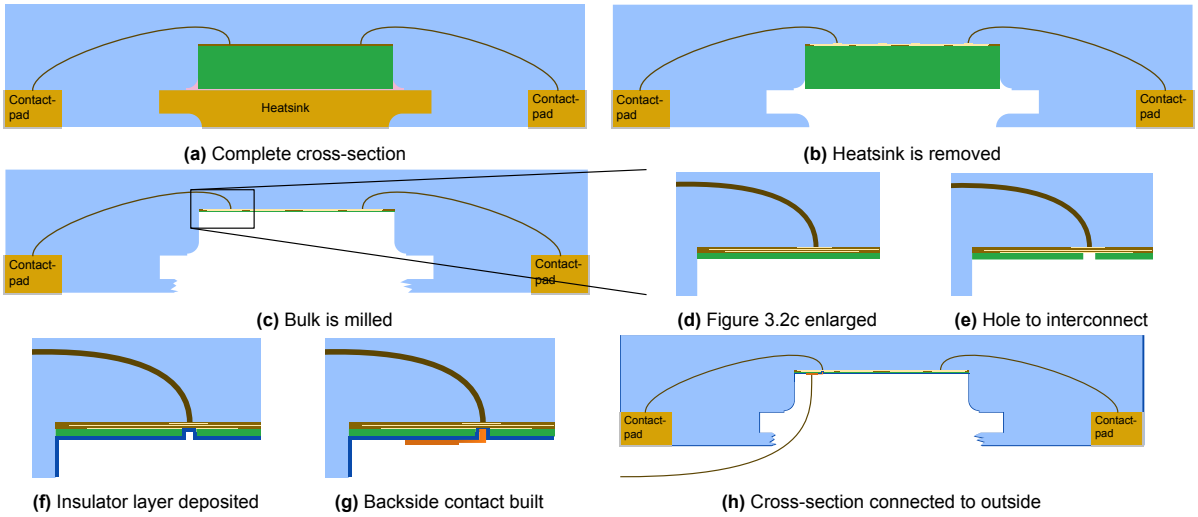


Figure 3.2: Cross-sections of the main concept

The starting point is an unknown chip, new from the factory. From this, the first step is to expose the backside of the die. Once this is done, a hole can be made and a pillar constructed, however, at this point, aspect ratios of around $1:200$ are found which are too high, especially for constructing the pillar. Therefore, the bulk should be milled until a thin slice of less than $10\mu m$ remains. In this slice, a small hole can be made, ideally fitting in an area without transistors. Before the pillar is made, a thin insulation layer should be deposited so that the bulk stays insulated from the pillar and the interconnect. Then this layer needs to be removed at the bottom of the hole after which the pillar can be constructed. Now, the connection to the bottom interconnect has been made.

What is left is making a connection to the outside to be able to read the interconnect. The proposed idea is to create a large pad on top of the bulk and in the vicinity of the hole. The pillar is then connected to this pad with a small trace. The pad can be used to connect an outside probe with for example a wire-bond. The pad is still insulated from the bulk due to the insulation layer from earlier. Using this approach it should be possible to get access to, connect to and read signals from the bottom interconnects.

3.1.2. Early observations

To prevent the project from reaching an unnecessary amount of experiments and getting sidetracked the following three observations were made. Together these observations provide the structure in which the recipe was developed.

First, it was observed that the recipe can be split up into three phases that can be partially carried out in parallel. These phases are the acquiring data phase, the creation phase, and the bonding phase. Each phase depends on the other two but in a non-restrictive manner. This improves the efficiency of the design process as time-consuming tasks will not block the entire development flow. Furthermore, iterating over only a phase or part of a phase is faster, which leads to the next observation, the importance of testing.

To ensure the best processes and steps are selected several tests must be implemented which can quantitatively compare different processes. However, as will become apparent further into the design, each process step comes with significant deviations due to the amount of manual labour. Combined with a limited availability of test samples, this meant that testing and interpreting the results was not trivial. Therefore, testing ranged from numerical comparisons to visual observations.

Finally, it was observed that the availability of machines and other equipment like chemicals or deposition materials influences the feasibility of the recipe. This project was done in collaboration with the NFI and the TU Delft which meant access to their facilities and labs. Although the goal is to provide a proof of concept and a reliable recipe, this limitation meant that not all options could be carried out or tested and better alternatives might exist. However, it should be noted that the NFI would like to be able to run the recipe in-house, thereby giving enough reason to try and use only the NFI labs when possible.

3.2. Acquisition of chip layout and technology

The acquisition of data is the first phase in designing the recipe. The goal of this phase was to get a better understanding of the physical properties of the chip and to determine which areas are suitable for backside contacts. This was done by gathering as much data as possible with the techniques described in section 2.2.

For this recipe, it was found that the data acquisition can be split up into four independent processes; outside observation, obtaining cross-section, extracting the die and exposing the backside. These processes will be discussed in the next sections where each section will be a process and each sub-section will be a step of that process. There is no particular order in which these processes should be done, however, doing the outside observation first helped with understanding the results of the other processes and creating the crosssection helped with extracting the backside.

3.2.1. Non-destructive observation

This process was the first part of the data acquisition and therefore not really a process as it was more about collecting general information. This process aimed to obtain a specific set of information which was used as the basis for the rest of the recipe. This set of information was constructed by iterating over the phases and steps and will be discussed next.

The first piece of information is the pin layout of the chip. This information was used to align the chip the same way every time. This sounds quite trivial however the chip was rotated and flipped a lot during processing so it was difficult to keep track of its orientation. It was therefore also helpful to create an upside-down pin layout since for backside contacting the chip will be accessed like this a lot. The pin layout was also used to build several test setups, the first being to test the chip's functionality before it entered any processing.

Secondly, an X-ray scan was performed. The goal of this step was to get a first impression of what is inside the package. This means information about how many dies, how they are stacked if there is more than one, what size they have and where they are roughly located. Furthermore, it is useful to know how they are connected to the outside. For example, for this project, the die was wire-bonded and thus it was important to find out how these wires travel inside the package and where the bonding pads were situated on the die. This information was used to confirm the results of the cross-section, to avoid hitting the wire bonds when milling and to get a first impression of the suitability of accessing the bond pads.

Finally, some general information was noted. This included the size of the package, the type and material of the output pins and the presence of a heat sink. This information was mainly used for making test setups although knowing the height of the chip is often required when working with load-locks or other machines that make use of vacuum chambers.

3.2.2. Cross-section fabrication

This is a fairly straightforward process of which the practical part is already described in section 2.2.1. This process was done in an hour or two and provided important parameters to tune other processes. First, the shape and size of the heatsink were observed and measured. In addition, a thin layer of glue should be found as well. This will be important in the next process. Next, the thickness of the bulk could be determined. This will be verified at a later stage but it gives a good indication of how much to mill and of how strong the die is. After that, a closer look was taken at the interconnects. The amount of metal layers can be found as well as the size and thickness of these layers. Furthermore, some of the vias were also visible. Together an estimate was made on the technology node of the die. This in turn gives some insight into the materials that might be used and some of the feature sizes that can be expected. Finally, with the use of EDX some of the materials used could be determined.

3.2.3. Die extraction

By extracting the die, the goal is to do measurements on the front- and backside of the die. Again the process is already described in section 2.2.1. It can be summarized as a quick nitric acid etch to remove the package and possibly a heatsink. This was done twice to have samples for both a front-side and back-side analysis.

For the front side a repetition of Alpad and Al etchant is used to expose the interconnects layer by layer. Alpad etches $SiOx$ and consists of ammonium fluoride and acetic acid while Al etchant etches

Al and consists of phosphoric acid and nitric acid. Another option is to start with RIE, although it is harder to get a uniform etch across the die. The results from front-side analysis are used in two ways. First of all to determine the layout of the die meaning the locations of memory blocks, the processing unit(s) and other typical modules like analog circuitry. Secondly, a more detailed analysis was done on specific blocks. Information on how they are wired, where the data traces lie and how certain elements are structured was obtained. This information is used to determine where to create a backside contact.

For the backside, the bulk was first milled after which choline was used to remove the remaining Si. The results from the backside mainly show the position of all the transistors. This can support earlier founding from the front side and indicate possible places for a hole.

3.2.4. Exposing backside

The final process in this phase is the exposing of the backside. This process should leave the die alive because it will also serve as the beginning of the next phase. That in turn means no room for error, precise processing and many iterations to improve the steps. The process can be divided into three steps, namely the removal of the package or heatsink, the removal of the glue and finally thinning of the bulk. The next sections will discuss them in more detail.

Removal of package or heatsink

As the title already indicates, this step has two possibilities depending on the packaging. Before going into detail of either option it is important to look at the whole process. From this, it can be observed that the final step, the thinning of the bulk, requires a flat and preferably intact bulk. This means that chemicals or physical removal of material in the previous steps should not interact with the Si bulk. Moreover, by looking at the whole recipe it can be seen that the output pins should also remain functional. This is important because they are located on the backside of the package and are therefore exposed to the processes. Having made these observations an approach can be designed for either possibility.

In the first case, there is no heatsink. This is the more challenging case of the two because during the packaging it is not ensured that the backside of the die is completely parallel with the backside of the package. To mill a hole would therefore require to level the package based on the die within which is not possible. Another approach would be to carefully etch away the package with a few drops of nitric acid. As the bulk has a very slow etch rate compared to the package, this etching will practically stop at the bulk, leaving it exposed. The problem however lies in controlling the sideways etch and preserving the output pins. A combination of milling, applying protection and careful acid etching could result in an exposed bulk. For this project, however, all the samples had a heatsink meaning this method was not developed further.

In the second case, the package has a heatsink. Using the earlier mentioned processes the shape and materials of the heatsink were determined. Furthermore, a thin layer of glue was found between the heatsink and the die. The heatsink could then be removed with Al etch, which is meant for aluminium etching but also etches copper quite fast. It does not however etch the package or the glue or bulk, thus leaving a perfectly exposed backside. One problem, however, is the shape of the heatsink, especially at the etches. This shape is shown in figure 2.3 which indicates that the heatsink is wider underneath the package. Removing the heatsink creates a cavity under the package which can become sufficiently deep so that the etchant no longer reaches the remaining copper. This resulted in contamination and might have caused short-circuiting due to the high thermal diffusion rate of copper, see section 2.3. Furthermore, copper is a relatively soft material which caused the mill to clog up which resulted in unpredictable milling depths. It is therefore quite important to remove all the copper. Apart from longer etching, there is no fast and practical way to reach the deep insides of the cavity. The best option is to either mill the overhanging packaging before etching or etch most of the heatsink, then mill the overhang after which a second etch is done. The etching itself can also be improved by making sure the etchant is in constant movement using for example a magnetic stirrer. This was not the case for this project which resulted in uneven etching and several replenishments. Finally, the Al etchant will also remove the output pins which means they have to be carefully shielded. A quick and easily appliable and removable material is used in dentistry to create imprints of the teeth. This material is a type of silicone which hardens quickly, has the adhesion of tape, does not react with many chemicals and is safe to use.

Removal of the die-attach

This step can be skipped if there is no heatsink but in case the heatsink is removed a thin layer of glue is left behind. Several ways were explored to remove the glue as it is assumed to be non-uniform and could influence the milling step. The glue itself was observed to be hardened, which made removing it quite difficult. Furthermore, as different batches were used, different types of glue were found. This is a common phenomenon as it is generally considered irrelevant which glue is used and therefore each manufacturer uses their glue.

A first attempt to remove the glue was made using soft polishing. The idea is that the glue is softer than the silicon and would therefore be removed without damaging the bulk too much. This was indeed the case although minor scratches were found and the process itself was quite time-consuming. In order to speed up the process a manual approach was used in which cotton swaps were used in combination with acetone to dissolve the glue. This turned out to be the preferred method because of its speed and preservation of the bulk. However, some of the glues encountered would not resolve into the acetone. In an effort to solve this problem brief etch with a droplet of nitric acid was used. This again was quite effective but with the side effect of being quite time-sensitive and messy. The latter is because the acid also etches the packaging material at a high rate. Finally, there is the possibility to use RIE, however, this was not tested because it would most likely not improve the already used methods.

Once the glue is removed a shiny, smooth Si bulk is visible. At this point, some tests were conducted. Firstly the functionality of the chip was tested. Secondly, a IR measurement was done to determine and confirm the thickness of the bulk. At this point, the bulk is too thick to be able to look through it with a sufficient resolution.

Removal of the Si bulk

The removal of the bulk is the first step which introduced a lot of divergence between the samples. As will become clear later, the thickness of the remaining bulk should be $5\mu m$ or less. Otherwise, the walls of the holes are too high, preventing the flow of gasses inside the hole and reducing the accuracy of the milling depth. Directly on top of the bulk (in this case directly underneath) are the transistors, meaning a slight error could kill the chip. Two options were considered to remove the bulk.

The first option was RIE because, with a slow etch rate, the final thickness of the bulk could be well-defined. However, earlier tests showed that for flat chips, the etch rate was not uniform across the whole chip and fluctuated in general. As the bulk is already surrounded by a wall of the packaging this nonuniformity would only get worse. Therefore the second option was chosen and optimized.

This second option makes use of a mechanical mill. The milling machine is able to level the chip within $0.5\mu m$ accuracy and can thin the bulk with around $3 - 4\mu m$ accuracy. To achieve the desired result many iterations were needed between milling and measuring the thickness. This is because the accuracy of the milling height goes down over time as the milling bit wears down and leftover copper parts clog up the bit. With some experience and many thickness measurements, this can be compensated and a thin backside can be achieved.

By now, the backside is thin enough but far from smooth. This leads to inaccurate measurements regarding thickness, lots of distortion in IR imagery and non-uniform processes further on. Therefore a polishing step is required to flatten the surface. This step can be done in the same machine and by moving from rough to fine paste a flat surface can be created. The polishing however does leave some minor scratches and visual stains. These marks form a unique pattern across the chip which is helpful to navigate across the surface and provides reference points for later processes. Furthermore, it is important to emphasise that the thickness and smoothness achieved in this step greatly influence the quality, predictability and therefore also yield of the next processes.

3.2.5. Determine areas of interest

The final process is to combine all the acquired data and determine which locations on the die are suitable for a backside contact.

Referring back to section 2.2.2, three areas of interest were determined; a connection pad, a remote signal or power line and a flash data line. The information found in the previous processes is combined to create a complete picture of the package and die. Using a combination of IR images, SEM and optical images, a clear overview of the different modules and the layout within the die can be formed. This can then be used to accurately pinpoint different areas that are suitable for a backside contact.

A possible approach would be to look at the layout first and determine the different modules and their functionality. Based on that, a decision could be made on which areas are suitable for a backside contact. The next step would be to zoom in on those areas and create a more detailed view of the structures and interconnects. This leads to a final precise point where the interconnect can be made. Finally, new images using IR or the SEM could be made to verify the findings.

In conclusion, at the end of this phase a detailed understanding of the most important parts of the die is aquired. This results in a range of possible places for backside contacts, each with their advantages and disadvantages. On top of that, the chip has been prepared for the next phase by thinning the bulk.

3.3. Invasive processing

The second phase of the recipe is the creation phase. In this phase, the backside contact is built in three sequential processes: the creation of a hole, the deposition of an insulation material and the creation of pillars. The goal of this phase is to create a via that is visible from the outside and connects to the interconnect on the inside. The next sections will describe the different processes in more detail.

3.3.1. Create holes in the Si bulk

The input of this process comes in two parts; an exposed and thinned backside of a die and the information of one or more areas where to create the holes. To ensure the holes are created in the right place the chip needs to be aligned correctly. Section 2.2.3 discussed the available options already. For this project only NIR was used because it meets the required accuracy, is able to create a link between the interconnects and the pattern on the outside, is fast in both setup and usage and is readily available. Once these images are made the sample is put into the SEM-FIB. This is one machine that can do both electron imaging and etching using an FIB.

The sample is then aligned using the SEM and the IR images. The FIB is linked to the SEM which means a hole can be made in the area indicated by the SEM. The hole is made using the FIB which makes use of a gallium source. The source ions are accelerated to the surface where they knock off the surface atoms, thereby milling the surface. In figure 3.4, this process is referred to as silicon etch. For large areas, however, this process is very slow. To speed it up, an insulator enhanced etch (IEE) can be used in which the chamber is filled with XeF_2 (Xenon difluoride) gas. This speeds up the process by roughly twenty times. This speed comes at the cost of precision and quality as the resulting hole is very rough around the edges and rarely uniform. This means that the optimal result can be achieved by combining both techniques.

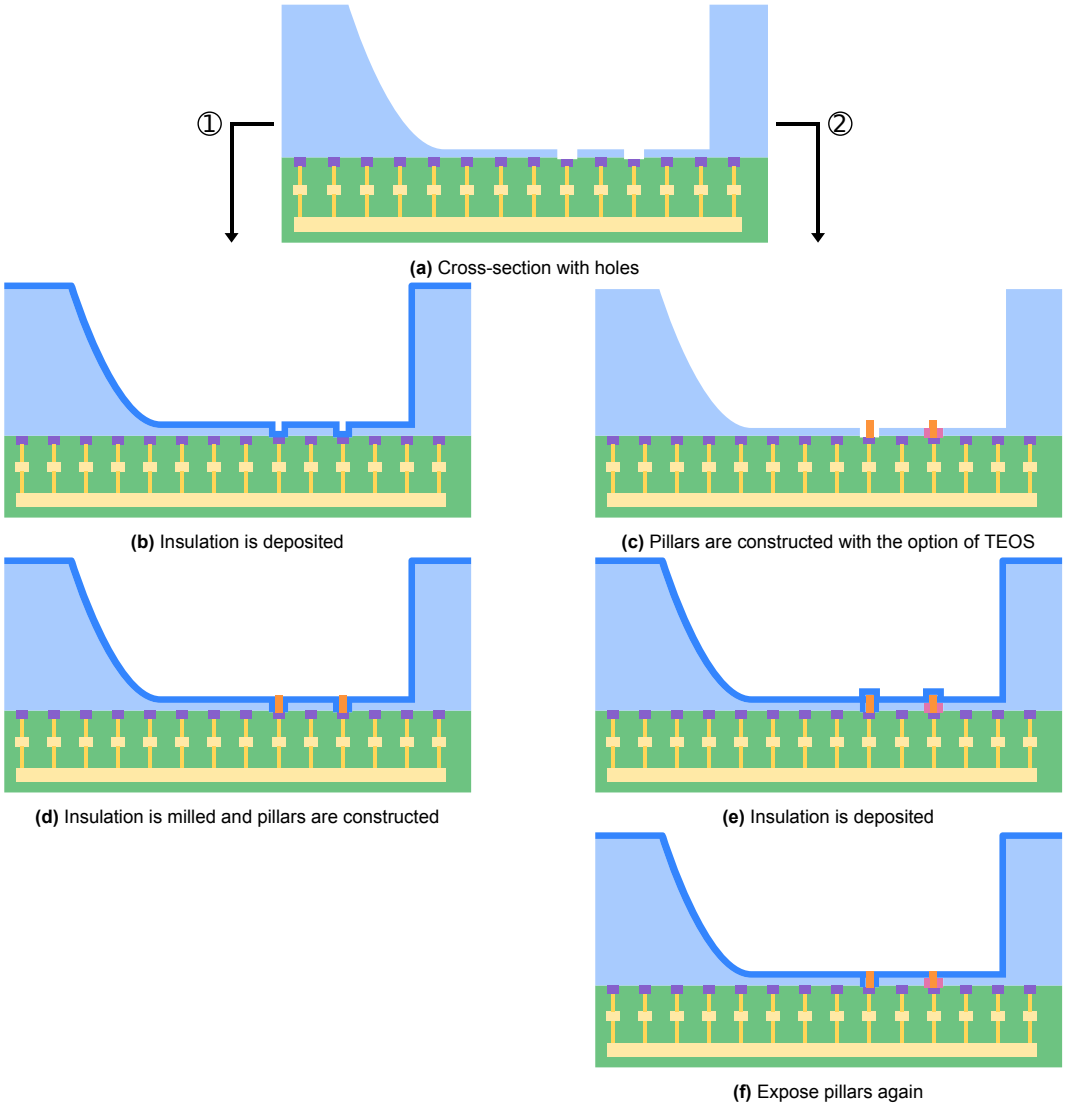


Figure 3.3: Example crosssections of the different paths

Once the holes are made, two different paths are possible depending on the order of the processes. Figure 3.3 shows both paths with option one in the left column and option two in the right column. These paths are also indicated by the numbers one and two in figure 3.4.

In the first path, the insulator is applied directly after the holes are made. This process is done in a cleanroom but more importantly in another lab. That means that the samples are taken out of the vacuum chamber, transported, coated with an insulating material, transported back and put back into the vacuum chamber. These transportation steps bring some contamination risks, like oxidation or dust as the interconnects will be exposed to air. After the chip is coated and placed back in the SEM, the interconnect is covered again by the insulation film. This needs to be removed after which the pillar can be built.

The second path goes to these processes in reverse order. At the start, the sample is still in the SEM-FIB which means the pillar can be constructed immediately. The SEM-FIB is a tool which combines both the SEM and the FIB using only one vacuum chamber. After the construction, it is possible to add some insulation around the pillar in the form of TEOS. The difference is shown in figure 3.3c where one hole has the TEOS; indicated in red. The sample is then taken out of the SEM-FIB and brought to the clean room for the deposition of an insulation layer. Once this is done the sample is put back into the SEM-FIB to find and expose the pillar again. Doing the steps in this order prevents the interconnect and the pillar if TEOS is applied from exposure to air and other contaminants. However, constructing the pillar while the thin bulk is unprotected might lead to diffusion of metal into the bulk, resulting in

decreased resistance and undesired behaviour.

3.3.2. Create pillars

Depending on the chosen path, first the sample needs to be aligned again and the insulation material needs to be removed at the bottom of the hole. After that, a choice has to be made regarding the pillar material. There are two options; platina and tungsten. Platina is softer, sticks better to the surface and is, therefore, easier to handle. Tungsten on the other hand is harder, is more complicated to apply but has lower resistance. In practice, platina was used because the conduction of the sample was low. This causes charging effects during the deposition which are much higher for tungsten as it requires higher currents to deposit. The die charges because it is partially in the package and it is hard to sufficiently connect the bulk to ground. One of the side effects of charging is a shift in the live image and working area. This effect is even greater once the insulation deposition has been done. These distortions make it challenging to form clear images and straight pillars.

The pillars are created layer by layer. This is a relatively fast process in which, depending on the size, roughly every 5 minutes $1\mu m$ of height is added. Furthermore, this process is able to achieve high aspect ratios. This is needed to ensure that the pillar is still visible after the TEOS and the insulation layer are added.

3.3.3. Deposite insulation

There are two processes available which can do the deposition of the insulation layer within the given temperature budget and given the contamination risks. These processes are ICPECVD and ALD and are discussed in section 2.1.1. Apart from the time and process constraints, it is also important to note the difference in the end products. ALD is designed for thin layers of a few nm's. Furthermore, ALD has a very high conformity which means that the steep walls of the hole are nicely covered with insulating material. ICPECVD on the other hand quickly deposits layers of a few hundred nm's, an order of magnitude thicker. Both results have sufficient insulation to prevent current flow.

The end product ideally has a smooth surface in which the holes or pillars are easily recognisable. The smoothness would help with the last phase in which a connection is made to the probe on the outside. To realign the chip in the SEM-FIB it is practical to have clear visible markers. This used to be the pattern originating from the polishing step. In this lies the trade-off between ALD and ICPECVD. With the high conformity and very thin film of ALD those patterns and the holes should still be visible. However, the surface is still quite rough and the edges of the hole are still steep. This makes it harder to deposit a trace and connect to the pillar inside the hole. The thicker film of ICPECVD helps smoothen these sharp etches but thereby also covers the pillar or makes it harder to find the exact spot on where to make the pillar.

Finally, these machines are kept under strict supervision in a cleanroom. This means these processes need to be planned carefully to ensure all requirements for access to the cleanroom are met and the right supervision is present. On top of that, these machines are designed to handle wafers. A carrier wafer should therefore be present. Fortunately, this also means multiple samples can be processed at once, meaning the administrative time can be reduced significantly by having multiple samples ready.

3.4. Establish external connection

The final phase is the connection phase. The goal of this phase is to connect the pillars to a probe on the outside. This enables the readout of the signals that are present on the interconnect thereby completing the goal of this project. To make this happen, pads, traces and bondwires are needed.

3.4.1. Create pads and traces

First of all, the bond wires need some place to bond to. The pillars are too fragile and small to bond to. This means each pillar needs a special pad, preferably nearby. This pad is then connected to the pillar with a small metal trace and used to bond the bondwire. As the sample is still in the SEM-FIB no additional aligning needs to take place and both the pad and trace can be deposited in the same process. This process is similar to the creation of the pillar which means that again a choice can be made between platina and tungsten. Once these features are deposited, the sample can be taken out of the SEM-FIB and placed in the probe for its final process.

3.4.2. Wirebonding

The final process is wirebonding for which specific equipment exists. There are four things to keep in mind before starting the wirebond. First, the materials on which the bonding happens should not be too soft. The bonding is made possible with the use of ultrasonic vibrations and pressure which should be absorbed by the wire and not the surface. Furthermore, the size of the bonding areas should be sufficiently large. This depends on the thickness of the wirebond and the resolution of the camera or microscope used in the bonding device. Finally, the space around the bonding area should be clear of high obstacles as the bonding needle needs a small space to move around. During testing, it was found that the path the wire needs to travel has no significant restrictions regarding height, distance and curvature. That being said, both bonding areas should be horizontal.

With that in mind, the wire bonds can be placed. The connections are made on one side to the deposited connection pad and on the other side to a specially designed probe. This probe is then able to buffer and amplify the signals for further processing.

Once all the wirebonds are placed the product is finished and measurements can be taken. To measure multiple samples, either multiple probes have to be made or the wirebonds have to be broken and the sample replaced. As these probes are not cheap. the latter option will likely be chosen. However, with careful handling of the samples, it should be possible to bond a wire multiple times in the same area.

3.5. Conclusion

The recipe has been summarized in figure 3.4. The flowchart can roughly be read from top to bottom and left to right, with the start and end indicated in red. The different processes are indicated in yellow and in green are the steps. Finally, the different phases are shown in different background colours.

This chart aims to clarify the possible routes that can be taken, the different options that can be explored and the parallel sections that are present. It also shows some repetition which indicates important processing steps. Finally, it should be clear that the larger part of the recipe is sequential, meaning every step depends on the quality and success of the previous steps. Thereby enforcing the importance of quality control, especially in the early steps and processes.

What is missing from the flowchart are the testing steps. These should occur at least after every process and where possible after each step. Depending on the state of the sample, different tests can be carried out. The different test setups are discussed in detail in section 4.2.

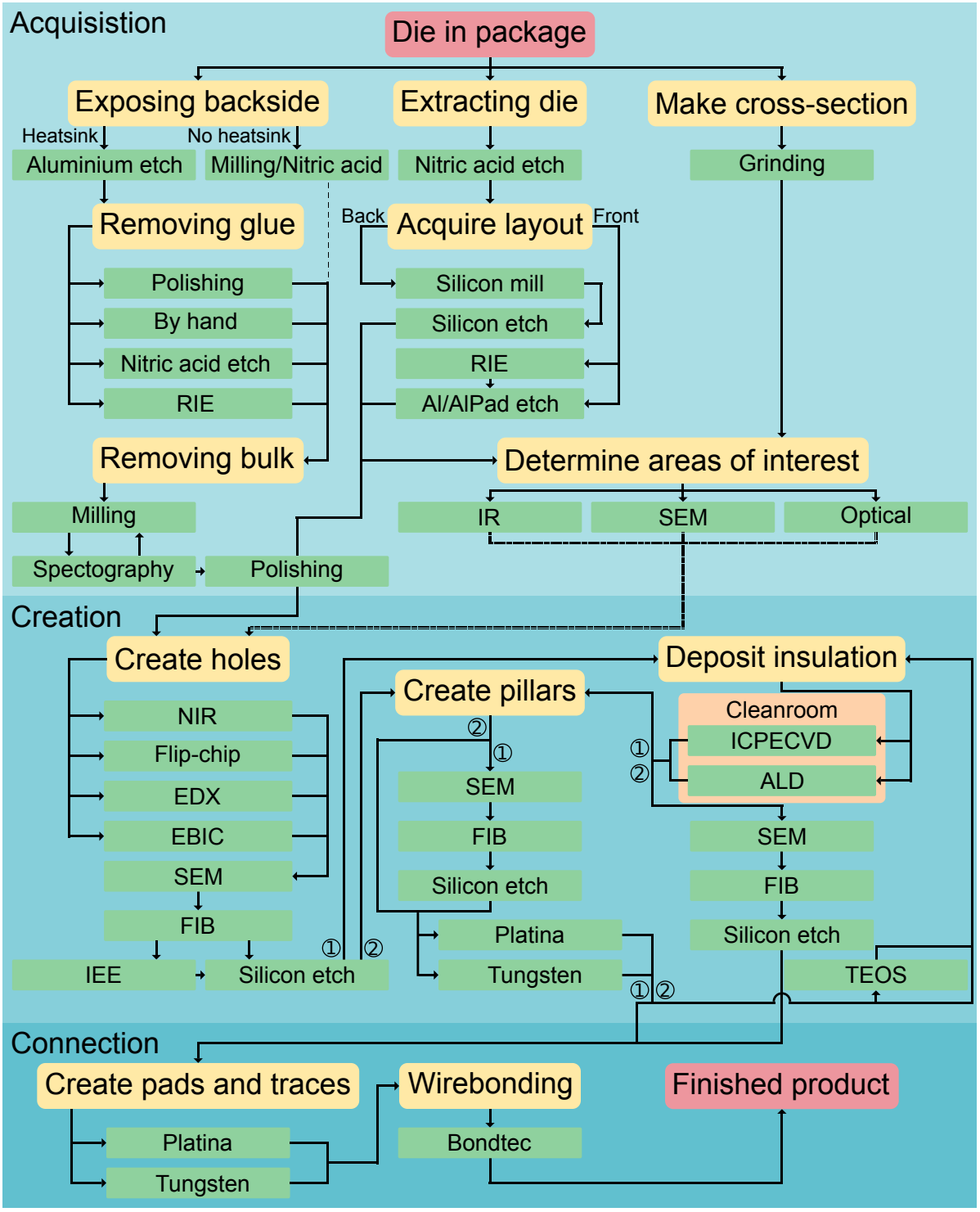


Figure 3.4: Recipe flowchart

4

Experiments

At this point, the general concept has been converted into a recipe. However, this recipe needs to be put into practice, refined and tested. In this chapter, the equipment and different setups used to implement the recipe will be discussed.

The chapter will be split up into two parts. The first part will discuss each process of the recipe, which equipment was used, the recipe for each step and some improvements that were implemented. The second part will discuss the different test setups and where they were used within the recipe.

4.1. Recipe processes

This section will go through the whole recipe as shown in figure 3.4. Each unique step will be discussed with the aim of making it possible to recreate the process. The processes were created using a combination of equipment manuals, help from experienced users and experimental iteration.

4.1.1. Cross-section fabrication

To make a cross-section a complete chip was taken and glued onto a temporary holder. It was placed on an angle and with more than half of the chip sticking out as described in section 2.2.1. This was then mounted on the Allied Multiprep, see figure 4.1.

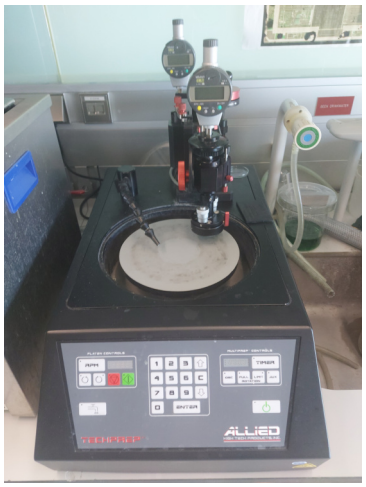


Figure 4.1: Multiprep

Two types of sandpaper were used, one with 400 grit and one with a finer 1200 grit. The first grit was used until the chip was ground down about one-third of the die. The finer grit was used for around 25 minutes to flatten the cross-section and to remove the deep scratches of the first step. At the end, the sample was cleaned using acetone and put into the SEM for imaging.

4.1.2. Die extraction

The extraction of a die from its package was done in a fume hood with nitric acid. A recipe was already developed by the people of the NFI and can be found in the appendix A. The basic setup can be seen in figure 4.2. After the nitric acid etch, the samples were put into a vacuum with some acetone to completely neutralise the acid including the acid that might have diffused into the die. Finally, the dies were cleaned one last time with acetone. The clean dies are then used in the next process.

Figure 4.2: Fume hood with equipment used

4.1.3. Acquire layout

As indicated in the flowchart, there are two ways to obtain the layout. Each gives a unique set of information as discussed previously in section 2.2.1. The next sections will discuss the steps that were taken to acquire the layout of the interconnects.

Frontside

To delayer the die from the front, the chip was mounted onto a stub using a permanent two-component glue, see figure 4.3. From this point, the process turned into trial and error because of the following reasons. First of all, the etchants that were used were quite old, meaning their etch rates were no longer well-defined. On top of that, the etch rates are temperature-dependent and since the temperature across a hot plate is not evenly distributed, the rates become even more undefined. Additionally, the etch rate along a vertical edge of two materials is higher than a flat surface of one material. Finally, the content of the die is not uniform. To summarize, there is no practical way to design a time-based recipe.

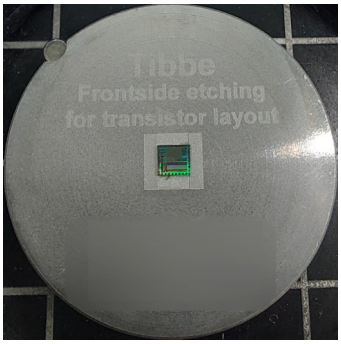


Figure 4.3: Die mounted for frontside layout analysis

Figure 4.4: Lext 3D surface-analysis tool

This means that every so often the process should be stopped and a photo should be taken to see if something has changed and to see if any layers of interest are exposed. These photos were made

using the Olympus Lext OLS5000, see figure 4.4. This is an optical microscope with an ultraviolet laser to measure the height. By combining both sensors the tool can create a 3D colour image of the surface. The tool has a moving stage on which the sample was placed. Navigation as well as zooming is done via the accompanying software. The software also provides options to stitch and do measurements on the surface heights. The 3D surface measurements are done using a $504nm$ laser and different focal points. The combination of a 3D surface and a high-resolution coloured image was incredibly useful in fine-tuning this process, acquiring valuable data for the chip layout and gaining a better understanding of the chip in general.

Besides the imaging, this process consists of two more steps; Al etch and AlPad etch. Both are done inside a fume hood on a hotplate at $45°C$. Each process can be described using a general structure.

AlPad etch Alpad etchant doesn't remove aluminium but does remove SiO_2 and SiN, albeit much slower. It is therefore important to use a ceramic cup. The Alpad etchant is placed roughly in the middle of the hotplate and preheated. When the Alpad etchant is applied, the whole stub is submerged in the etchant. This is left for around 5 minutes to remove the $SiOx$, after which the sample is taken out to verify the process and take images. When the sample is taken out, first the solution is diluted with water followed by cleaning with acetone. Since a thin layer of SiN protects the top layer, reaching the first metal layer can take a few cycles.

Al etch The Al etchant is also preheated on the hotplate. A similar process applies, however, a wall was made with dentist paste around the die to save on etchant and therefore cost. The paste itself is immune to the etchant and can be easily removed without leaving residues. Since the metal is also exposed at its edges and smaller, the recommended time between checks is only 1 minute.

In total around 20 cycles were carried out before the bottom metal layer was reached. Each cycle an image was taken. This took the better part of a day. Combining all the pictures taken resulted in a good understanding of the 3D structures within the die.

Backside

To get a fingerprint of the transistor layout from the backside, the bulk needs to be removed. The heatsink was already removed with nitric acid when the die was extracted out of the package. Removing the bulk is done in two steps. First, the majority is milled away with the Allied X-prep. This machine, shown in figure 4.5, is a milling tool which is able to very accurately remove material. The options and settings are discussed in more detail in section 4.1.6. This machine has an adapter to place stubs inside. Therefore, the chips were placed on a stub as shown in figure 4.6.

Figure 4.5: X-prep

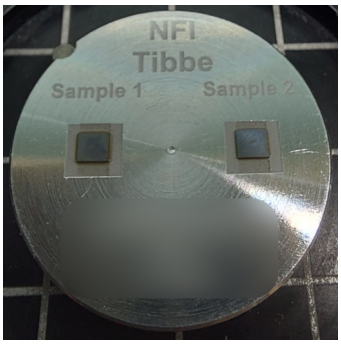


Figure 4.6: Sample setup for backside layout analysis

After the milling, a more refined process should be used to remove the remaining silicon. For this, a silicon etchant like choline hydroxide can be used which has a high selectivity for Si compared to $SiOx$. This leaves the gate oxide intact and therefore an indication of where the transistors are located.

As there were no refined recipes available to remove the remaining bulk and all necessary information about the layout was already gathered using the other methods, it was decided to skip this step.

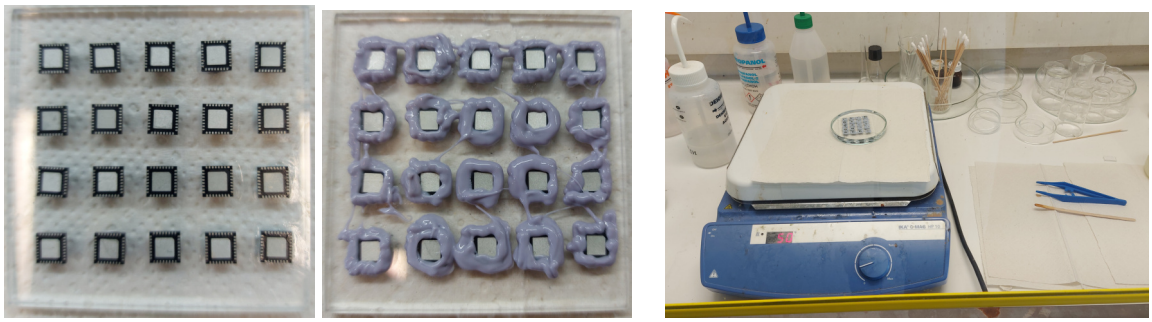
4.1.4. Exposing backside

As discussed in the previous chapter, there are two options depending on whether there is a heatsink present or not. For this project, a package with a heatsink was chosen because it provides a clear path to the ability to level the die. This is because the heatsink can be etched chemically without damaging the die or package. When there is no heatsink present the packaging is wrapped around the die to protect the backside. Removing only the packaging at the back of the die without damaging the bulk, surrounding packaging or other features requires more finesse and repetitions.

The heatsink can be removed using Al etchant, like in section 4.1.3. This etchant, consisting of phosphoric acid and nitric acid is capable of removing both aluminium and copper. Figure 4.7 shows the setup steps.

First, the sample was placed on a glass plate and glued into place using wax. This is then placed inside a Petri dish, on top of a hotplate at $50°C$ and inside a fume hood. It is important to protect the in and output pins since they are also made from aluminium and copper. This was done using dentist paste.

The etchant is transparent at the start and becomes more and more green as it reacts with the copper. In the beginning, the process was very slow as the copper was still protected by the aluminium. As the aluminium was slowly etched away, the process sped up. The etchant was replaced five times, starting with two hours in between down to half an hour before it was saturated.



(a) Samples on glass plate **(b)** Dentist pasted applied **(c)** Setup in fume hood

Figure 4.7: Setup steps for heatsink etching

4.1.5. Removing the die-attach

After the heatsink has been removed, a thin layer of hardened thermal conductive glue remains. Different methods were tried to remove this glue which are described next.

First, to remove the glue using polishing the X-prep was used. The final polishing step of section 4.1.6 was used to preserve the bulk as much as possible. That section will also provide a more detailed description of polishing using the X-prep.

Secondly, a droplet of nitric acid was used. This works similarly to the die extraction from section 4.1.2 where the recipe is slightly altered since only one drop of acid will be enough. Instead of placing the chip inside the beaker, a small amount of acid is taken out using a pipet. With the pipet in one hand and bottle of acetone in the next a drop of acid was applied. One second later the die was sprayed with acetone to immediately stop any further reactions from the acid. Referring back to the original recipe, this meant sections A.2.3 and A.2.3 were skipped and the final cleaning step could be carried out thereafter.

A third option would be to use the RIE tool which in theory could also remove the glue. However, in practice, it was found that this would be rather time-consuming because of the required setup end processing time and unnecessarily complex. The methods mentioned above were already found to be effective and faster. Moreover, the final method proved to be even better. That being said, if the glue couldn't be removed using the other methods, this is still a viable option.

The final method was using a cotton swap and some acetone to manually remove the glue. This was done inside a fume hood and in a Petri dish. The chips were put under a microscope to verify that

all the glue was removed. Out of the four methods, this was the quickest and safest with no visible damage to the bulk. It should be noted that this method did not work on all glue types. One glue in particular was impossible to remove using all previously mentioned methods.

4.1.6. Removing Si bulk

The thinning of the bulk is done in multiple steps, all done within the X-prep. The X-prep is able to level the sample based on a specified rectangular area. It does this by measuring four points; top left, top right, bottom left and the center. The accuracy achieved is a maximum difference between the measured points of $0.5\mu m$ of tilt of the horizontal plane.

The milling is done by moving both the sample and the mill. Many different mills can be inserted, depending on the size, speed and type of material that needs to be removed. Furthermore, different programs can be executed which also influence the quality. Finally, it is important to choose the same reference points for the area to be able to compare different programmes and mills. The next paragraphs will go through the different steps of milling. Note that the testing step discussed in section 4.2.4 is carried out after each milling step.

First, a larger area than the visible backside is selected to remove the package overhang and possible copper residue, see figure 4.8a. The parameters used are summarized in the list below. During this step, only a small amount of bulk is removed.

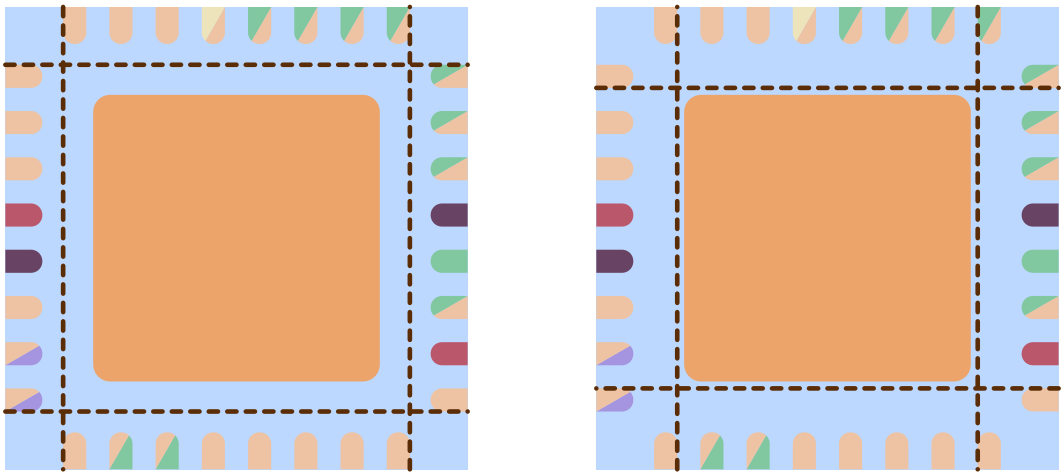
- cutting depth: $50\mu m$ in 2 passes
- tool: coarse plated diamond mill with a diameter of $1.5mm$
- rpm: 70000
- program: 75% overlap, front-back snake pattern and no notch selected
- xy-rate: $4mm/s$

The second step is to remove the major part of the bulk. The settings were as follows and the selected area can be found in figure 4.8b. It is important that the mill contains no copper residue as this is expected to cause shorts when the bulk is thinned. This can be done by cleaning them using for example Al etch, grinding the bottom contaminated part or using a new mill.

- cutting depth: 11 times $24.1\mu m$ plus 1 final pass $5\mu m$ which is in total $270.1\mu m$
- tool: fine metal diamond bond mill with a diameter of $0.7mm$
- rpm: 70000
- program: 85% overlap, alternating front-back and left-right snake pattern and no notch selected
- xy-rate: $4mm/s$ for the first 11 passes, $1mm/s$ for the final pass

Depending on the wear and tear or the severity of clogging up of the milling tool the actual thickness might deviate from the specified values. Therefore, a third step can be required. This final, more precise mill was done with the same area as step 2. Depending on the tests, this step is repeated until a thickness of around $5\mu m$ is achieved.

- cutting depth: 1 pass of $5\mu m$ plus 1 final pass of $0\mu m$
- tool: fine metal diamond bond mill with a diameter of $0.7mm$
- rpm: 70000
- program: 85% overlap, alternating front-back and left-right snake pattern and no notch selected
- xy-rate: $4mm/s$ for the first pass, $1mm/s$ for the final pass



(a) Selected area for removing residue

(b) Selected area for thinning bulk

Figure 4.8: Selected areas for milling

The fourth step is the start of the polish. Polishing requires a different program. Instead of a mill a metal tip is used on which a small circular felt is attached. Three different pastes were used containing glass balls with diameters of $15\mu m$, $6\mu m$ and $1\mu m$. The felt grinds the balls over the surface, thereby polishing the area into a smooth surface. The same area was selected and the following parameters were used for the first polishing step. The different overlap percentages and patterns ensure the polishing is done evenly and doesn't leave significant scratch marks. Between each step, the sample was cleaned with demineralized water. As running demi water might cause tribal charging it is best to submerge the sample in a beaker of demi water.

- paste: $15\mu m$ diamond paste
- area: $450\mu m$ margin to the edge
- tool: $1.5mm$ diameter and $1.25N$ as downwards force
- rpm: 5000
- program: 3 left-right snake pattern passes with 85% overlap and 1 left-right snake pattern pass with 75% overlap
- xy-rate: $2mm/s$

The fifth step is a more refined polish. A new tip with a clean felt was used. Most parameters stayed the same except the paste and program.

- paste: $6\mu m$ diamond paste
- program: 3 front-back snake pattern passes with 85% overlap and 1 front-back snake pattern pass with 75% overlap

The sixth and final step is the final polishing step. Again a new tip with a clean felt was used and only the paste and program were changed. The outline option in the program gives an additional pass at the edge of the selected area.

- paste: $1\mu m$ diamond paste
- program: 9 passes in an alternating front-back and left-right snake pattern pattern with an outline and 95% overlap

After these steps are executed, the bulk is thinned and ready for further processing. The substances used to cool the mill and the paste do not damage any other parts of the chip and can be cleaned easily with a wooden chopstick and water. Finally, although the tool indicates it will mill to the specified depth, in practice this is not accurate and depends on several parameters as discussed in section 3.2.4. It was therefore beneficial to mill in smaller steps the first time and increase the cutting depth every iteration. The remaining height is verified using the Allied X-vision which is discussed in section 4.2.4.

4.1.7. Determine areas of interest

To determine the areas of interest the samples from the previous processes were put under a SEM, the Lext and an NIR microscope. A short description of each device, its settings and abilities will now be given, starting with the SEM.

A SEM will generate an image by scanning a bundle of electrons over a sample in a high vacuum, see figure 4.9. These electrons interact with the atoms present at the surface, generating secondary electrons. The energy and amount of these secondary electrons indicate the material and its structure. The resulting image will be grey-scale but using this technique magnifications of 50000x are possible. Some experience is required to create a clear image because there are many variables to tune. Furthermore, it is essential to have a conducting sample. Without proper conduction, the electrons will be trapped inside. This charges the sample which means other electrons are pushed in random directions, thus creating a distorted image. As the chip is most of the time inside an insulating package additional steps have to be taken to make sure the electrons can flow away.

First of all a test was carried out to deduce which pins can be connected to ground while keeping the chip functioning properly. By connecting as many pads as possible to ground more potential paths are created for electrons to disperse. Secondly, part of the bulk that was away from the areas of interest was connected to ground. Two options were available to make these connections. The first option was copper tape which has conducting glue on one side. The second option was carbon paint. As the name implies, this is paint which contains carbon to make it conductable. It was found that although the carbon paint was a better conductor and more reliable, it also connected parts that weren't supposed to be connected. Furthermore, removing the paint proved to be impossible as there would always remain some residue that caused a short. Therefore, the paint was only used as a one-way solution and most of the time copper tape was applied. Another option would be to use the nanoparticle printer of the Else Kooi Laboratory (EKL) at the Delft University of Technology.

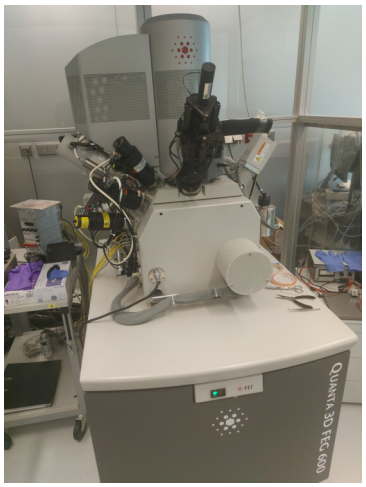
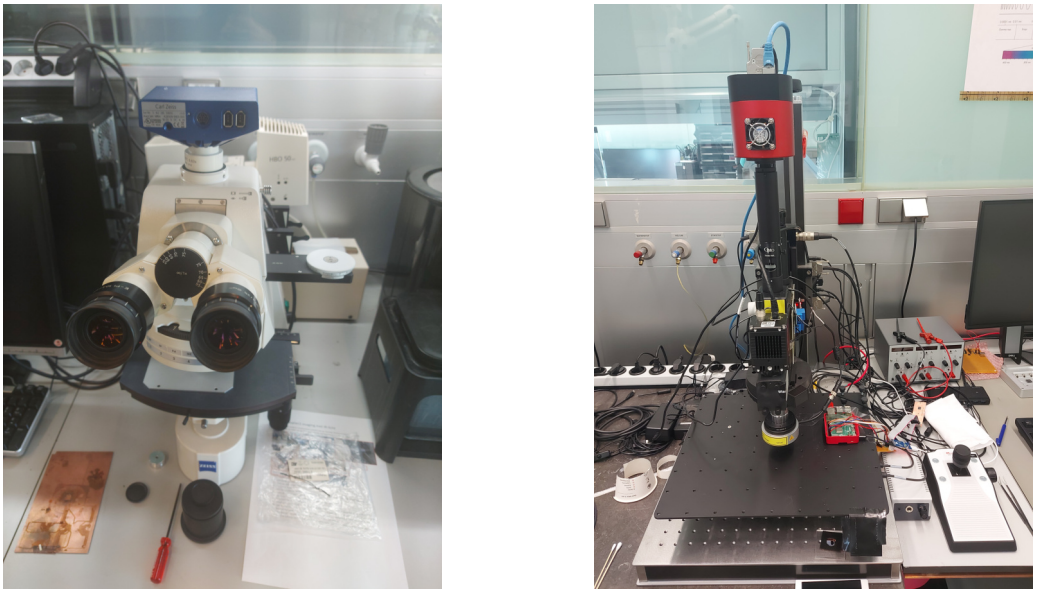


Figure 4.9: SEM-FIB tool

The second measurement tool used was the Lext. This tool has already been described in section 4.1.3. The coloured images proved especially useful by giving more context to the grey-scaled images from the SEM.

Finally, reflection-based NIR microscopes were used. The benefits and implementation of these tools have already been discussed in section 2.2.1. The operation of such a microscope is similar to that of a normal optical microscope. The only difference is that the source is now in the NIR spectrum instead of the observable light spectrum. This means that a detector was needed to create an image which was displayed on a computer screen. As the image acquisition already happens in the digital domain, the setup was expanded to contain a movable stage and stitching software. Figure 4.10 shows both IR microscopes.



(a) Standard IR microscope

(b) IR microscope with moving stage

Figure 4.10: IR microscopes

4.1.8. Creation of holes in the Si bulk

To create the holes, the sample was loaded into the FIB-SEM. Once a high vacuum was reached, the sample was aligned. As described in section 3.3.1 a combination of IR photos and SEM images was used to align the die and zoom in on the area of interest. From that point, the tool can be switched to the FIB mode. The settings of the FIB are similar to that of the SEM except current indicates the strength of the FIB instead of voltage with the SEM.

While tuning the FIB it is important to note two things. Firstly, the ion source has only a limited amount of working hours. This means that to use the source efficiently it should only be turned on if it is used for milling or taking pictures. Secondly, every time an image or video is taken with the FIB the sample is bombarded with ions. This means the sample is already being milled. To ensure fine features or thin layers stay intact, the number of images should be kept to a minimum, especially at high currents.

The FIB is equipped with a gas injection system (GIS). In practice, this means a needle can be placed, just above the area of interest to eject a type of gas into the chamber. These needles are placed under an angle and are fragile. Careful consideration has to be made to ensure the needle doesn't hit anything. When the area of interest is close to the edge and therefore close to a wall of the packaging, the sample might need to be rotated to avoid a collision. Finally, it is important to retract the needle when navigating over the sample.

Once the FIB is calibrated correctly and all its systems are warmed up it is time to create the holes. The steps are described below. It is important to note that the thickness of the bulk at this point is not known exactly and might vary across the die. This in turn means that a repetitive and robust recipe is not possible. Fortunately, it is possible to activate the Everhart–Thornley detector (ETD) during the milling operation, which collects the secondary and backscatter electrons resulting in a live feed. This helps to adjust or stop the process when necessary.

The creation of a hole was generally done in two steps. The first step is creating a relatively large hole to thin the bulk around the area of interest even more. An example of such a process is summarized using the following list where the area is defined as width by height by depth and time in minutes and seconds. In general, a balance needs to be found between the area, current and time. Note that large currents reduce the quality and large areas increase the time or current. Additionally, the time should be chosen to not be too long because of the limited working hours and long waiting times, but also not too short because the process might need to be stopped halfway through.

- Process: IEE
- Area: $30\mu m$ x $20\mu m$ x $7\mu m$

- Current: $50nA$
- Time: 7min 18s

The second step is the Si etch. This is done without any gas, just using the gallium ions. This results in cleaner etches and better control over the depth but with longer etching times. This means a smaller surface was selected. Furthermore, this step was done several times before the desired result was achieved. An example process is listed below.

- Process: Si
- Area: $2\mu m$ x $2\mu m$ x $1\mu m$
- Current: $0.30nA$
- Time: 1min 29s

4.1.9. Create pillars

Following route ② from figure 3.4, the next step is to create pillars which is done using the FIB. If route ① was followed, the chip needs to be aligned and etched as described in section 4.1.8. There are two materials available in the FIB; platina and tungsten. For this project, it was chosen to use platina because it is easier to work with and depositing a pillar inside a hole with low conductivity is already challenging.

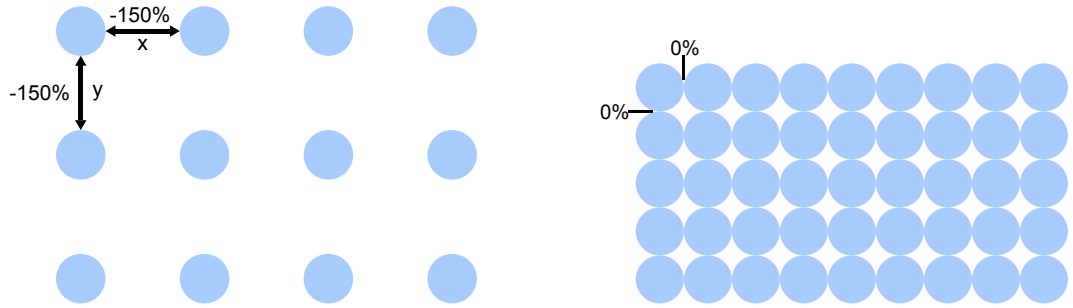
To deposit platina a current of $2 - 6\mu A/\mu m^2$ is advised. On top of that, the gas methylcyclopentadienyl(trimethyl) platinum is inserted into the chamber. The rest of the process to create a pillar of around $3\mu m$ is listed below.

- Process: Pt deposition
- Area: $1\mu m$ x $0.5\mu m$ x $9\mu m$
- Current: $10pA$
- Time: 16min 02s

Continuing on route ②, there is the option to deposit TEOS. Creating a sufficient insulating film for possibly $5V$ requires four more steps. The optimal current to deposit TEOS is $1 - 10pA/\mu m^2$. The area on which the TEOS needs to be deposited can be selected using a polynomial meaning the pillar can be excluded. The process is described below.

1. Depositing $300\mu m$ insulation
    - Process: TEOS deposition
    - Area: polygon of approximately $600\mu m^2$
    - Current: $1nA$
    - Time: 11min 18s
    - Other: dwell time is $200ns$ and 0% overlap in both the x and y axis
2. Waiting
    - Wait until the gas is pumped out of the vacuum chamber and the high vacuum pressure ($\sim 5e - 6mbar$) is restored
3. Improve resistance by cleaning
    - Process: IEE
    - Area: same polygon of approximately $600\mu m^2$
    - Current: $1nA$
    - Time: 0min 7s
4. Repeat steps 1-3 one more time

The overlap percentage when depositing TEOS determines the thickness and quality of the insulation layer. Figure 4.11 shows how the numbers should be interpreted. To deposit TEOS, the beam is aimed at a spot, indicated by a circle, and shoots ions for the amount of dwell time specified. It then moves to the next location in a horizontal scan. With iterative testing, it was found that the best result was achieved using 0% overlap. Having negative percentages resulted in long deposition times while positive percentages were not accepted by the software.



(a) -150% overlap **(b)** 0% overlap

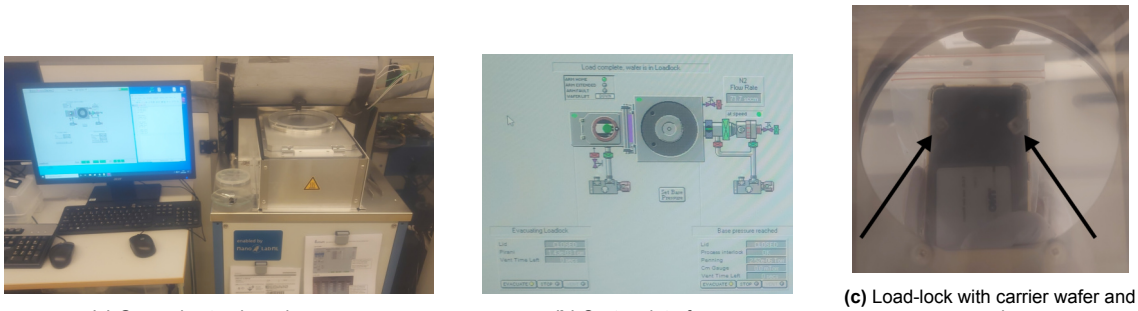
Figure 4.11: Overlap percentage explained

4.1.10. Deposit insulation

The deposition of a thin insulation layer is the only step that cannot be done in the NFI lab and was done in the Kavli cleanroom of Delft University of Technology. Two tools were used; the ALD tool and ICPECVD tool. These tools and the used recipes will be discussed next.

ALD

The Oxford ALD, shown in figure 4.12a, exists in two rooms. One room where the user normally is and one maintenance room. Located in the first room is the computer on which the recipes can be created and the machine can be monitored, see figure 4.12b. Besides the computer is the load-lock in which the carrier wafer with the samples on top was placed, see figure 4.12c. Before the recipe can be run the correct valves for the gasses need to be opened. This was done in the maintenance room.



(a) General setup in main room **(b)** System interface **(c)** Load-lock with carrier wafer and samples

Figure 4.12: ALD tool setup

There were two recipes used; one $14.5nm\ AlOx$ deposition and one deposition with $5nm\ HfOx$ and $7nm\ AlOx$. Important to note is that the tool needs to be cleaned before and after the process. This is done by depositing a film with only an empty wafer inside, thereby creating a protective coating with the same specifications as the desired coating. Eventually, this coating is removed with a pre-programmed cleaning process. This means that in total a deposition can take the better part of a day. Fortunately, multiple samples can be placed on one wafer and the machine can be left alone.

The different recipes as summarized in the tables 4.1 and 4.2. The columns indicate the different steps with on each row the characteristics of each step. The first deposition took one hour and 30 minutes. The second and third depositions took two hours and one hour respectively. To change the thickness of the layer, the amount of loops was adjusted. These recipes were standard recipes with known and tested deposition rates.

Table 4.1: Recipe for $14.5nm\ AlOx$ ALD at $105°C$

	Stabilize	Gas flow stabilization	Repeat (200x)	$Al(CH_3)_3$ Pulse	$Al(CH_3)_3$ Purge	H_2O Pulse	H_2O Pulse	Loop	Evacuate chamber
Temperature	$105°C$	$105°C$	$105°C$	$105°C$	$105°C$	$105°C$	$105°C$	$105°C$	$105°C$
Time	10min	2min	-	0.020s	12s	0.25s	12s	-	-
N_2	Closed	Closed	-	$70sccm$	$70sccm$	$70sccm$	$70sccm$	-	Closed

Table 4.2: Recipies for $HfOx$ and $AlOx$ ALD at $105°C$

(a) First part of combined recipe for $5nm\ HfOx$ ALD at $105°C$

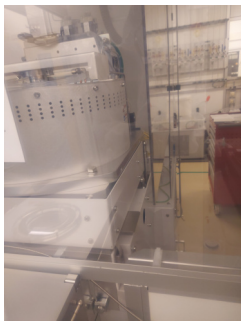
	Pump to pressure	Preheat	Repeat (50x)	TDMAH dose	TDMAH purge	H_2O Pulse	H_2O Purge	Loop	Evacuate chamber
Temperature	$105°C$	$105°C$	$105°C$	$105°C$	$105°C$	$105°C$	$105°C$	$105°C$	$105°C$
Time	-	-	-	0.45s	59s	0.1s	59s	-	-
N_2	Closed	Closed	-	$100sccm$	$100sccm$	$100sccm$	$100sccm$	-	Closed

(b) Second part of combined recipe for $7.5nm\ AlOx$ ALD at $105°C$

	Stabilize	Gas flow stabilization	Repeat (100x)	$Al(CH_3)_3$ Pulse	$Al(CH_3)_3$ Purge	H_2O Pulse	H_2O Purge	Loop	Evacuate chamber
Temperature	$105°C$	$105°C$	$105°C$	$105°C$	$105°C$	$105°C$	$105°C$	$105°C$	$105°C$
Time	10min	2min	-	0.020s	12s	0.25s	12s	-	-
N_2	Closed	Closed	-	$70sccm$	$70sccm$	$70sccm$	$70sccm$	-	Closed

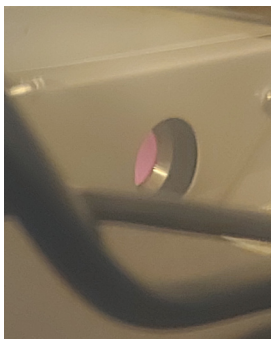
ICPECVD

The PlasmaPro100 ICPECVD was used for the ICPECVD. Like the ALD tool, this tool is also divided over two rooms. For this tool, everything can be done with the computer and the load-lock. Figure 4.13 shows the tool and the load-lock. During operation, a purple glow from the plasma can be observed coming from the window, see figure 4.13c.



(a) ICPECVD tool

(b) Loadlock with carrier wafer and sample



(c) Purple hue resulting from the plasma

Figure 4.13: ICPECVD tool setup

There is a cleaning process before and after the recipe but in total it takes only about an hour to deposit tens of nanometers of insulator. The samples on the wafer were placed at a slight angle to reduce the chance of gliding of the wafer. Two recipes were programmed on the computer; one for SiO_2 and one for SiN. The SiO_2 should attach better to the bulk and introduce less stress while the SiN ideally is more conformal and has a smoother surface. The recipes can be found in table 4.3 and table 4.4. Each column represents a step of the recipe.

The height of the film is determined by the time spent in the deposition step. The deposition rates were found iteratively and were $59nm/min$ at $150°C$ for SiO_2 and $45nm/min$ at $100°C$ for SiN. A film thickness of around $400nm$ should provide enough electrical insulation and a surface onto which new features could be built. This meant a deposition time of $7min$ for SiO_2 and $8.5min$ for SiN.

Table 4.3: Recipe for ICPECVD with SiO_2 at $150°C$

	Heat-up	Gas stab	Strike 1	Strike 2	SiO_2 deposition	Purge	Pump
Time	4m 0s	20s	10s	10s	7m 0s	30s	30s
APC	$20.0mTorr$	$12.0mTorr$	$12.0mTorr$	$8.00mTorr$	$8.00mTorr$	Open	Open
ICP RF	Off	Off				Off	Off
• Power Demand			$1300W$	$1300W$	$1300W$		
• Reflective Power			$250W$	$150W$	$150W$		
• Toll time			10s	10s	10s		
SiH_4	Off	$16.0sccm$	$16.0sccm$	$16.0sccm$	$16.0sccm$	Off	Off
O_2	Off	$60.0sccm$	$60.0sccm$	$60.0sccm$	$60.0sccm$	Off	Off
Ar	$10.0sccm$	$10.0sccm$	$10.0sccm$	$10.0sccm$	$10.0sccm$	$10.0sccm$	Off

Table 4.4: Recipe for ICPECVD with SiN at $100°C$

	Heat-up	Gas stab	Strike 1	Strike 2	SiO_2 deposition	Purge	Pump
Time	3m 0s	20s	8s	8s	8m 30s	1m 0s	30s
APC	$50.0mTorr$	$10.0mTorr$	$10.0mTorr$	$5.00mTorr$	$2.00mTorr$	Open	Open
ICP RF	Off	Off				Off	Off
• Power Demand			$1200W$	$1200W$	$500W$		
• Reflective Power			$150W$	$150W$	$150W$		
• Toll time			10s	10s	10s		
SiH_4	Off	$26.0sccm$	$26.0sccm$	$26.0sccm$	$26.0sccm$	Off	Off
O_2	Off	$50.0sccm$	$25.0sccm$	$25.0sccm$	$25.0sccm$	Off	Off
Ar	Off	Off	Off	Off	Off	$10.0sccm$	Off

Once an insulation film was deposited the samples were brought back the the NFI lab and placed back inside the SEM-FIB. They were then aligned using the SEM and the remaining markings such as the edge of the holes and the pads on the outside. Depending on the chosen route, only the pillar needs to be exposed again or the process 'Create pillars' from section 4.1.9 should be executed.

4.1.11. Create pads and traces

This step is basically an adaptation of the 'Create pillars' process from section 4.1.9. Instead of building a tall pillar, a long trace was deposited. This was tried with both platina and tungsten.

The deposited structure consists of two parts; the trace and the bonding pad. The dimensions of the trace were chosen to be $6\mu m$ in width, $2\mu m$ in height (twice the standard height) and around $130\mu m$ in length. These values were chosen to keep the deposition time manageable at around ten minutes, the pad sufficiently away from the hole and the resistance as low as possible. Two different sizes of pads were tried; one of $50\mu m$ by $50\mu m$ and one of $25\mu m$ by $25\mu m$. Both are again $2\mu m$ thick.

Table 4.5 describes the different processes run. It is important to note that for tungsten the optimal deposition current is $70 - 100pA/\mu m^2$ and for platina $2 - 6pA/\mu m^2$. The expected resistance was calculated using the area, height and equation 4.1.

$$R_{expected} = \frac{L}{WH} * \rho \text{ where } \rho_W = 2\Omega * \mu m \text{ and } \rho_{Pt} = 10 - 20 \approx 15\Omega * \mu m \tag{4.1}$$

Table 4.5: Pads and traces deposition

#Feature	Size [$W * L * H$][μm^3]	Current [nA]	Time [min:s]	Overlap [%]	$R_{expected}$ [Ω]
Tungsten					
1 pad	$50 * 50 * 0.6$	15	3:22	0	3.3
2 pad	$50 * 50 * 3$	65	3:55	0	0.67
3 pad	$25 * 25 * 5$	50	2:07	0	0.67
4 pad	$25 * 25 * 0.6$	1	12:32	0	3.3
5 pad	$25 * 25 * 0.6$	1	12:32	$+70$	5.6
Platina					
1 pad	$50 * 50 * 2$	15	11:13	0	7.5
2 pad	$50 * 50 * 2$	15	11:13	0	7.5
3 pad	$25 * 25 * 2$	3	14:02	0	7.5
4 pad	$25 * 25 * 2$	3	14:02	0	7.5
5 trace	$6 * 126 * 2$	5	10:18	0	158
6 trace	$5.5 * 150 * 2$	5	11:25	0	205

4.1.12. Wirebonding

To complete the recipe, the chip is wire-bonded to the external probe. This is done using the Bondtec, see figure 4.14a. Figure 4.14b shows the printed circuit board (PCB) of the external probe which serves as a reference.

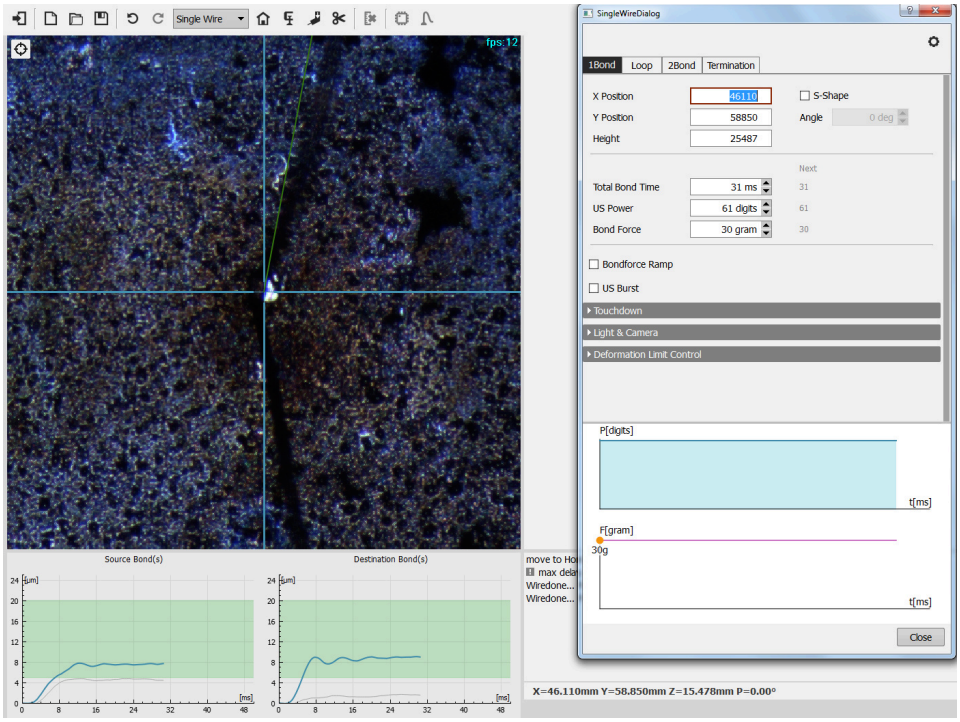
The Bondtec tool has many options, making it possible to bond the wires from the pad upwards and at an angle to the probe. These options are shown in figure 4.14c. The top left shows the view of the camera. This is used to navigate to the bonding places. Below the camera is an estimation of the bonding quality. The bond should be successful as long as the graph stays within the green band. Finally, on the right, are the parameters which determine the bond.

For this bond only wedge bonding was available since the wires needed for ball bonding were not present. The wire used was an aluminium wire with a diameter of $17.5\mu m$. The expected resistance of this wire can be calculated using equation 4.2. To create more reliable bonds the power and force were increased by 50% from its standard of $61digits$ and $30g$. Finally, the angle was adjusted to prevent the bonding head from hitting the package edge. As the bonder needs some margin to move and place the wedge bond it is safer to bond parallel to the closest wall and adjust the angle accordingly.

$$R_{expected} = \frac{L}{A} * \rho \text{ where } \rho_{Al} = 0.0265\Omega * \mu m \text{ and } A \text{ is the crossectional area} \tag{4.2}$$

(a) Bondtec tool

(b) External probe mounted on the Bondtec



(c) Bonding options

Figure 4.14: Wirebonding setup

4.2. Testing

This section aims to discuss the different testing setups that were used throughout the whole recipe. It is important to test the chip before and after to ensure the viability of each process step. Several setups have already been discussed in section 4.1 like the IR microscopes, the images of the SEM and the Lext. These tools were also used for testing and verification purposes. However, these tools fail to measure the functionality of the chip.

This section starts with three different setups which are able to connect to the chip, verify its functionality and perform read and write functions to the flash. This is followed by a section on a setup to measure different resistances in the distinct parts of the backside connection. Finally, the X-Vision is discussed which was used to determine the thickness of the bulk.

4.2.1. BeeProg2

The first functionality tests were performed with the Beepgrog2. This device is plugged into a computer and an adapter is placed on top which supports the HVQFN packaging format, see figure 4.15. On the computer the Elnec software is run which is able to read and write to the chip. The chip is placed in the adapter and pressed against pins that connect to the pads on the chip.

This setup has the advantage that chips are easily placed and tested and therefore relatively fast. However, the adapter finds it hard to detect the chip if the pads on the chip are not perfectly intact. This happens already during the etching of the heatsink, thereby making this setup unreliable after this process. On top of that, the Elnec program is only able to verify the chip's functionality if each pin is connected to the output. This should not be necessary since, according to the datasheet, only six pins were required to operate the target chip. These shortcomings led to the implementation of the second test setup.

Figure 4.15: BeeProg2 with DIL32/MLF32-1 adapter

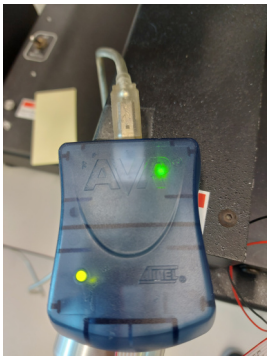


Figure 4.16: AVRISP mkII

4.2.2. AVRISP mkII

The AVRISP mkII is a similar device to the BeeProg2 in the sense that it serves as a controller between the computer and the chip. The AVRISP mkII, however, is more customizable. It has a six-wire cable as an interface onto which self-made adapters can be attached. Furthermore, it has an open-source graphic user interface (GUI) and a huge library with hundreds of supported devices. Using this setup meant only the six necessary pins needed to be connected. Figure 4.16 shows the slightly transparent casing.

That gave rise to the question of how many of the other pins could be connected to ground while preserving the full functionality of the chip. Connecting as many pins to ground would improve the functionality of the SEM and FIB. Using the datasheet figure 4.17 was constructed. The six mandatory pins are indicated in yellow. The brown line depicts a possible wire connecting as many pads to ground as possible.

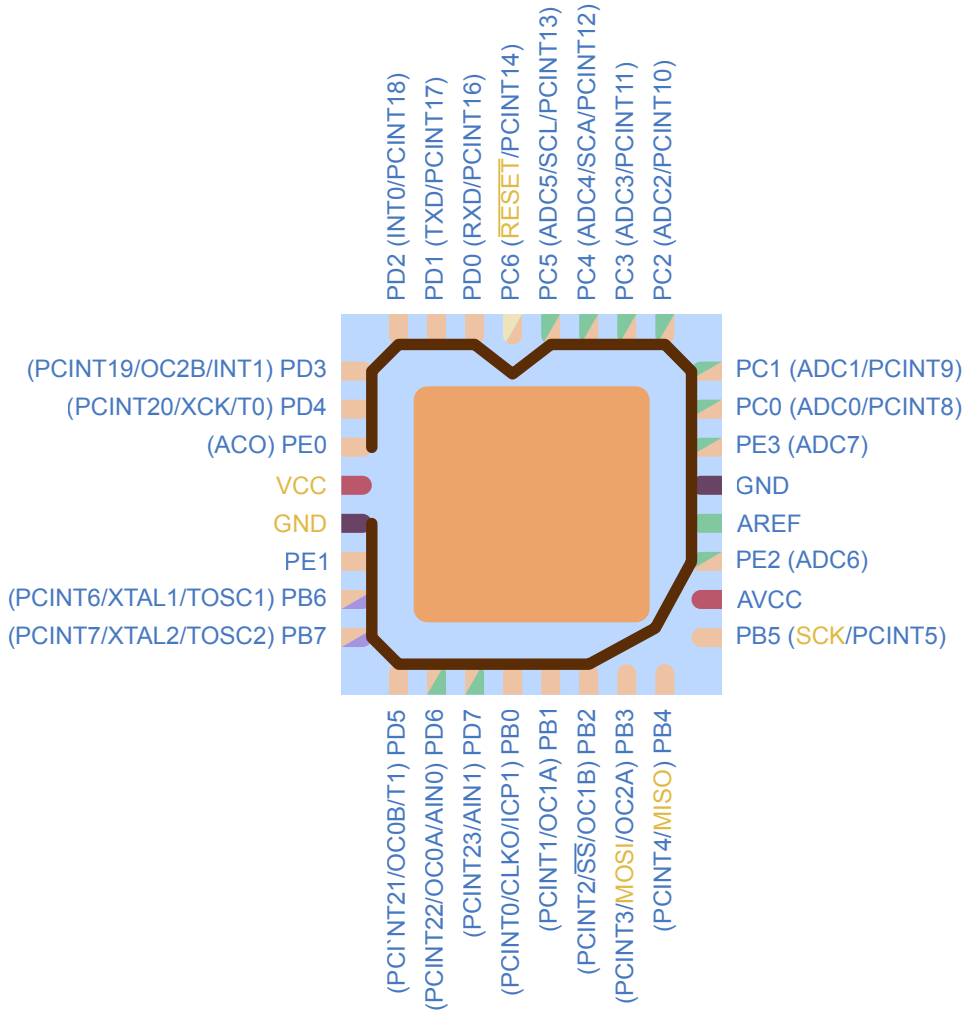
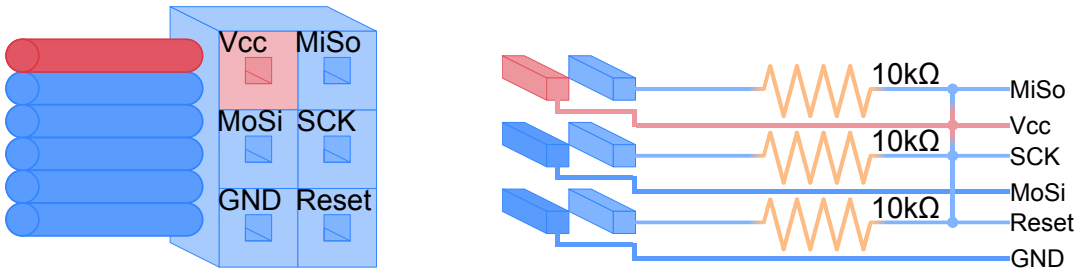


Figure 4.17: Chip layout including ground wire

With this information, two adapters were built; one for the probestation and one for the SEM. The schematic of these adapters is shown in figure 4.18 where Vcc and ground are also connected to an external voltage source set to $3.5V$. Their implementation is discussed next.



(a) AVRISP mkII interface

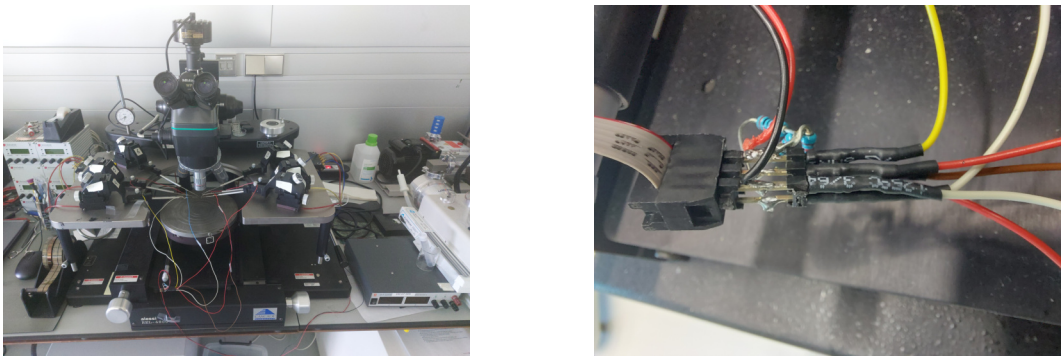
(b) Adapter schematic

Figure 4.18: AVRISP mkII general setup

Probe station

Once the chip has undergone some processing steps, the pads are not pristine anymore. Meaning they were damaged, missing parts, partly oxidized and gathered dust. That resulted in the use of the probestation where tiny needles are placed on the good parts of the pads. This is done manually, meaning it is slower than the BeeProg2. To speed up the process, the samples were placed in the same orientation and secured using a vacuum underneath. This setup is shown in figure 4.19a.

Next, an adapter was made which is shown in figure 4.19b. This adapter sits between the probes of the probe station and the AVRISP mkII and consists of three resistors in this case. Once all the signals are connected, the chip functionality can be verified and flash can be read and written.



(a) Probe station

(b) Adapter for probe station

Figure 4.19: Probe station setup

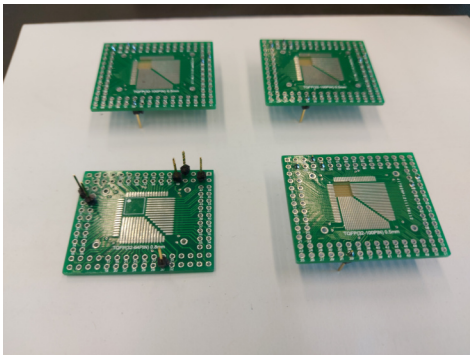
SEM

For the SEM more components were needed to build a test setup. The aim of this setup was to be able to operate the chip while it was inside the SEM. This in turn would create the option to do voltage contrast imaging (VCI). With VCI, a voltage is applied to a specific trace or circuit. This voltage causes a change in electric potential at the surface which influences the electrons from the electron beam. If the electrons are trapped the image will show a dark trace, when they are repelled a more white trace appears. Additionally, switching the voltage source on and off can be seen in real time. This would be useful in detecting interesting areas on the chip or in fault analysis. For example, it would be possible to see if the pillars are connected to the desired trace.

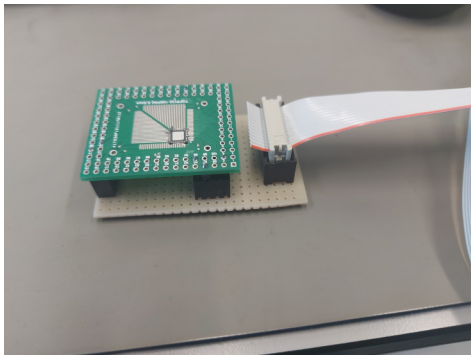
There already was a cable going into the SEM which had a female connector on both sides. That meant that an adapter had to be built on each side. The adapter outside of the SEM is shown in figure 4.20d. This adapter implements the required resistors, creates connectors for the voltage source and reroutes the pins to the correct inputs of the AVRISP mkII interface.

The adapter on the inside of the SEM required more work. A general PCB was chosen and a hole of exactly the size of the chip was created. This had two major advantages. First, the chip was lowered, thereby lowering the overall height of the structure. Secondly, because the chip was kept in place by the hole and the pads were lowered to the correct height, it was quicker and easier to solder the chip to the PCB. This PCB was then placed on top of another PCB. This second PCB was mounted inside the SEM and connected to the correct pins of the chip to the cable going out of the SEM.

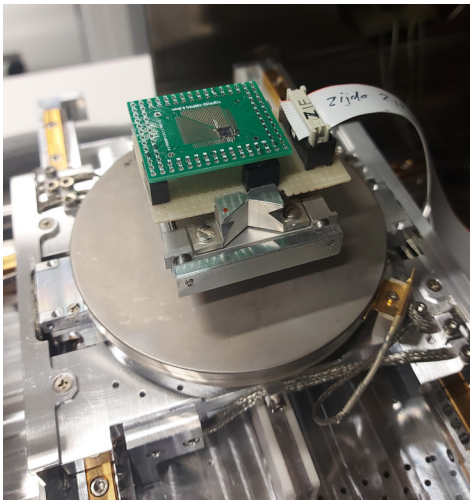
Figure 4.20 shows these different adapters. This setup fits the size constraints of the SEM while being quick to install. On top of that, chips are easily exchanged by replacing the top PCBs.



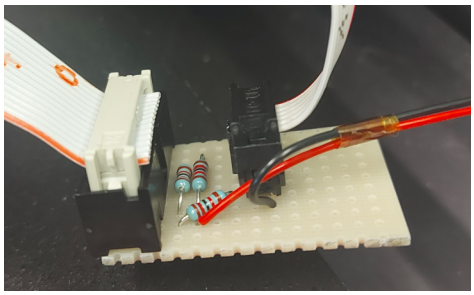
(a) PCB adapter with chip



(b) Adapter at the bottom of the SEM



(c) Full setup inside SEM



(d) Adapter outside the SEM

Figure 4.20: SEM adapters

4.2.3. Resistance measurements

One quantitative measurement that can be done is measuring the resistance of the backside contact. This is important to keep as low as possible to preserve the signal integrity. As will become clear in the results section, there was a problem with a short between Vcc and Gnd. The next sections will discuss both instances.

Multimeter

To measure if there is a short the Fluke 189 was connected to the Vcc and Gnd pad, see figure 4.21. According to the datasheet, the resistance is calculated by injecting a small current and measuring the corresponding voltage. To make sure no latch-up or other shorts are created during the measurement it was advised to limit the range to $5M\Omega$.

Figure 4.21: Fluke 189 multimeter

Backside contact

The backside contact can be divided into four resistive parts connected in series. Figure 4.22 shows where three of these resistances are situated with additionally the resistance from the bond wires. The Vcc pad was chosen as a suitable place for these measurements for the following reasons.

First, the Vcc pad has many vias over a large area, making it possible to build larger structures on top of it. Secondly, the contact point on the backside will be close to the wire bond on the front side, thereby making the unknown internal resistance as small as possible. Finally, in contrast to the Gnd pad, the Vcc pad should be insulated from the bulk which means that the effect of the insulation layer can be tested at the same time.

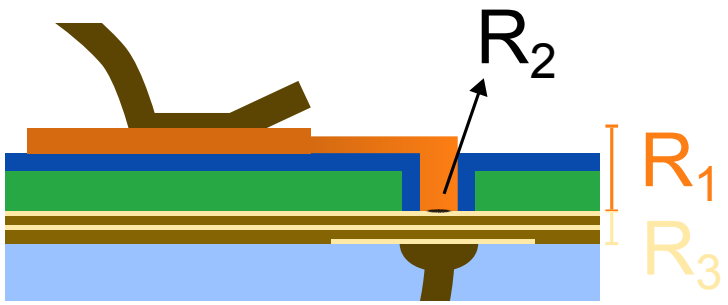
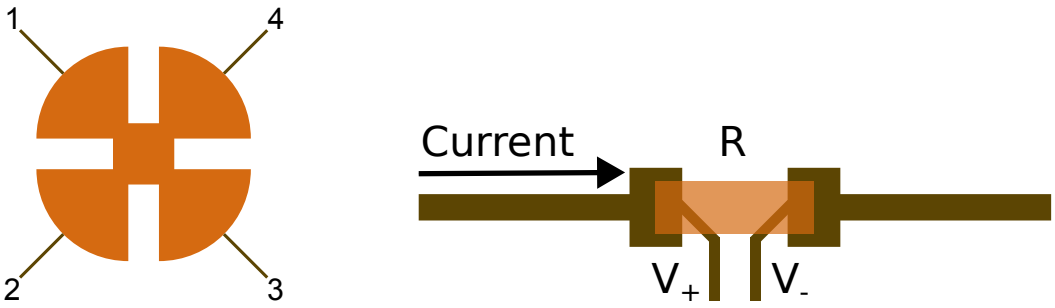


Figure 4.22: Resistances in signal path

The three resistances indicated in figure 4.22 are the following. R_1 indicates the resistance of the deposited platina or tungsten. This includes the pillar, trace and bonding pad. Next, R_2 indicates the contact resistance between the pillar and the interconnect. Finally, R_3 indicates the resistance of the interconnects and vias within the die. The three resistances each require their own measuring method.

Starting with R_3, this resistance is almost impossible to measure and depends largely on the chip design. However, because the Vcc pad was chosen, this resistance is likely to be small and certainly much smaller than R_1 and R_2, therefore it is assumed that $R_3 << R_1||R_2$.



(a) Van der Pauw structure

(b) Kelvin connection

Figure 4.23: Resistance measurement structures

For R_1 it is possible to calculate its resistance with reasonable accuracy. This is done using the electrical linewidth measurement (ELM) method [43]. This method uses sheet resistance ($R_\square$) to calculate a structure's resistance based on the number of unity sheets that fit inside the structure. To get the sheet resistance a Van der Pauw structure can be created, see figure 4.23a. Two neighbouring corners are connected to a current source which generates a current flow through the resistance. This generates a voltage which is measured on the other two corners. When the same resistance is found for currents in both the horizontal and vertical direction, the sheet resistance can be calculated using equation 4.3.

$$R_\square = \frac{\pi R}{ln(2)} \tag{4.3}$$

The average resistance of the structure can then be determined as $\rho = R_\square * t$, where t is the structure thickness in meters. From this point, equation 4.2 can be used with the correct ρ, length and cross-sectional area.

When the structure consists only of a line, the Kelvin connection can be used. This is a simpler version of the Van der Pauw structure, see figure 4.23b. By placing the voltage sense connections immediately adjacent to the resistance and by having a measurement tool that accepts almost no current flow, the voltage drop in these connections is negligible. In this case, only a horizontal current and Ohm's law are applied.

Assuming that the resistance of the bond wires and R_1 are calculated using the ELM method and R_3 is neglected, leaves R_2 unknown. This will be the remaining resistance and can be determined by averaging multiple measurements. Doing it this way leaves room for errors and unknown resistances. A possible solution could be to make a reference sample of many platina pillars on aluminium traces. Averaging these results would remove the unknown resistance and give a reference value which can be used in the real application.

4.2.4. X-vision

It is important to measure the thickness of the bulk regularly as the milling depth of the X-prep can vary during the process. The fastest and most practical option to do this is the Allied X-vision, see figure 4.24. Since the same company develops both tools, the samples can be exchanged between the tools without calibrating and levelling each time.

The X-vision makes use of a focussed IR source. This is projected onto the surface, thereby creating a unique signal based on the refractive index of the material. This signal is then analysed, resulting in a thickness. Multiple measurements are done in a grid pattern to get a better understanding of the thickness across the whole surface.

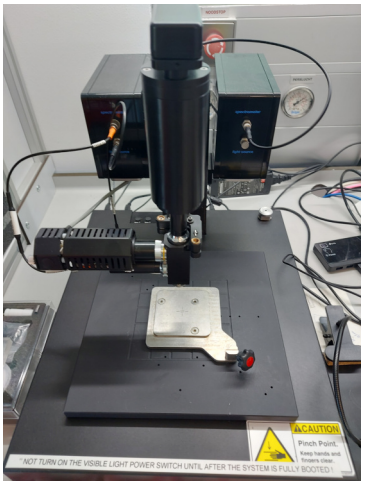


Figure 4.24: X-vision

4.2.5. IR heat camera

At this point, all the IR measurements and imagery were done in the NIR or short wave IR range. However, when a short is created and a large current goes through, this generates a heated spot. This, if captured, could pinpoint the place of the short, thereby giving insight into where the fault is and possible causes. To get a thermal image a different part of the spectrum should be detected, as indicated in figure 4.25.

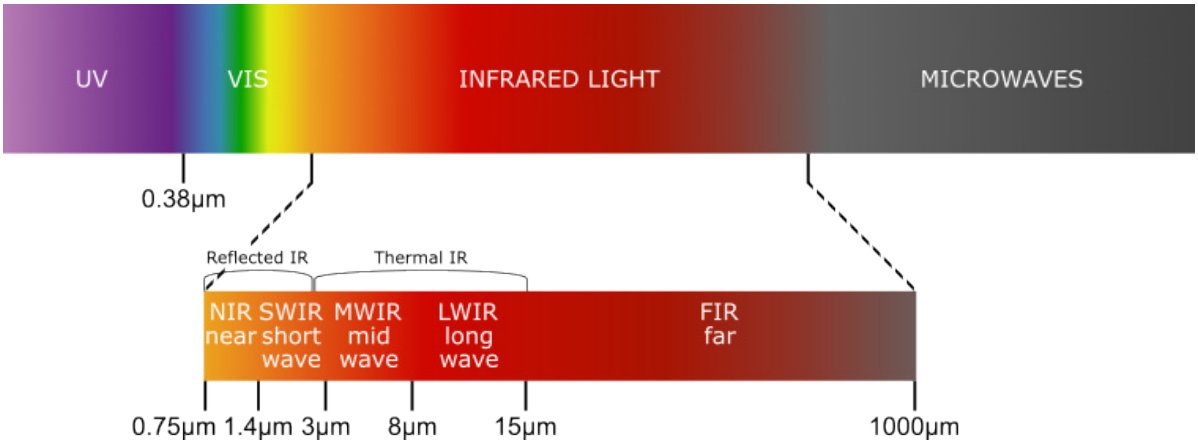
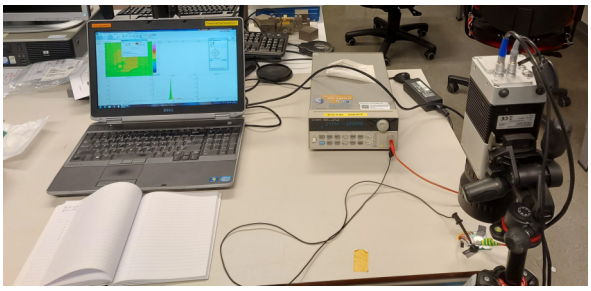
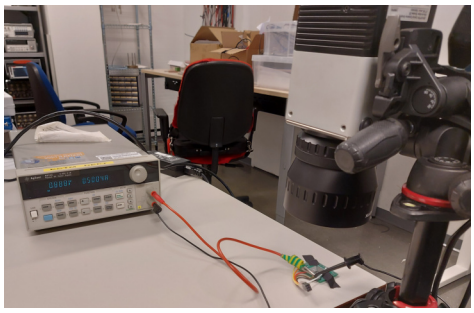


Figure 4.25: IR spectrum

To detect the thermal IR waves a long wave IR camera from Jenoptik was used. This camera was able to capture a temperature range from 0 to $500°C$. The setup is shown in figure 4.26.



(a) Overview of laptop, voltage source and IR camera



(b) Chip placed under camera

Figure 4.26: IR thermal imaging setup

5

Results and evaluation

This chapter will cover the results from the experiments described in chapter 4. Furthermore, these results are discussed and evaluated. Some of these improvements have been implemented in a second or third iteration, resulting in a more detailed evaluation. There were over 50 samples used in many processes consisting of several steps. As most of the results are based on imagery, showing all of them is therefore impractical. Therefore a selection was made containing the most representative or noteworthy results.

The structure of this chapter is similar to chapter 4, where each phase of the recipe will be discussed in a separate section. The tests that were carried out in between will be discussed at the corresponding place. This means that the results are discussed in chronological order. Additionally, a complete summary of all the samples and which processes they went through can be found in appendix B.

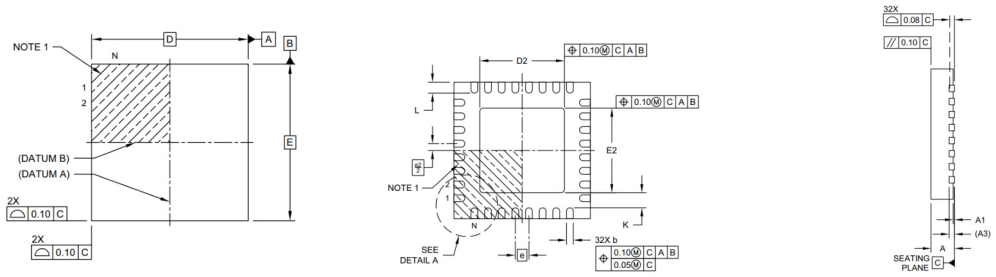
Finally, the sample on which the project is based is confidential. Fortunately, this only impacts some of the images, where complete overviews of the chip have been omitted. All results that are not directly traceable to this chip are published in full.

5.1. Acquisition of chip layout and technology

As discussed before, this phase is about gathering information about the chip design, layout and other properties like materials used and temperature range. This section aims to provide the necessary information and results to identify the characteristics of the chip. Each section adds additional information, some more useful than others.

5.1.1. Outside observation

Three main pieces of information were gathered in a non-destructive way; the pin layout, general measurements of the size and an X-ray scan. The pin layout was already shown in figure 4.17. Figure 5.1 and table 5.1 were taken from the datasheet which indicates the size of the package and the relative position of the pads. It was found that these dimensions were very accurate, as expected. Finally, figure 5.2 shows the X-ray images which show where the die is situated and how the bond wires curve. Additionally, some SEM images were made later to put the X-ray images into context.

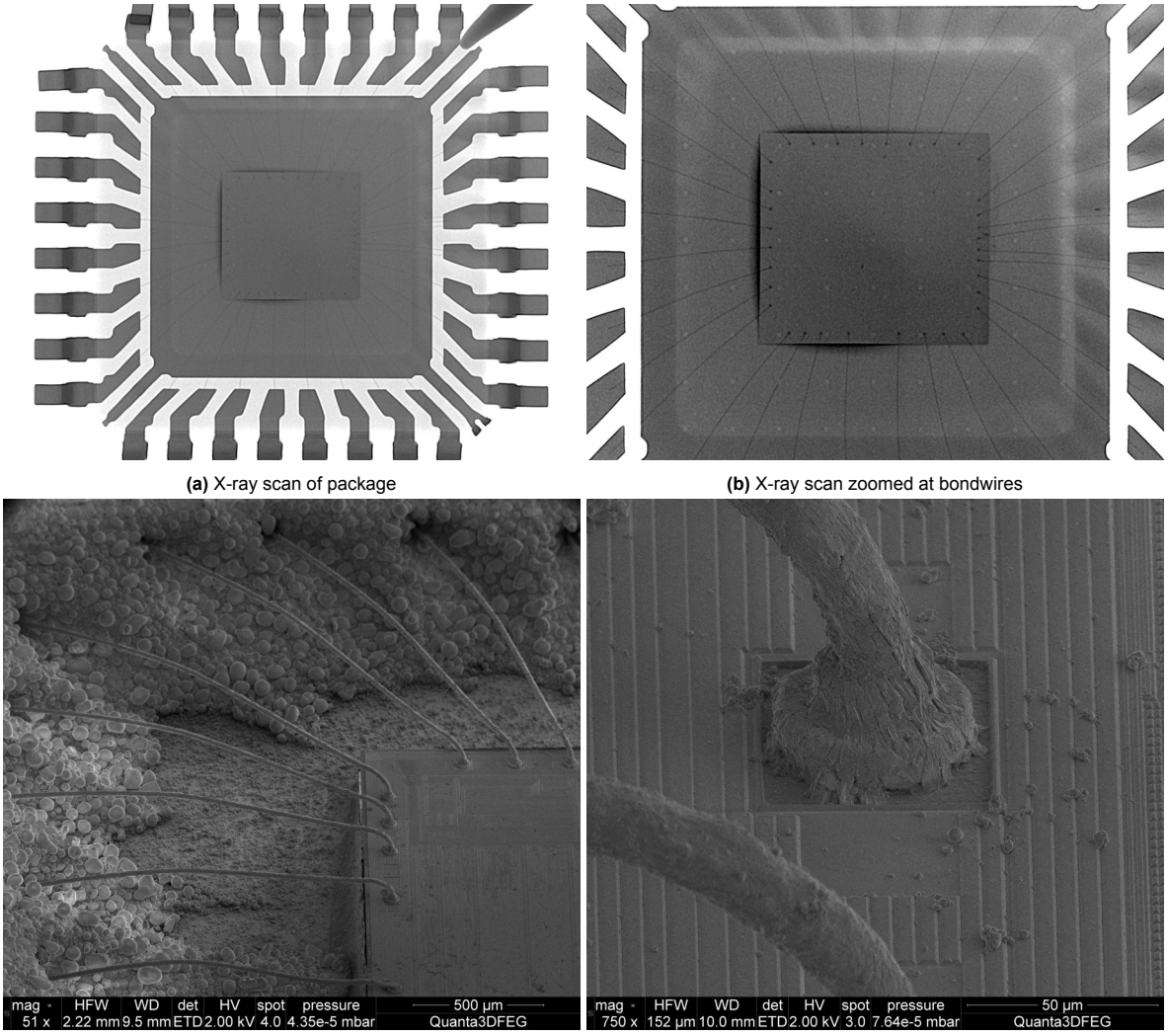


(a) Top view **(b)** Bottom view **(c)** Side view

Figure 5.1: Sizing and package from datasheet

Table 5.1: Packaging dimensions from datasheet

Units		[mm]		
Dimension Limits		MIN	NOM	MAX
Number of terminals	N		32	
Pitch	e		0.50	
Overall height	A	0.80	0.85	0.90
Standoff	A1	0.00	0.02	0.05
Terminal thickness	A3		0.203	
Overall length	D		5.00	
Exposed pad length	D2	3.00	3.10	3.20
Overall width	E		5.00	
Exposed pad width	E2	3.00	3.10	3.20
Terminal width	b	0.18	0.25	0.30
Terminal length	L	0.30	0.40	0.50
Terminal-to-exposed-pad	K	0.20	-	-



(a) X-ray scan of package

(b) X-ray scan zoomed at bondwires

(c) SEM image of exposed bondwires

(d) SEM image of bond wire bonding to pad

Figure 5.2: X-ray scans with additional SEM images

The pin layout was very important for testing and orienting the chip. The same holds for the measurements of its size as it proved useful in several steps of the process and gave a solid reference

point for other size measurements. Finally, knowing how the pads were bonded to the die and how the bond wires were routed within the package meant that milling and etching could be done safely without worrying about damaging the connections.

Finally, the datasheet also offered information about temperature and voltage ranges. As the temperature range was thought to be on the conservative side, additional tests were carried out. Samples were placed on a hot plate for varying amounts of time and temperature and tested before and after. These results are shown in table 5.2. From these results, it can be concluded that both a new chip and a thinned chip survive temperatures of $200°C$ for at least three hours. This prevents further processes from being limited by the temperature budget.

Table 5.2: Temperature test results

Bulk thickness per sample [μm]	Initial test	After 3 hours and 30 minutes at $150°C$	After 3 hours at $200°C$
4	No shortcircuit current, successful flash and fuses verification	Successful verification	Successful verification
16	$0.43A$ shortcircuit current, successful flash and fuses verification	Chip could not be detected	Chip could not be detected
40	No shortcircuit current, successful flash and fuses verification	Successful verification	Successful verification
280 (new)	No shortcircuit current, successful flash and fuses verification	Successful verification	Successful verification

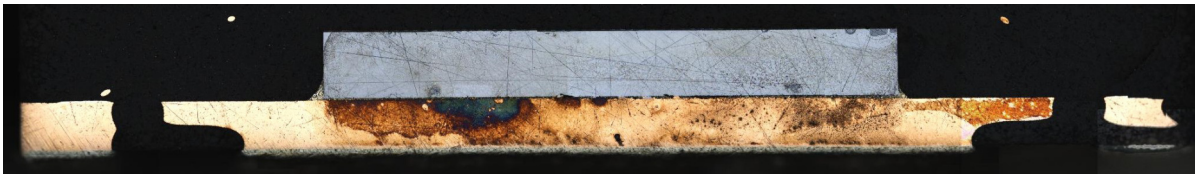
5.1.2. Obtaining cross-section

The obtained cross-section is shown in figure 5.3 with enlarged versions below. In figure 5.3a the inside of the package is clearly visible. In the bottom corners, the pads are visible consisting largely of copper with a protective layer of aluminium. In the middle sits the heatsink, having similar materials but an interesting extension underneath the package. On top sits the die with in silver the bulk. The interconnects sit on top of this bulk but are not visible yet. Between the die and heatsink is some conductive glue visible. This image confirms the assumptions from the previous chapters.

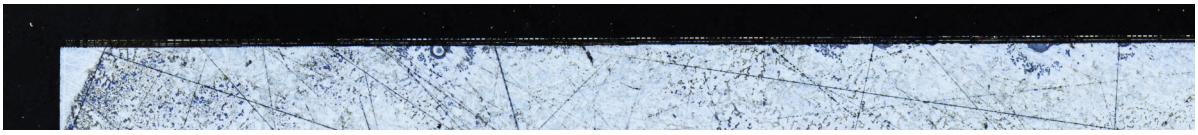
Figure 5.3b shows a zoomed-in image of the top left corner of the die. When looking closely at the top of the bulk, a faint shimmer of the interconnect layers is visible. The main purpose of this image is to give a sense of the scale of the interconnects and the bulk.

Figures 5.3c and 5.3d were made using the SEM and indicate some of the distances that are present within the chip. Most important are the thickness of the bulk and the width of the bottom interconnect. Using these images, the bulk is estimated to be around $280\mu m$ which will be used as a reference when the bulk is thinned. The width of the interconnect can be used to determine the technology node that was used for fabrication. Based on these measurements, this is likely to be the $130nm$ node. Depending on the node, different processing steps, structures and materials are used. Knowing the technology node helped define recipe steps and placed findings in context.

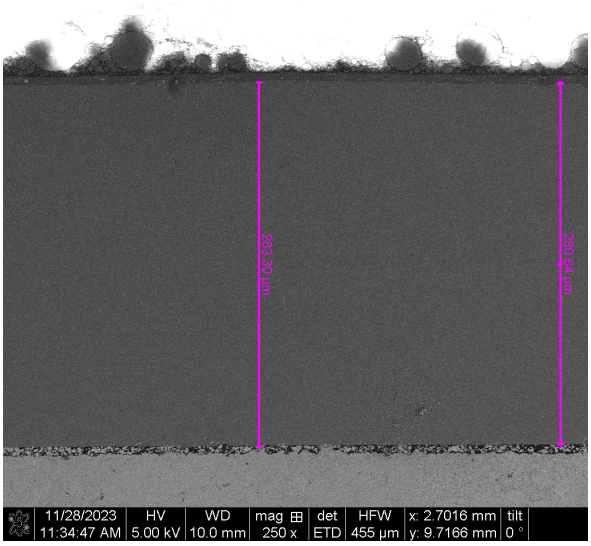
Finally, figure 5.3e shows a larger overview of the interconnects and vias. From this picture, it is clear that the die contains four metal layers. There are also hints of transistors, mainly in the form of wells from STI.



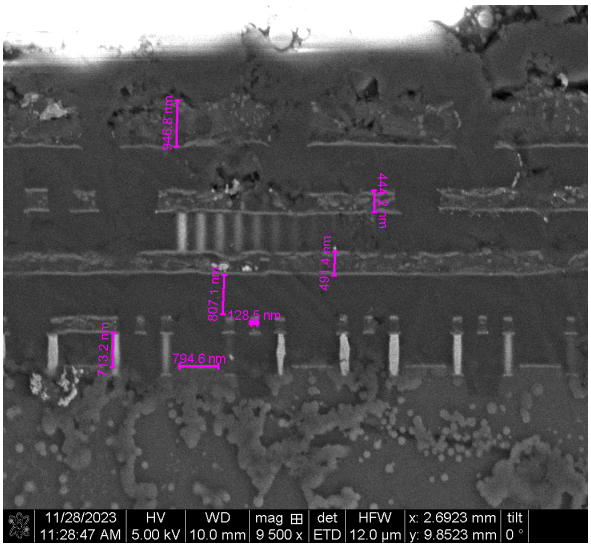
(a) Cross-section overview made with the Lext



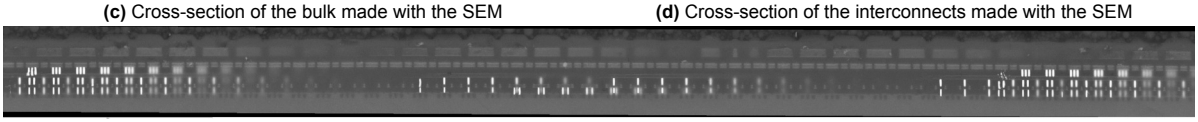
(b) Zoomed version of top left corner



(c) Cross-section of the bulk made with the SEM



(d) Cross-section of the interconnects made with the SEM

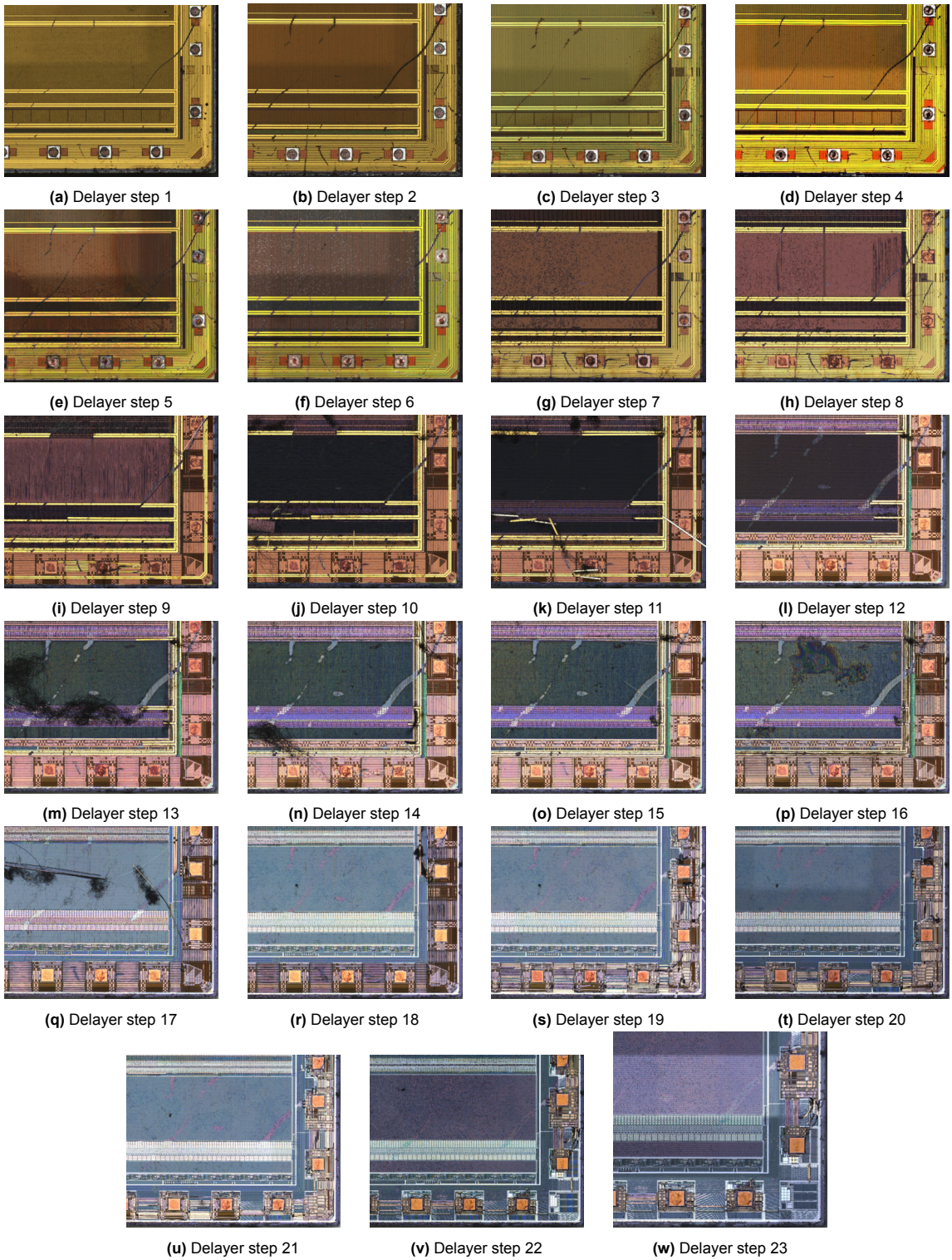


(e) Cross-section overview of interconnects made with the SEM

Figure 5.3: Crosssections made with the Lext and SEM

5.1.3. Extracting the die

The result after extracting the die can be found in figure 4.3. The die is already mounted on a stub using wax and ready for further processing. The next step was to delayer the frontside. Figure 5.4 shows the result after each intermediate step, going from the top metal layer to the bottom metal layer.



(a) Delayer step 1 **(b)** Delayer step 2 **(c)** Delayer step 3 **(d)** Delayer step 4

(e) Delayer step 5 **(f)** Delayer step 6 **(g)** Delayer step 7 **(h)** Delayer step 8

(i) Delayer step 9 **(j)** Delayer step 10 **(k)** Delayer step 11 **(l)** Delayer step 12

(m) Delayer step 13 **(n)** Delayer step 14 **(o)** Delayer step 15 **(p)** Delayer step 16

(q) Delayer step 17 **(r)** Delayer step 18 **(s)** Delayer step 19 **(t)** Delayer step 20

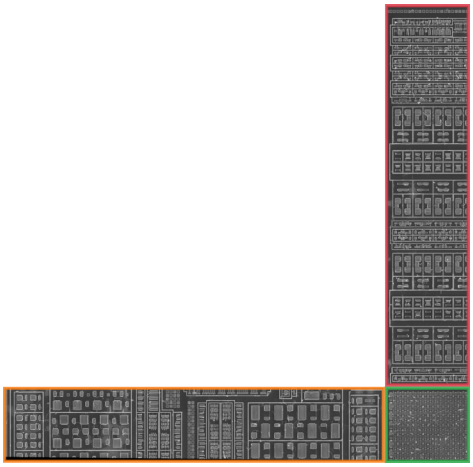
(u) Delayer step 21 **(v)** Delayer step 22 **(w)** Delayer step 23

Figure 5.4: Delayering results of the frontside lower right corner

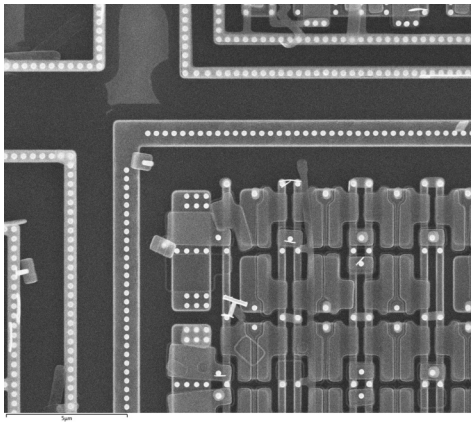
These figures gave useful insight into the layout of the chip at the different metal layers. Furthermore, from the different colours it became clear which materials were used. The slow etching in the beginning indicates the use of $SiNx$ on top as well as the yellow colour. The black colour indicates the insulator in between the traces, which are shown in purple or blue. In between some interconnects came loose which indicated the end of that metal layer. This is especially clear in figure 5.4m where a ball of hair lies on top.

This lower right corner of the die contains part of the flash, which is one of the areas of interest. At the edge, larger interconnects are visible which contain the higher power signals. Finally, the large round objects are the connection pads, thereby having all three areas of interest in one image.

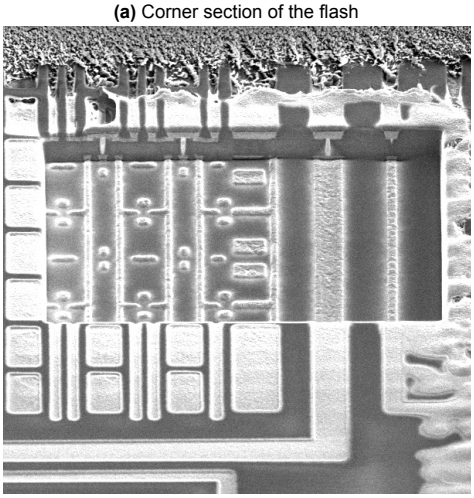
To complete the front side analysis, the die was placed inside the FIB to create more detailed images of the flash and the Vcc pad. These cross-section images were then reconstructed in a 3D environment. This made it possible to reverse engineer the signal paths of these components.



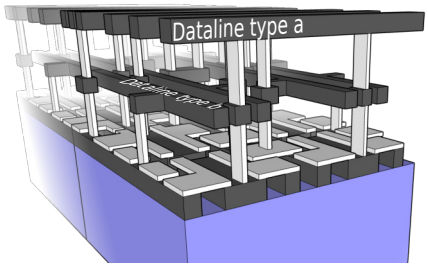
(a) Corner section of the flash



(b) Flash transistors top view



(c) Flash transistors crossection



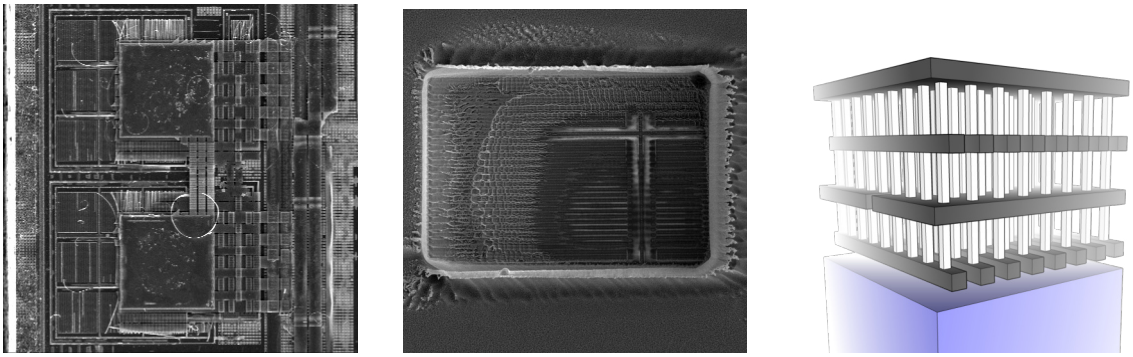
(d) Flash transistors 3D model

Figure 5.5: Flash analysis using the SEM

Figure 5.5a shows the upper left corner of the flash. Indicated in green is part of the flash where the ones and zeros are stored. These read and write operations are described in section 2.1.1 and their circuit logic is indicated in red and orange. This circuitry is repeated across the sides of the flash except for the rightmost side and operates all the individual word and bit lines.

Looking closer at the individual transistors results in figure 5.5b. Again the top left corner is shown, containing several unit elements of flash transistors. These transistors were then hollowed out to create a cross-section. This is shown in figure 5.5c with a resulting 3D model in figure 5.5d.

This model gives insight into the layout of the data paths and where the transistors are situated. From this model, it was determined that data line a is probably the word line and data line b the bit line. Accessing one of these lines would be the ultimate goal of the backside contacting methodology. For this project, however, the focus was first shifted towards the Vcc connection pad.



(a) Vcc pad overview **(b)** Vcc pad cross-section **(c)** Vcc pad 3D model

Figure 5.6: Vcc pad analysis using the SEM

From the cross-sections and model shown in figure 5.6, it is clear that the Vcc structure underneath the connection pad is built for good conduction. The structure consists of many vias and is full of traces or even larger surfaces. Furthermore, it is insulated from the bulk and relatively large. Finally, going too deep into the structure will not result in limited functionality of the chip, meaning it is an ideal location for attempting backside contacting.

5.1.4. Exposing backside

In this section, the results of exposing the backside and thinning the bulk are discussed. This process was arguably the most challenging as it caused many short-circuited samples. Around 40 samples were processed and divided over two batches. The results are separated for each step to better pinpoint potential causes of failure.

Heatsink

Two batches of chips were made. The first contained 36 chips, the second 20 chips. Figure 5.7 shows the results of the second batch. There are several observations that are important to note.

First, a good indication of whether the chemical reaction is taking place is the change of colour of the etchant from transparent to light green, see figure 5.7a. Furthermore, small bubbles of gas can be observed, forming on the heatsink and rising up.

The second observation is shown in figure 5.7b, where there is a clear difference in etch rate between samples. Fortunately, the etchant stops at the die meaning this is not really a problem and longer etching will resolve this issue.

Looking at the last two figures 5.7c and 5.7d, this is indeed the case. However, increasing the etching time does result in some side effects. The first is the limited protection of the dentist paste. In figure 5.7c one of the chips got exposed because the dentist paste was no longer sticking to the glass and chip surface. As a result, all the connection pads were etched away, rendering the sample useless. Secondly, even if the dentist paste remained in place, due to the long time span, the etchant was able to find its way to most of the pads anyway. Additionally, over time, the etchant was able to etch some of the wax used to keep the chips in place. The fact that this is related to the etching time was confirmed by the fact that the first batch was etched only half as long as the second and did contain significantly fewer side effects. Meaning, that only a handful of pins were damaged but not completely removed and all the chips and paste stayed in place.

That being said, the longer etching removed all the copper without leaving residues which proved to be a significant benefit during the thinning process. In order to remove all the copper, two additional observations and resulting adjustments were made.

First, the etch rate of copper seemed much higher than that of aluminium, resulting in under-etching. This in turn resulted in a floating thin aluminium film with some copper attached as residue. Removing these pieces using a toothpick meant that the etch rate increased and that it took longer for the etchant to be saturated.

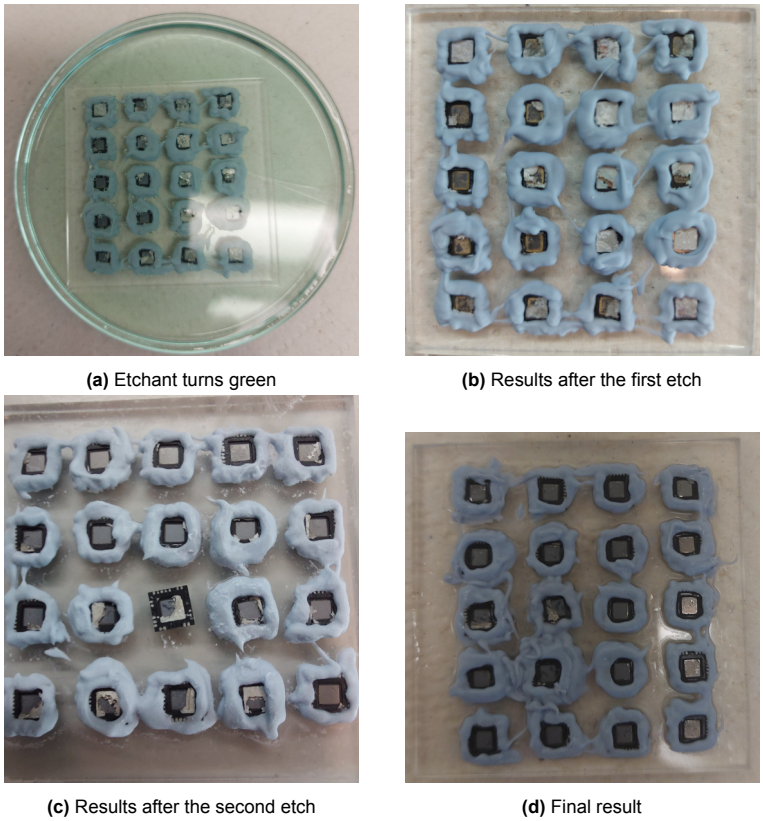
Secondly, due to the shape of the heatsink, the bottom edge is protected by the packaging leaving a small ring of copper that is slow to etch. By carrying out a small milling step before the final etch, this

packaging can be removed exposing the remaining copper. This final adjustment made sure all the copper was removed and allowed for visual confirmation.

Finally, some adjustments were suggested to improve the etch rate and reduce the amount of etchant used. Starting with the possibility of adding flow in the etchant as it was observed that some reactants were depositing back on the heatsink, thereby shielding it from further etching. A flow might also help mitigate the different etch rates across the sample. However, such a device was not available in the lab. In an attempt to prevent the deposition, the samples were placed upside down in the etchant. This, however, created air bubbles underneath the heatsink limiting the chemical reaction altogether. For further research, an approach with a small beaker could be tried out where the samples are placed on a small strip of glass which is placed vertically.

These results and observations can be summarized in the following list.

- Green etchant and bubbles are good indicators that the chemical reaction has taken place.
- Different etch rates are not problematic as long as the connection pads are well-protected.
- The protection should be able to withstand long etching times while being removable.
- Enough wax should be used to keep the chips in place.
- Residue should be removed in between etching steps to improve efficiency.
- The overlapping packaging should be milled before the last etching step.
- Removing all copper is necessary to ensure the correct milling depth can be achieved.



(a) Etchant turns green

(b) Results after the first etch

(c) Results after the second etch

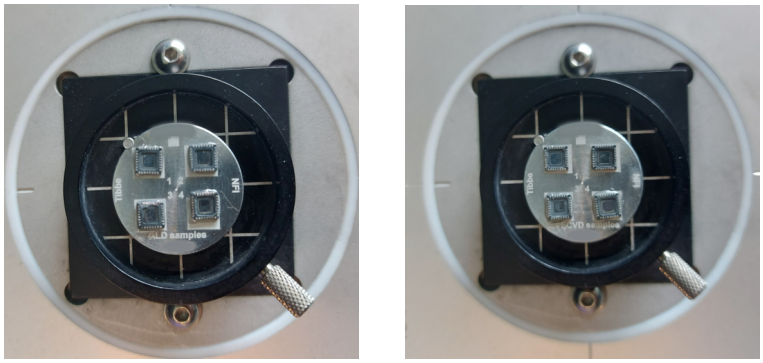
(d) Final result

Figure 5.7: Results of the heatsink removal process

Glue removal

The removal of the glue using a cotton swap and acetone was straightforward. As the bulk is still around $280\mu m$ and inside the package it can withstand all the forces that are applied during this step. After the glue is removed, the samples are tested in the probe station. It should be noted that at this point, all the samples that had the six necessary pads still intact and thus were able to be tested all passed successfully. The samples with the most intact pads were selected and placed on a stub. Some of the broken pads might be repairable, however, in many cases, entire pads were missing which would

be almost impossible to repair. Figure 5.8 shows two stubs each containing four samples with their heatsink removed, no copper left and a shiny backside of the bulk.

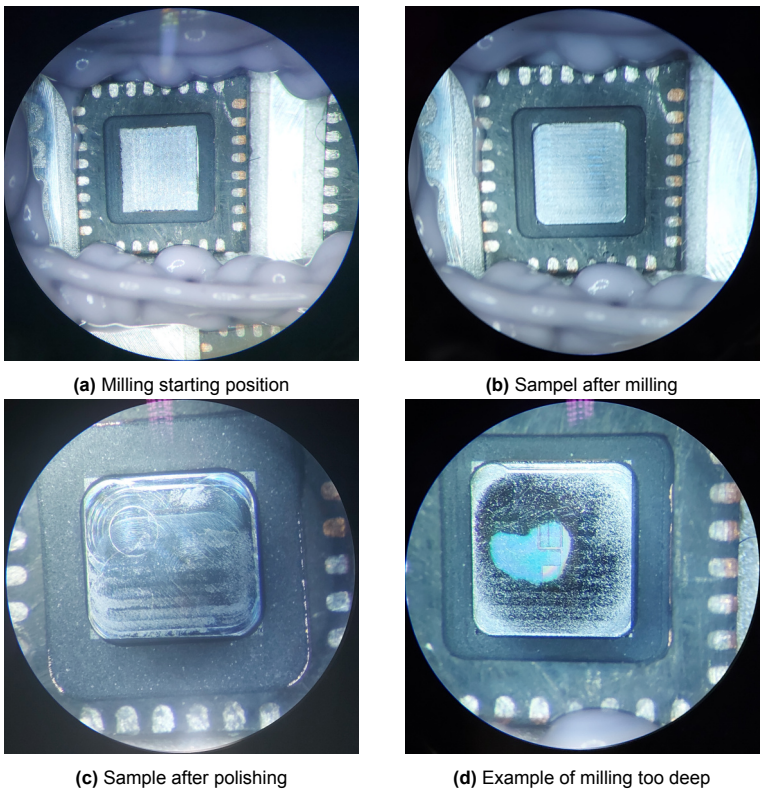


(a) Four samples for ALD **(b)** Four samples for ICPECVD

Figure 5.8: Stubs ready for bulk thinning

Milling

The results of the milling step can be divided into three separate subjects; milling and polishing, testing the Si thickness and short circuit problem analysis. First, figure 5.9 shows the start, milling and polishing results. The fourth figure shows what it looks like when the mill mills too deep. As the first batch contained some copper residue which clogged the mill, introducing inaccuracies, this happened in every other sample. This inconsistency also resulted in smaller step sizes and many tests in between. With all the copper removed in the second batch, these problems were resolved and only a singular precise mill was carried out.



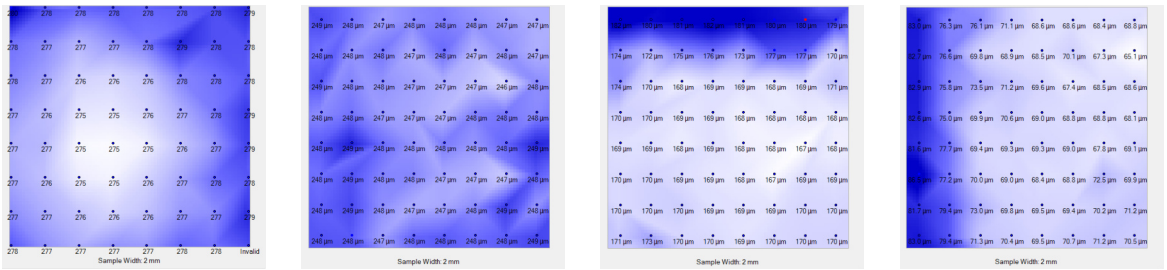
(a) Milling starting position **(b)** Sampel after milling

(c) Sample after polishing **(d)** Example of milling too deep

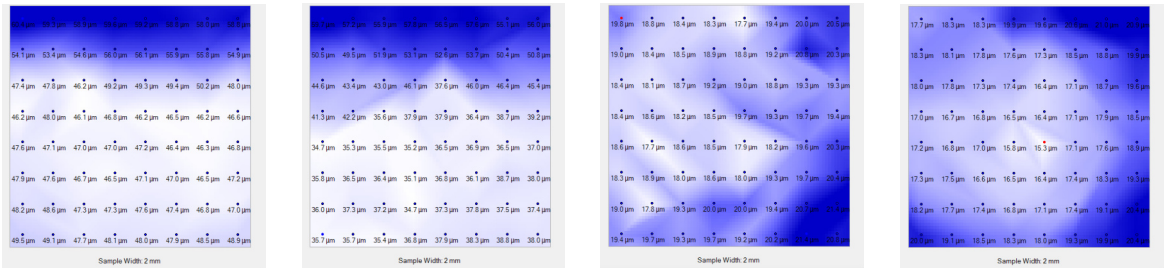
Figure 5.9: Milling steps results

In between milling steps, height maps were created using the X-vision which are shown in figure 5.10 and 5.11. The first shows how the process started, with many small steps. The second indicates

the reduction in steps to the bare minimum. At the start of each sample, a baseline measurement was taken. The values resulting from these measurements varied a lot because of the rough surface. This meant that they were only used as a reference and could not be used as exact measurements. When a measurement point states *invalid* it means that the embedded statistical software already flags the measurement as inaccurate.

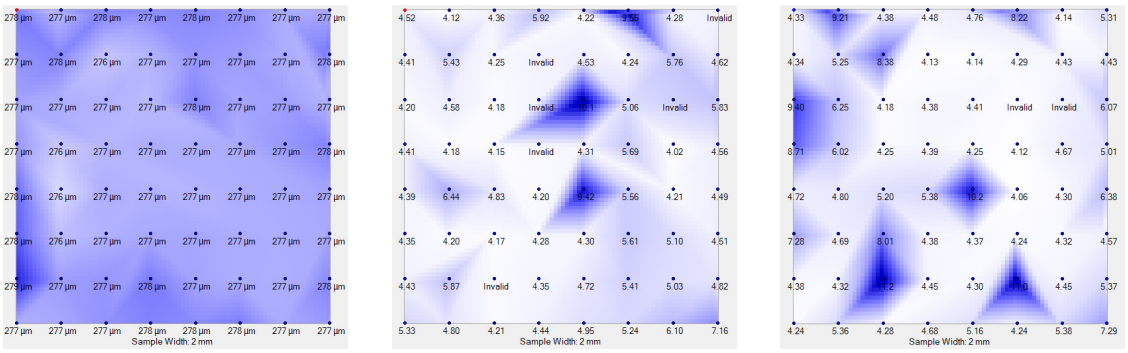


(a) Baseline measurement **(b)** Height after removal $\sim 50\mu m$ **(c)** Height after removal $\sim 100\mu m$ **(d)** Height after removal $\sim 100\mu m$



(e) Height after removal $\sim 25\mu m$ **(f)** Height after removal $\sim 15\mu m$ **(g)** Height after 2 steps of polishing **(h)** Height after final polish

Figure 5.10: Si height test results



(a) Baseline measurment **(b)** After the removal of $\sim 270\mu m$ **(c)** After polishing

Figure 5.11: Si height test results of the improved process

Comparing the two figures shows that the improved process is faster, contains fewer steps and is therefore less prone to error. Additionally, the final thickness is considerably thinner. This, however, did not solve the problem of the die short-circuiting, which will be discussed next.

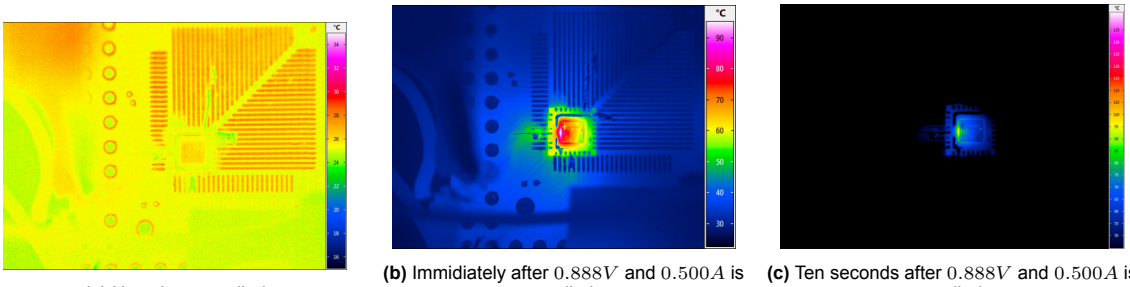
Short circuit problem analysis

At this point, the thinned chips were placed under the probe station to test their functionality. It turned out, that only half of the chips survived the thinning of the bulk. The results are shown in table 5.3. This table also shows that chips with a thickness of around $5\mu m$ almost always contain a short. Additionally, the initial resistance is incredibly low compared to the standard resistance of $5.3M\Omega$. Since a significant number of samples suffer from short circuits and part of the challenge is to make the recipe as reliable as possible, a more detailed analysis was carried out to find out what caused these chips to break down.

Table 5.3: Bulk thinning resistance and verification results

Bulk thickness per sample [μm]	Resistance between Vcc and Gnd before power is connected [Ω]	Functionality test
~ 278 (new)	$5.3M$	OK, flash and fuses
~ 80	-	OK, flash and fuses
~ 40	-	OK, flash and fuses
~ 15	-	OK, flash and fuses
$5 - 20$	130	OK, flash and fuses
8.44	$0.68M$	OK, flash and fuses
~ 4	130	OK, flash and fuses
$5 - 15$	7	BAD flash, OK fuses @ $0.25A$ shortcircuit current
$10 - 24$	-	BAD, shorted
~ 10	3.55	BAD, shorted
10.0	8.8	BAD, shorted
$5 - 10$	-	BAD, shorted
5.19	$0.67k$	BAD, shorted
5.05	$1k$	BAD, shorted
~ 5	2.5	BAD, shorted
5.0	13.5	BAD, shorted
4.79	$2k$	BAD, shorted
4.7	5.8	BAD, shorted

First, it was found that the immediate cause of the breakdown was a short between the Vcc and Gnd rails of the chip. To determine possible causes of this short, first, the location of the short should be determined. This was done using the thermal camera as discussed in section 4.2.5 resulting in figure 5.12.



(a) No voltage applied

(b) Immidiately after $0.888V$ and $0.500A$ is applied

(c) Ten seconds after $0.888V$ and $0.500A$ is applied

Figure 5.12: IR heat camera results

The spot where most of the heat is concentrated is right on the edge of the die, between the Vcc and Gnd connection pads. At this point, the main suspect for causing these shorts was the diffusion of the copper residue into the thin Si bulk. This would explain why some samples were 'lucky' while most samples were not. By milling with a contaminated mill this diffusion process would be amplified even more.

However, by removing all the copper in the second batch, this hypothesis could be ruled out while the shorts persisted. Possible explanations for the short now became more complex to solve. The following options were considered. Thinning the bulk to a few μm might destroy some unknown functionality of the bulk. Samples that are closer to $10\mu m$ seem to survive more often than those that are thinned to $5\mu m$ or lower. Another possibility is that at some point during the milling, the sample is charged, thereby destroying either a electrostatic discharge (ESD) protection circuit or creating a latch-up. Solving this would require grounding the mill and careful cleaning.

Concerning the charging possibility, a short analysis was done on the impact of the polishing steps. This turned out to have no effect. Samples that worked before the polishing also worked afterwards.

Finally, to determine whether a latch-up was created during the milling a resistance measurement was carried out. In here the multimeter from section 4.2.3 was connected to the Vcc and Gnd pad. The

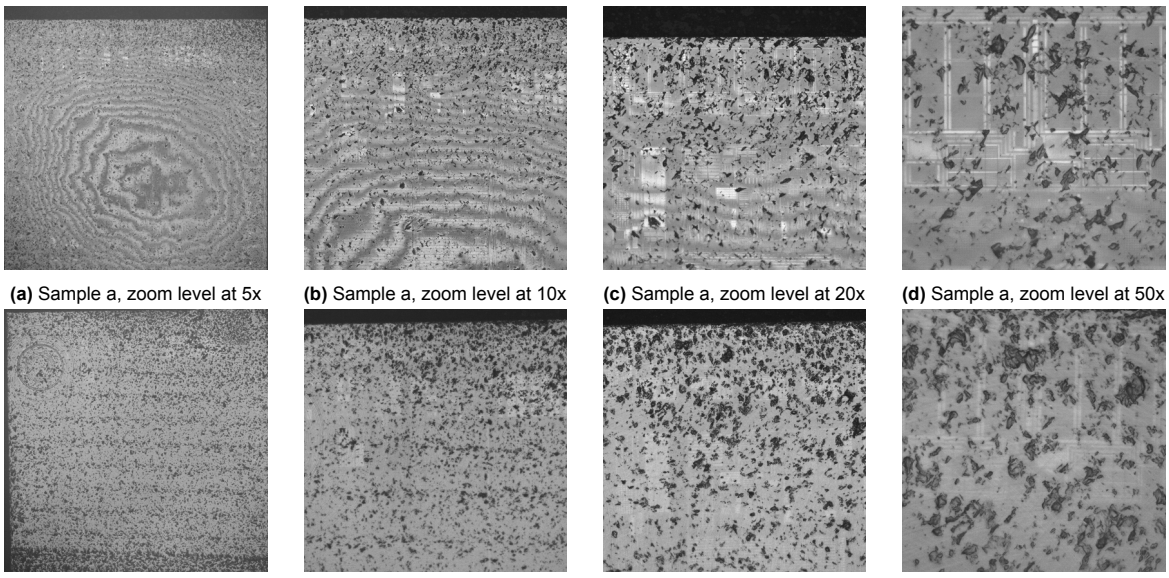
found resistances are shown in table 5.3. The samples that were shorted all had resistances around $1 - 5\Omega$. Samples which had more resistance fell back to this value when around $0.6A$ was applied. Once a sample was observed having such a resistance it never went back to its old value.

Further research will have to determine to what extent the bulk height has an impact on the creation of these shorts or if it occurs due to charging. Additionally, doing tests on other chips might indicate that it is the design of this chip that makes it more prone to these shorts.

5.1.5. Determine areas of interest

Using all the results shown in this section, it was found that the Vcc pad was the most suitable for the first attempt of a backside contact. After a successful connection was made to the Vcc pad further steps could be taken to either move to the next area of interest or improve for example the resistivity. However, as will become clear later, a complete successful cycle of the recipe is already challenging.

With the area of interest verified, NIR images were made for more information regarding the layout but mainly for navigation within the SEM and FIB. The results of two samples are shown in figure 5.13. The IR image is not completely equal to that of the SEM which meant that multiple images with increased zoom levels (5, 10, 20 and 50x) made navigating significantly easier. These images were made for each sample that needed processing in the SEM due to the need for the unique pattern of residue.



(a) Sample a, zoom level at 5x **(b)** Sample a, zoom level at 10x **(c)** Sample a, zoom level at 20x **(d)** Sample a, zoom level at 50x

(e) Sample b, zoom level at 5x **(f)** Sample b, zoom level at 10x **(g)** Sample b, zoom level at 20x **(h)** Sample b, zoom level at 50x

Figure 5.13: NIR images

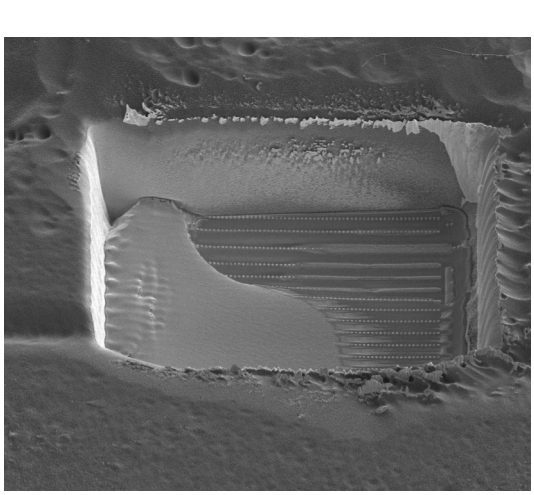
5.2. Invasive processing

This section covers all the results obtained during the creation phase. This phase is mostly carried out within the SEM and FIB which meant that the process variations and adaptability were limited to the tools' abilities. In general, it was observed that although the process is carried out behind the computer, a lot of manual action was required to steer the process in the right direction. Furthermore, the use of copper tape to ensure proper conduction is mandatory as without the ion and electron beams become unstable. The results discussed in the next section were obtained after multiple days of practice and testing.

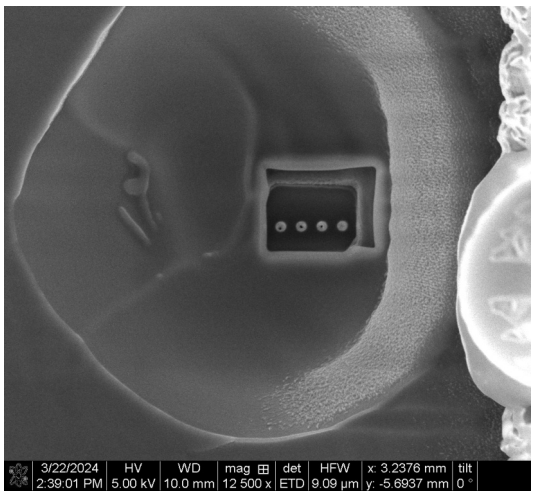
5.2.1. Create holes

The most challenging part of creating the holes was knowing when to stop. Ideally, this is right above the interconnect. In practice, however, this proved to be difficult because of two reasons. First of all, the thickness of the remaining bulk fluctuates multiple μm as is indicated in table 5.3. This makes defining a standard process impossible. Secondly, the traces only appear on the live feed when they are already exposed.

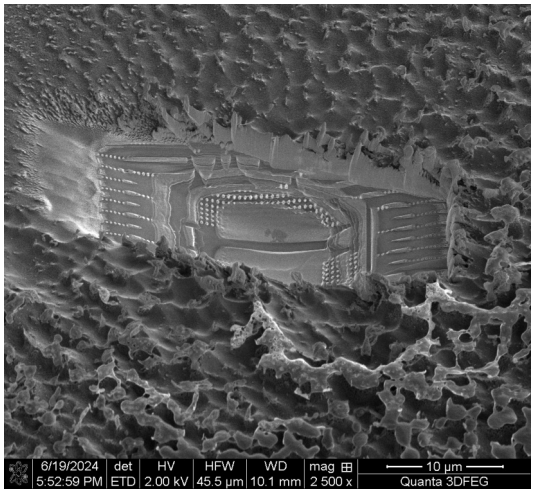
Figure 5.14 shows four examples of holes. In figures 5.14a and 5.14c the hole is too deep and the interconnect and vias are exposed. Figures 5.14b and 5.14d show the correct execution. In figure 5.14b, a larger hole was made with IEE which was stopped right above the interconnects. A more precise and clean hole was then made using only Ga ions. In the final figure a small hole was made directly to the interconnects. This was only possible because the process was adjusted based on the incorrect hole in figure 5.14d



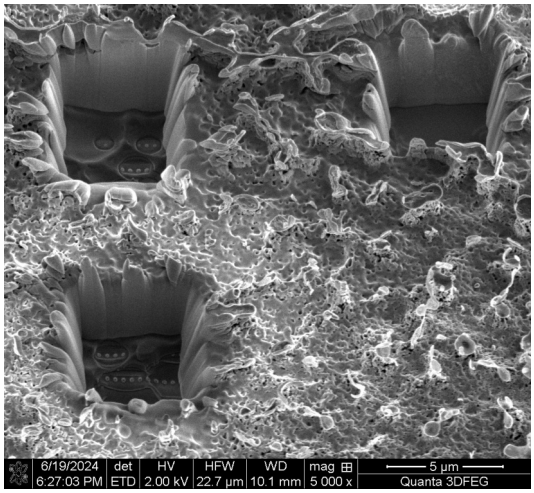
(a) Hole too deep, example 1



(b) Correct hole, example 1



(c) Hole too deep, example 2



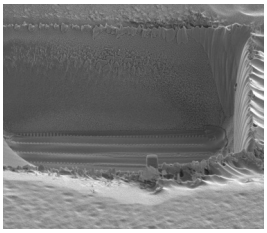
(d) Correct hole, example 2

Figure 5.14: Holes to the Vcc connection pad

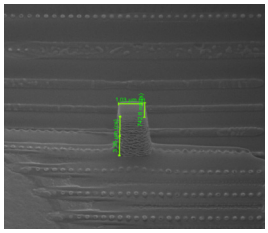
5.2.2. Create pillars and deposit TEOS

The pillars were created using Pt with the volume defined to be $1\mu m$x$0.5\mu m$x$9\mu m$. In practice, this resulted in pillars of $1\mu m$x$0.7\mu m$x$3.35\mu m$, meaning about one-third of the height and a bit wider. This is likely due to incorrect calibration of the software and wear and tear of the aperture used to limit the source ions. Nevertheless, these pillars fit the size requirements and are still visible after the TEOS deposition.

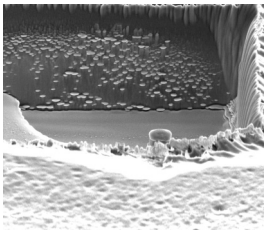
Figure 5.15 and 5.16 each show a sample on which a pillar is built. Afterwards, these pillars and holes are covered with TEOS. It was later found that by using polygons the pillar could be excluded from the TEOS deposition, however, this was not tried during this project. The TEOS layer is clearly visible but has a rough surface, especially at the vertical edges. This could become problematic when traces are deposited for wire bonding the pillar to the outside.



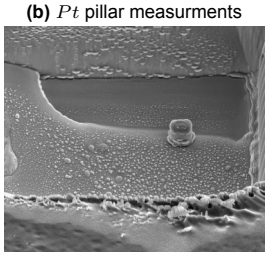
(a) Pt pillar overview



(b) Pt pillar measurments

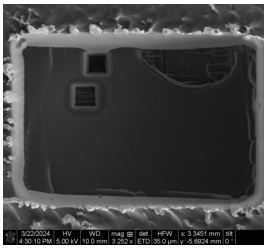


(c) TEOS deposition overview

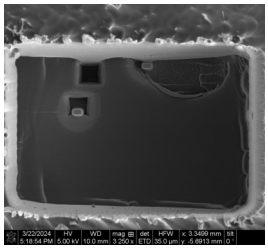


(d) TEOS deposition at a different angle

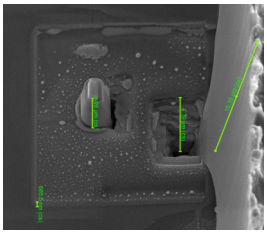
Figure 5.15: Pillar and TEOS of sample a



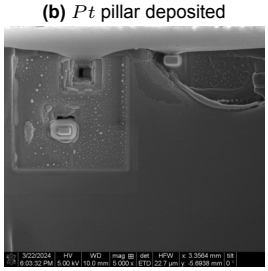
(a) Holes overview



(b) Pt pillar deposited



(c) TEOS deposition measurments



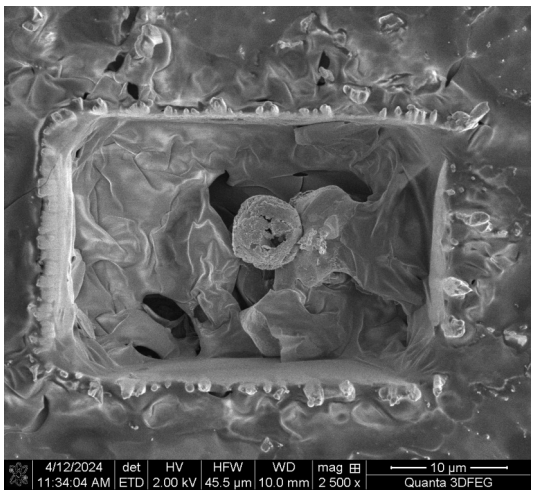
(d) TEOS deposition overview

Figure 5.16: Pillar and TEOS of sample b

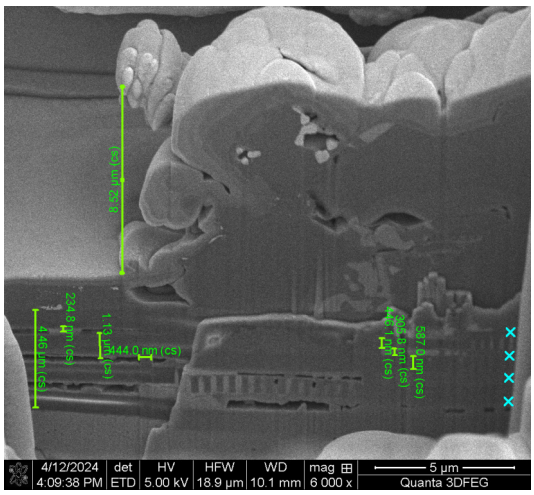
In conclusion, depositing the pillars and TEOS is relatively straightforward. The most important is to connect the sample to the stage to prevent charging and to make sure the pillar is high enough. Further tests could be done using tungsten as it should have a lower resistance but requires even better grounding of the sample. Lower resistance might be required if the driving strength is too low or voltage amplitudes are too low to be detected.

5.2.3. Deposite insulation

The samples from figures 5.15 and 5.16 were brought to the cleanroom and placed in the ICPECVD tool and ALD tool respectively. This resulted in the overviews and cross-sections shown in figure 5.17 for ALD and figure 5.18 for ICPECVD.

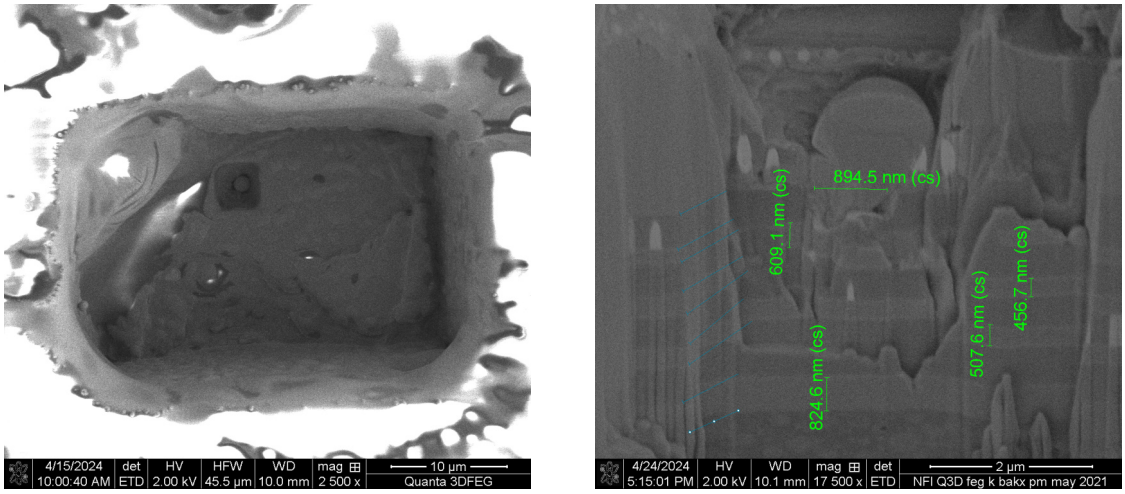


(a) Overview after $AlOx$ ALD



(b) Cross-section after $AlOx$ ALD

Figure 5.17: $AlOx$ ALD results



(a) Overview after ICPECVD

(b) Cross-section after ICPECVD

Figure 5.18: SiO_2 ICPECVD results

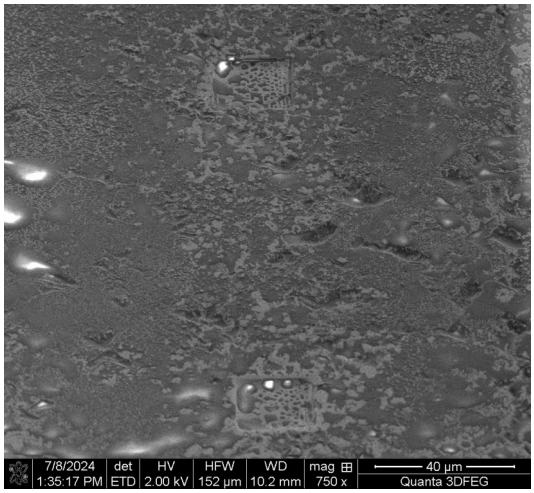
Starting with the ALD results in figure 5.17, it can be observed that this deposition did not go as expected. Several abnormalities were noted. First, the overview shows some flakes, tears and loose parts. Additionally, holes are observed in the cross-section where interconnects would be expected. The cross-section also shows a thin outer layer which indicates some insulator was deposited. The pillar, however, was not found during the entire cross-section.

These observations left the question of how the interconnects disappeared. It could not have been due to the cross-section process as this was done before for the flash and Vcc pad without any problems, see figures 5.5c and 5.6b. Previous tests with Pt and TEOS deposition meant that those processes could also be excluded. That left the ALD process as the main cause.

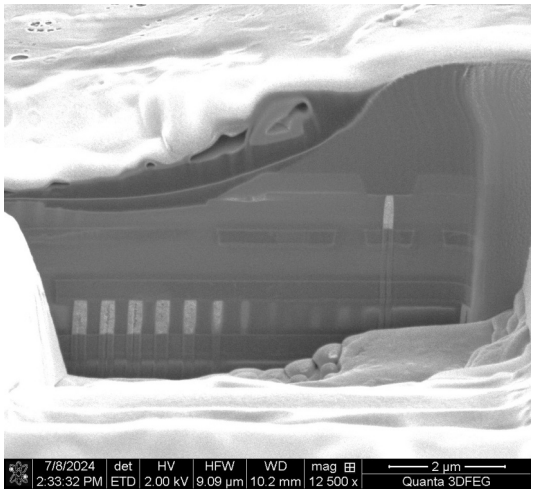
During ALD, Al_2O_3 is grown by exposing the sample alternating between H_2O and TMA ($Al(CH_3)_3$). Normally this is done on a relatively flat wafer which is heated up by the stage. The sample however contains a significant cavity and is thermally insulated from the stage by the remaining package. Together this might result in incomplete layers which in turn can lead to a hyperbolic reaction once the sample is taken out of the high vacuum. This might explain the artefacts seen in figure 5.17. Another option could be that some compounds that resulted from an irregular deposition reacted with the ions from cross-sectioning thereby removing some of the interconnects.

A possible solution would be to increase the temperature from $105°C$ to $150°C$. Alternatively, a layer of $HfOx$ could be deposited before a layer of $AlOx$ is deposited. The results of the latter are shown in figure 5.19. From the white reflections, it is clear that this process was much more successful in insulating the bulk. For a clean cross-section a layer of platina was deposited. This is indicated in figure 5.19b. Between the platina and the bulk is a thin gray line which would be the insulation layer. As it looks intact and relatively uniform it can be concluded that this deposition was indeed an improvement.

Finally, the fact that all pillars have disappeared. This indicates that creating the pillars after the insulation ALD deposition will be more reliable. On top of that, the addition of TEOS has little to no effect on the whole process and insulation characteristics and could therefore be omitted.



(a) Overview after $HfOx$ and $AlOx$ ALD



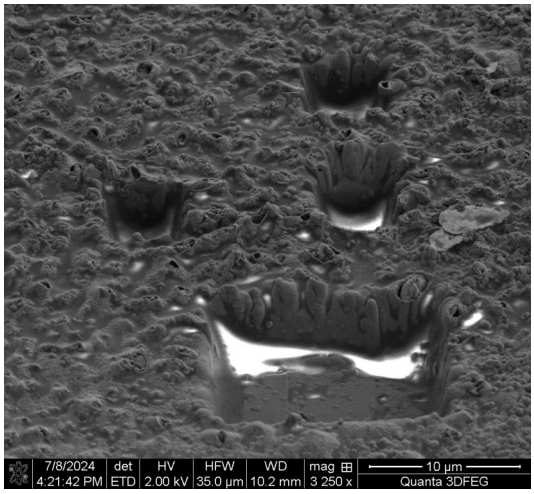
(b) Cross-section after $HfOx$ and $AlOx$ ALD

Figure 5.19: Improved ALD results

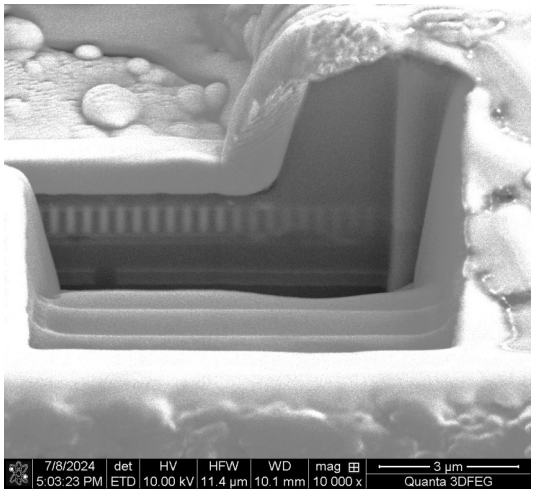
Figure 5.18 shows the results after ICPECVD. Comparing the overview with the ALD overview of figure 5.17a immediately shows that there is a lot more charging happening. This indicates that the sample surface is well insulated and a poor conductor. Furthermore, this layer should be a factor of 10x thicker making further processing less prone to break this layer. This is indeed the case as is visible in the cross-section by the thick light gray layer on top of the bottom interconnects. Unfortunately, also in the ICPECVD modified samples, the cross-section did not show any pillars. Meaning that they disappeared either by lack of adhesion or were destroyed during the deposition.

A second test was performed using SiN to see if this resulted in a smoother surface which in turn could improve the quality of the wire bond pads. These results are shown in figure 5.20. As this sample did not contain any TEOS or pillars it is a lot cleaner. The black line that appears as a shadow between the top layer and the bulk is the insulation layer. It looks uniform and intact. The top layer is platina, used to make a clean cross-section but also to indicate how the pad for wire bonding might be located. These results do not indicate a definitive preference for either SiO_2 or SiN. Further tests must determine which would be better insulating, more durable or better matched with other processes.

In conclusion, both insulation depositions were able to cover the steep walls of the hole and deposit a layer at the bottom. Although ICPECVD proved to be the better option, getting the ALD to work might be needed when the hole sizes are reduced since it should have higher conformity. Finally, having clean holes without pillars and TEOS deposition reduced the artefacts of the insulator deposition and improved the quality overall.



(a) Overview after SiN ICPECVD



(b) Cross-section after SiN ICPECVD

Figure 5.20: Alternative ICPECVD results

Electrical tests were performed after the samples were coated with the insulator. These results are shown in table 5.4. First, the resistance between Vcc and Gnd was measured. After the voltage source was set at $3.5V$ and connected to the chip another resistance measurement was carried out. Finally, the AVRmkII was attached to determine the chip functionality. As the pads were also covered with an insulator they had to be cleaned. This was done using a scalpel which was a fast and effective method.

Table 5.4: Vcc to Gnd resistance and functionality test results after insulation deposition

#Sample	Description	Initial R	R after power connected	Functionality test
1	ALD	21Ω	138Ω	OK
2	ALD, pillar + TEOS	1Ω	1Ω	BAD, short
3	ICPECVD	11Ω	1Ω	BAD, short
4	ICPECVD, pillar + TEOS	1Ω	1Ω	BAD, short
5	ICPECVD	$0.13k\Omega$	78Ω	BAD, no detection

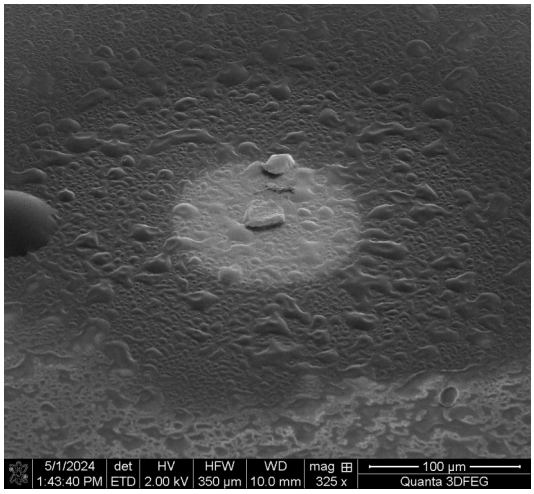
The samples were only tested after the hole was made and the ALD or ICPECVD deposition was applied. This means the cause of the malfunction can originate from several different sources. For example, creating a hole can already damage the structure, reducing the internal resistance to the point that a short is formed. Adding a pillar and TEOS before insulator deposition also seems to have a negative influence. Furthermore, from this limited sample size, it could be concluded that the ALD or ICPECVD do not affect the possibility of a short as in both cases high resistances were measured after the deposition. In conclusion, these results show a significantly reduced resistance, compared to the default resistance of $5.3M\Omega$ but they also prove that it is possible to have a functioning chip after an ALD deposition.

5.3. Establish external connection

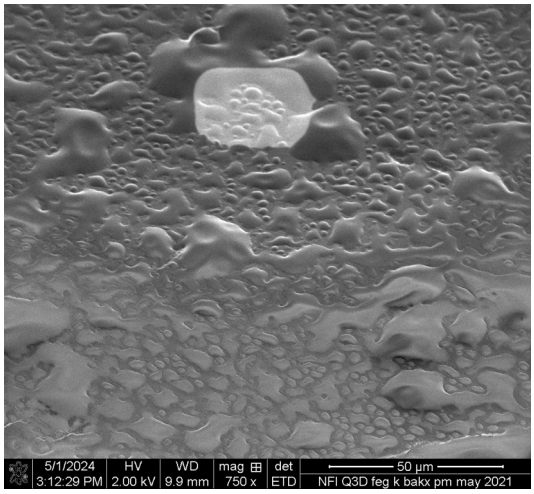
This section discusses the results of the final phase of the recipe. As only one sample survived all the previous steps it was decided to disconnect this phase from the previous two. In practice, this meant that the traces were not yet connected to the pillars.

5.3.1. Create pads and traces

Both tungsten and platina were tried out to create the pads and traces for the wirebonding. The depositions shown in figures 5.21 and 5.22 were done on a layer of $SiOx$ deposited using ICPECVD to make the test as representative as possible.

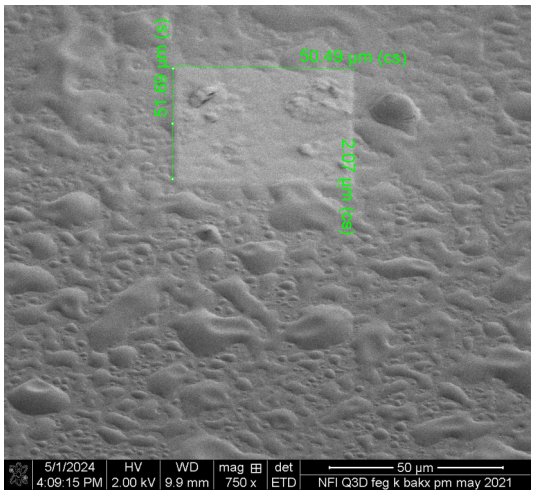


(a) Overview after first W deposition

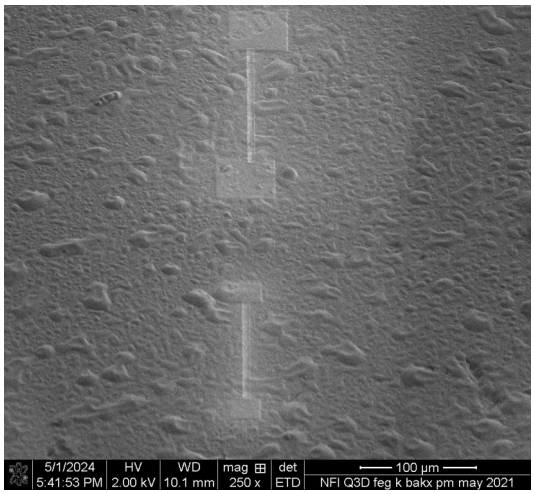


(b) Overview after many attempts of W deposition

Figure 5.21: Results of tungsten deposition of traces and pads



(a) Deposition of Pt pad



(b) Overview of Pt deposition of pads and traces

Figure 5.22: Results of platina deposition of traces and pads

Figure 5.21 shows the attempt to deposit tungsten. As is clear, there is no deposition but also no milling. This is remarkable since tungsten deposition requires highly energised ions which without tungsten atoms gas would mill easily through the insulator. Testing the tungsten deposition on a standard sample proved that the setup was working correctly and the insulation was too high for this technique to be feasible.

Fortunately, the Pt deposition was successful as shown in figure 5.22. Two traces connecting four pads were created. Further tests like pulling on a connected bond wire showed that their adhesion to the surface is good. Additionally, the roughness of the surface did not impact the process. Note that the input height was three times the required height as the deposition also in this case was less than the software predicted.

The expected resistance of these structures can be calculated using equation 4.1. Using a length of 200μ, a width of 5.5μ, a height of $2\mu m$ and $\rho_{Pt} \approx 15\Omega$ a resistance of $R_{expected} = 273\Omega$ was found. This can vary a lot depending for example on the quality of the deposition and the roughness of the surface.

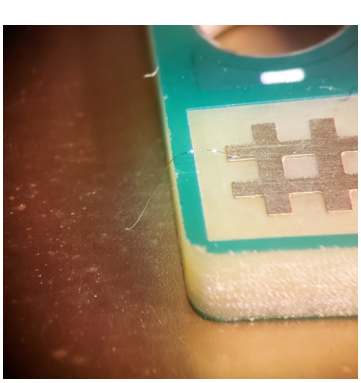
Although this gives some insight into the expected resistances, further tests like depositing the Van der Pauw structure could help determine these more precisely. Additionally, grounding the insulation layer closely to the to-be-deposited area might make tungsten deposition possible. For now, this depo-

sition met all the requirements needed to move to the next step; wire bonding.

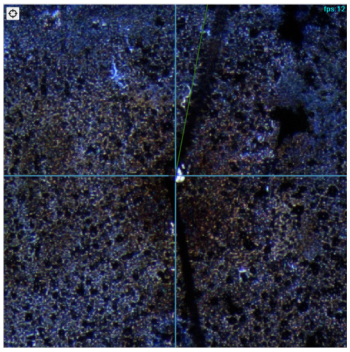
5.3.2. Wire bonding

The wire bonding was surprisingly straightforward. There were two limiting factors; the movement of the needle and the resolution of the camera. After some tests regarding different heights and curved connections, an attempt was made on the deposited pads. It was expected that the bonding attachment would be too weak because the platina pad and chip would absorb most of the ultrasonic energy. This turned out to not be the case and a strong wedgebond could be made. Furthermore, after removing the bond, it was possible to create a new bond at the same place. Attempts at bonding on the insulator failed, indicating the need for a bonding pad.

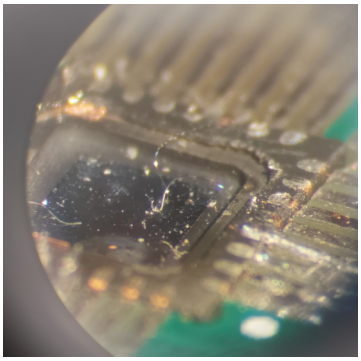
Figure 5.23 shows a height test, the wedge bond and a successfully bonded wire. The quality of the bond was also recorded by the software and displayed in a graph, see figure 5.23d. The applied force and ultrasonic power were adjusted to make sure the graph ended in the green area. This would indicate a strong bond which turned out to be an accurate prediction.



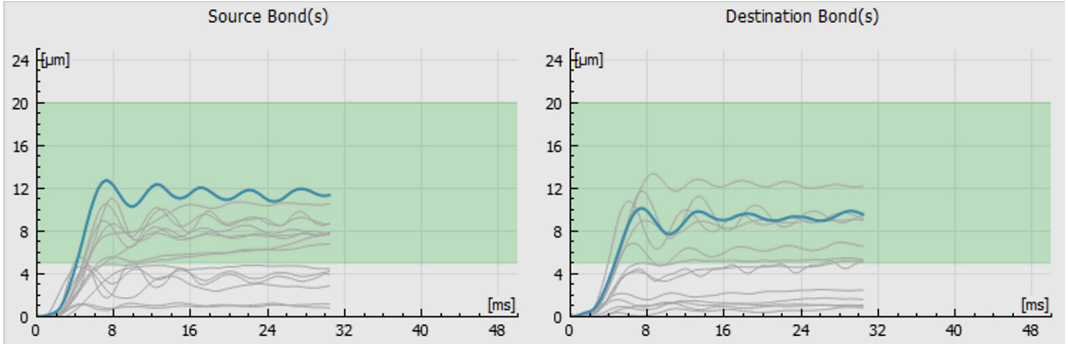
(a) Test bond in ideal situation



(b) Wire bond on connection pad viewed from the camera



(c) Two successful bonds from the deposited pads to the outer edge



(d) Grahps that indicate the quality of the bond

Figure 5.23: Results of the bond wirering

Finally, some resistance measurements were done using the Fluke multimeter. Using equation 4.2 the theoretical resistance for an aluminium bond wire with a length of $3mm$ and a diameter of $17.5\mu m$ was found to be $R_{expected} = 0.33\Omega$. When measured this turned out to be $\sim 0.68\Omega$, which is in the same order of magnitude. Measuring the resistance across two bond wires each connected to the 50x$50\mu m$ pad at the end of the trace resulted in a resistance of $86k\Omega$. This is significantly larger than the combination of the 273Ω found in section 5.3.1 for the pads and traces and the $\sim 1.36\Omega$ for the bond wires. More experiments and more precise measurements should be carried out to determine what causes this high resistance.

6

Conclusion and future work

This thesis discussed the different techniques that are used in chip manufacturing and looked at the possibilities in which they could be used in reverse engineering and chip forensics. Based on these findings a concept recipe was made to create a backside contact. This concept was then developed into a recipe containing multiple phases, processes and steps. Each process and step was then tested and if deemed necessary iterated upon to improve the yield and quality.

By doing so, the practical objectives of getting to know the available techniques and technology for chip forensics and getting to know the chip fabrication technology and layout were met. By designing and testing the recipe, the proof of concept of the method of backside contacting was given. In general, it can therefore be concluded that the method of backside contacting is a promising new technique within the field of chip forensics. That being said, there are three obstacles which need to be considered before this method can be applied in real scenarios. These are listed below.

- With new technologies aiming to use the bulk for BPDN, the way to the bottom interconnect may become obstructed and aspect ratios may become too high.
- Research into short circuits and other possible artifacts, when the bulk is thinned. To understand the origin and subsequently increase the reliability and yield.
- This thesis only showed that the individual steps are possible. Going through the whole recipe, the longest a chip survived was up to and including ALD. This means a complete cycle is yet to be achieved.

The next section will discuss the recipe in more detail. Addressing processes that went well and the ones that need improving. Furthermore, some shortcomings are mentioned regarding techniques and testing setups. In the final section, some additional recommendations are given concerning future research and experiments.

6.1. Recipe review

The proposed recipe, shown in figure 3.4, is meant to be the proof of concept. The workflow and structure of the recipe turned out to be great in avoiding delays and providing clear loops and processes to iterate upon. The various phases provided clarity regarding when the targets were achieved and when additional work was required. Additionally, each process provided clear benchmarks and results against which future improvements could be compared. Finally, this structure allowed for easy incorporation of additional experiments like RIE.

6.1.1. Acquisistion of chip layout and technology

The acquisition phase was a success. This means all the necessary information was gathered and a good understanding of the chip layout, technology and operation was developed. As it was the first time, the more information that was gathered the better. In future research, however, depending on the experience some of these steps like an X-ray scan can be omitted. This was done during the thesis as well when the frontside delayering in combination with the IR images and SEM cross-sections already

provided the necessary information and the backside delayering was therefore skipped. Taking the time to do the acquisition carefully and documenting it well helped build a foundation on which to fall back during the remaining processes.

This phase also included thinning the backside. This process turned out to be the most problematic as it created a short circuit between the Vcc and Gnd rail, thereby breaking all chip functionalities. Many attempts were made to solve this problem, like keeping the bulk at least $10\mu m$ thick, removing all the copper and trying out different milling patterns and mills. None seemed to make a difference. Possible causes now include a charge built up either during milling or cleaning thereby damaging potential ESD protection circuits or creating a latch-up. More research should be done in the particular Vcc and Gnd structures around the pads where the short seems to occur. Better measurement setups to see the location and creation of the short can be built involving better IR imaging or VCI. Measuring the built up of charge during the milling or cleaning process might prove to be difficult but grounding the sample and mill during processing might be a solution to these problems.

6.1.2. Invasive processing

There are several conclusions that can be drawn from the experiments regarding the creation of holes, pillars and the deposition of the insulator. First, the creation of holes is challenging, mainly due to the varying thickness of the bulk after milling. Additionally, it was found that a bulk of $5\mu m$ or less was easier to handle and smaller, more precise holes could be made. Higher aspect ratios caused inconsistencies in the etch rate. Furthermore, to create a deep hole within a manageable time, higher currents were needed, resulting in lower accuracy of the depth.

The deposition of the insulator proved to be challenging at first but showed promising results after only one iteration. As there are no manual adjustments involved, this process is rather stable and reliable. That being said, a proper way to test the quality of the deposition is yet to be developed. Apart from the cross-section from the SEM there is no other way to compare the different processes.

Finally, the creation of pillars underwent the most tests of the three processes. In the end, straight tall pillars could be deposited. Later it turned out that the deposition of both the pillars and TEOS only meant a more chaotic result after ALD or ICPECVD. A cleaner workflow resulted from depositing the pillars afterwards. There is much to be desired regarding measurements and testing of this process. Some methods for resistance measurements have been proposed in section 4.2.3 but were not put into practice. Adequate testing of the resistance does require professional testing equipment like probe needles inside a SEM.

In general, these steps are the most time-consuming and need a lot more testing where possible. Furthermore, attempts in the other areas of interest, like the insulated traces or the flash were not made since most of the chips had already broken down in earlier steps. That being said, the results are promising and show potential for the creation of holes in at least the $130nm$ technology node and possibly even smaller. That is under the condition that the bulk can be milled to at least $5\mu m$ since higher aspect ratio holes create too much variance.

6.1.3. Establish external connection

The connection phase was the fastest and most straightforward phase. Although the deposition of tungsten was unsuccessful, the experience gathered from the pillar deposition made the deposition of the pads and traces easy. Bonding the wires to the pads also went without problems.

Although the traces were deposited separately and were therefore connected similarly to what would be a pillar, a complete test with the pillar was never carried out. Furthermore, early resistance measurements show a high resistance emerging from either the trace or a contact. With low voltages and low drive strengths of the measured signals, this resistance should be kept as low as possible. More experiments and improvements should therefore be made to better understand the structure and the origin of this resistance.

6.2. Recommendations

As this was a first attempt at creating a connection, the options felt limitless. Each problem could be solved in multiple ways and testing them all was not possible within the timeframe of this thesis. This means that each process could be improved upon, in reliability, practicality and quality. To be able to create backside contacts in the latest mobile phone chips, smaller holes, lower resistance and more

precise methods are needed. The following list outlines the optimal follow-up steps and highlights the primary obstacles that must be addressed first.

- Research the cause of the short circuit phenomenon appearing after thinning the bulk.
- Complete a whole cycle, irrespective of short circuits or not to prove its functionality to a limited extent.
- Carry out experiments regarding the measurement of the three resistances mentioned in section 4.2.3.
- Attach a probe and find the limits of the contact on the signal voltage and frequency to see if the signals are detectable.
- Improve the milling process and try different processes like RIE. The resulting die should have no shorts and the bulk should be a maximum of $5\mu m$ thick.
- Decrease the hole size and observe the change in the quality of insulation and the resistance to see if the current methods are also applicable in smaller technology nodes.
- Try the two other areas of interest; the insulated interconnect line and a flash data line.

References

[1] *Doorbraak in onderscheppen van cryptocommunicatie*, Nov. 2018. [Online]. Available: https://www.om.nl/actueel/nieuws/2018/11/06/doorbraak-in-onderscheppen-van-cryptocommunicatie.

[2] S. Hamdioui, J.-L. Danger, G. D. Natale, F. Smailbegovic, G. v. Battum, and M. Tehranipoor, "Hacking and protecting IC hardware," in *Design, Automation & Test in Europe Conference & Exhibition (DATE), 2014*, New Jersey: IEEE Conference Publications, 2014, pp. 1–7, ISBN: 9783981537024. DOI: 10.7873/DATE.2014.112. [Online]. Available: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6800313.

[3] A. Oberai and J.-S. Yuan, "Efficient Fault Localization and Failure Analysis Techniques for Improving IC Yield," *Electronics*, vol. 7, no. 3, 2018, ISSN: 2079-9292. DOI: 10.3390/electronics7030028. [Online]. Available: https://www.mdpi.com/2079-9292/7/3/28.

[4] R. Naveenkumar, N. M. Sivamangai, A. Napolean, and V. Janani, "A Survey on Recent Detection Methods of the Hardware Trojans," in *2021 3rd International Conference on Signal Processing and Communication (ICPSC)*, 2021, pp. 139–143. DOI: 10.1109/ICSPC51351.2021.9451682.

[5] C. Shepherd, K. Markantonakis, N. van Heijningen, *et al.*, "Physical fault injection and side-channel attacks on mobile devices: A comprehensive analysis," *Computers & Security*, vol. 111, p. 102 471, 2021, ISSN: 0167-4048. DOI: https://doi.org/10.1016/j.cose.2021.102471. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167404821002959.

[6] D. Lu and C. Wong, Eds., *Materials for Advanced Packaging*, 2nd ed. Cham: Springer International Publishing, Nov. 2017, pp. 373–419, ISBN: 978-3-319-45097-1. DOI: 10.1007/978-3-319-45098-8.

[7] L. Van Gemert, T. Kamphuis, R. Van Der Meulen, and E. Israel, *HEATSINK VERY-THIN QUAD FLAT NO-LEADS (HVQFN) PACKAGE*, Jan. 2017.

[8] J. Liu, *Imec Introduces Intermetallics and Airgaps in Advanced Interconnect Metallization Schemes*, Jul. 2021. [Online]. Available: https://www.imec-int.com/en/press/imec-introduces-intermetallics-and-airgaps-advanced-interconnect-metallization-schemes.

[9] R. S. Feigelson, "Crystal growth History: Theory and melt growth processes," *Journal of Crystal Growth*, vol. 594, p. 126 800, 2022, ISSN: 0022-0248. DOI: https://doi.org/10.1016/j.jcrysgro.2022.126800. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022024822002883.

[10] J. Appels, E. Kooi, M. M. Paffen, J. J. H. Schatorje, and W. H. C. G. Verkuylen, "Local oxidation of silicon and its application in semiconductor-device technology," *PHILIPS RESEARCH Reports*, vol. 25, no. 2, pp. 118–132, Apr. 1970.

[11] S. Sadhukhan, S. Acharya, T. Panda, *et al.*, "Evolution of high efficiency passivated emitter and rear contact (PERC) solar cells," *Sustainable Developments by Artificial Intelligence and Machine Learning for Renewable Energies*, pp. 63–129, Jan. 2022. DOI: 10.1016/B978-0-323-91228-0.00007-0.

[12] A. Chatterjee, J. Esquivel, S. Nag, *et al.*, "A shallow trench isolation study for 0.25/0.18 /spl mu/m CMOS technologies and beyond," in *1996 Symposium on VLSI Technology. Digest of Technical Papers*, 1996, pp. 156–157. DOI: 10.1109/VLSIT.1996.507831.

[13] M. Lisker, S. Marschmeyer, M. Kaynak, and I. Tekin, "Sub-Atmospheric Chemical Vapor Deposition of SiO2 for Dielectric Layers in High Aspect Ratio TSVs," *ECS Transactions*, vol. 35, no. 2, p. 95, 2011, ISSN: 1938-5862. DOI: 10.1149/1.3568851. [Online]. Available: https://dx.doi.org/10.1149/1.3568851.

[14] E. Fortunato, P. Barquinha, and R. Martins, “Oxide Semiconductor Thin-Film Transistors: A Review of Recent Advances,” *Advanced Materials*, vol. 24, no. 22, pp. 2945–2986, Jun. 2012, ISSN: 0935-9648. DOI: https://doi.org/10.1002/adma.201103228. [Online]. Available: https://doi.org/10.1002/adma.201103228.

[15] H. Brune, “Growth Modes,” *Encyclopedia of Materials: Science and Technology, Sect. 1.9, Physical Properties of Thin Films and Artificial Multilayers*, pp. 3683–3693, 2001. [Online]. Available: https://infoscience.epfl.ch/record/135795?v=pdf.

[16] E. Lusky, Y. Shacham-Diamand, G. Mitenberg, A. Shappir, I. Bloom, and B. Eitan, “Investigation of channel hot electron injection by localized charge-trapping nonvolatile memory devices,” *IEEE Transactions on Electron Devices*, vol. 51, no. 3, pp. 444–451, 2004, ISSN: 1557-9646. DOI: 10.1109/TED.2003.823245.

[17] M. Lenzlinger and E. H. Snow, “Fowler□Nordheim Tunneling into Thermally Grown SiO2,” *Journal of Applied Physics*, vol. 40, no. 1, pp. 278–283, Jan. 1969, ISSN: 0021-8979. DOI: 10.1063/1.1657043. [Online]. Available: https://doi.org/10.1063/1.1657043.

[18] R. Bez, E. Camerlenghi, A. Modelli, and A. Visconti, “Introduction to flash memory,” *Proceedings of the IEEE*, vol. 91, no. 4, pp. 489–502, 2003, ISSN: 1558-2256. DOI: 10.1109/JPROC.2003.811702.

[19] W. F. Brinkman, D. E. Haggan, and W. W. Troutman, “A history of the invention of the transistor and where it will lead us,” *IEEE Journal of Solid-State Circuits*, vol. 32, no. 12, pp. 1858–1865, 1997, ISSN: 1558-173X. DOI: 10.1109/4.643644.

[20] G. E. Moore, “Cramming More Components Onto Integrated Circuits,” *Proceedings of the IEEE*, vol. 86, no. 1, pp. 82–85, 1998, ISSN: 1558-2256. DOI: 10.1109/JPROC.1998.658762.

[21] *Smaller, better, faster: imec presents chip scaling roadmap*, Feb. 2023. [Online]. Available: https://www.imec-int.com/en/articles/smaller-better-faster-imec-presents-chip-scaling-roadmap.

[22] Y. Liu, X. Duan, H.-J. Shin, S. Park, Y. Huang, and X. Duan, “Promises and prospects of two-dimensional transistors,” *Nature*, vol. 591, no. 7848, pp. 43–53, 2021, ISSN: 1476-4687. DOI: 10.1038/s41586-021-03339-z. [Online]. Available: https://doi.org/10.1038/s41586-021-03339-z.

[23] K. Y. Kamal, “The Silicon Age: Trends in Semiconductor Devices Industry,” *Journal of Engineering Science and Technology Review*, vol. 15, pp. 110–115, May 2022. DOI: 10.25103/jestr.151.14.

[24] M. Jurczak, N. Collaert, A. Veloso, T. Hoffmann, and S. Biesemans, “Review of FINFET technology,” in *2009 IEEE International SOI Conference*, IEEE, Oct. 2009, pp. 1–4, ISBN: 978-1-4244-4256-0. DOI: 10.1109/SOI.2009.5318794. [Online]. Available: http://ieeexplore.ieee.org/document/5318794/.

[25] H. Mertens, R. Ritzenthaler, A. Hikavyy, *et al.*, “Gate-all-around MOSFETs based on vertically stacked horizontal Si nanowires in a replacement metal gate process on bulk Si substrates,” in *2016 IEEE Symposium on VLSI Technology*, IEEE, Jun. 2016, pp. 1–2, ISBN: 978-1-5090-0638-0. DOI: 10.1109/VLSIT.2016.7573416. [Online]. Available: http://ieeexplore.ieee.org/document/7573416/.

[26] C. Tan and H. Zhang, “Two-dimensional transition metal dichalcogenide nanosheet-based composites,” *Chemical Society Reviews*, vol. 44, no. 9, pp. 2713–2731, 2015, ISSN: 0306-0012. DOI: 10.1039/C4CS00182F. [Online]. Available: http://dx.doi.org/10.1039/C4CS00182F.

[27] H. Jagannathan, B. Anderson, C.-W. Sohn, *et al.*, “Vertical-Transport Nanosheet Technology for CMOS Scaling beyond Lateral-Transport Devices,” in *2021 IEEE International Electron Devices Meeting (IEDM)*, IEEE, Dec. 2021, pp. 1–26, ISBN: 978-1-6654-2572-8. DOI: 10.1109/IEDM19574.2021.9720561. [Online]. Available: https://ieeexplore.ieee.org/document/9720561/.

[28] M. K. Gupta, P. Weckx, P. Schuddinck, *et al.*, “A Comprehensive Study of Nanosheet and Forksheet SRAM for Beyond N5 Node,” *IEEE Transactions on Electron Devices*, vol. 68, no. 8, pp. 3819–3825, 2021, ISSN: 1557-9646. DOI: 10.1109/TED.2021.3088392.

[29] H. Mertens, R. Ritzenthaler, Y. Oniki, *et al.*, "Forksheet FETs with Bottom Dielectric Isolation, Self-Aligned Gate Cut, and Isolation between Adjacent Source-Drain Structures," in *2022 International Electron Devices Meeting (IEDM)*, IEEE, Dec. 2022, pp. 1–23, ISBN: 978-1-6654-8959-1. DOI: 10.1109/IEDM45625.2022.10019497. [Online]. Available: https://ieeexplore.ieee.org/document/10019497/.

[30] J. Ryckaert and S. B. Samavedam, "The CMOS 2.0 revolution," *Nature Reviews Electrical Engineering*, vol. 1, no. 3, pp. 139–140, 2024, ISSN: 2948-1201. DOI: 10.1038/s44287-023-00016-3. [Online]. Available: https://doi.org/10.1038/s44287-023-00016-3.

[31] T.-H. Le, C. Canovas, and J. Clédière, "An overview of side channel analysis attacks," in *Proceedings of the 2008 ACM Symposium on Information, Computer and Communications Security*, ser. ASIACCS '08, New York, NY, USA: Association for Computing Machinery, 2008, pp. 33–43, ISBN: 9781595939791. DOI: 10.1145/1368310.1368319. [Online]. Available: https://doi.org/10.1145/1368310.1368319.

[32] E. Peeters, F.-X. Standaert, and J.-J. Quisquater, "Power and electromagnetic analysis: Improved model, consequences and comparisons," *Integration*, vol. 40, no. 1, pp. 52–60, 2007, ISSN: 0167-9260. DOI: https://doi.org/10.1016/j.vlsi.2005.12.013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167926005000647.

[33] A. Barenghi, L. Breveglieri, I. Koren, and D. Naccache, "Fault Injection Attacks on Cryptographic Devices: Theory, Practice, and Countermeasures," *Proceedings of the IEEE*, vol. 100, no. 11, pp. 3056–3076, 2012, ISSN: 1558-2256. DOI: 10.1109/JPROC.2012.2188769.

[34] B. Debraize and I. M. Corbella, "Fault Analysis of the Stream Cipher Snow 3G," in *2009 Workshop on Fault Diagnosis and Tolerance in Cryptography (FDTC)*, 2009, pp. 103–110. DOI: 10.1109/FDTC.2009.33.

[35] S. Ali, X. Guo, R. Karri, and D. Mukhopadhyay, "Fault Attacks on AES and Their Countermeasures," in *Secure System Design and Trustable Computing*, C.-H. Chang and M. Potkonjak, Eds., Cham: Springer International Publishing, 2016, pp. 163–208, ISBN: 978-3-319-14971-4. DOI: 10.1007/978-3-319-14971-4{_}5. [Online]. Available: https://doi.org/10.1007/978-3-319-14971-4_5.

[36] R. L. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and public-key cryptosystems," *Communications of the ACM*, vol. 21, no. 2, pp. 120–126, Feb. 1978, ISSN: 0001-0782. DOI: 10.1145/359340.359342.

[37] J. Fan and I. Verbauwhede, "An Updated Survey on Secure ECC Implementations: Attacks, Countermeasures and Cost," in *Cryptography and Security: From Theory to Applications: Essays Dedicated to Jean-Jacques Quisquater on the Occasion of His 65th Birthday*, D. Naccache, Ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 265–282, ISBN: 978-3-642-28368-0. DOI: 10.1007/978-3-642-28368-0{_}18. [Online]. Available: https://doi.org/10.1007/978-3-642-28368-0_18.

[38] X. Yuan, J. Anderson, L. Yang, J. K. Joshuva, and J. Land, "Survey of Recent Hacking Events," in *2019 SoutheastCon*, IEEE, Apr. 2019, pp. 1–6, ISBN: 978-1-7281-0137-8. DOI: 10.1109/SoutheastCon42311.2019.9020605. [Online]. Available: https://ieeexplore.ieee.org/document/9020605/.

[39] V. Korchnoy, "Investigation of Choline Hydroxide for Selective Silicon Etch from a Gate Oxide Failure Analysis Standpoint," Oct. 2002, pp. 325–331. DOI: 10.31399/asm.cp.istfa2002p0325.

[40] T. Onuki, R. Ono, A. Suzuki, H. Ojima, J. Shimizu, and L. B. Zhou, "A Thin Silicon Wafer Thickness Measurement System by Optical Reflectmetry Scheme Using Fourier Transform Near-Infrared Spectrometer," *Advanced Materials Research*, vol. 797, pp. 549–554, 2013, ISSN: 1662-8985. DOI: 10.4028/www.scientific.net/AMR.797.549. [Online]. Available: https://www.scientific.net/AMR.797.549.

[41] O. Breitenstein, J. Bauer, M. Kittler, T. Arguirov, and W. Seifert, "EBIC and luminescence studies of defects in solar cells," *Scanning*, vol. 30, no. 4, pp. 331–338, Jul. 2008, ISSN: 0161-0457. DOI: https://doi.org/10.1002/sca.20112. [Online]. Available: https://doi.org/10.1002/sca.20112.

[42] W. J. Parker, R. J. Jenkins, C. P. Butler, and G. L. Abbott, “Flash Method of Determining Thermal Diffusivity, Heat Capacity, and Thermal Conductivity,” *Journal of Applied Physics*, vol. 32, no. 9, pp. 1679–1684, Sep. 1961, ISSN: 0021-8979. DOI: 10.1063/1.1728417. [Online]. Available: https://doi.org/10.1063/1.1728417.

[43] V. Rangelov, M. Sarstedt, J. Somerville, T. Marschner, R. Jonckheere, and A. Poelaert, “Exploring capabilities of electrical linewidth measurement (ELM) techniques,” *Microelectronic Engineering*, vol. 57-58, pp. 673–681, Sep. 2001, ISSN: 0167-9317. DOI: 10.1016/S0167-9317(01)00455-5.

A

Extracting die using nitric acid

A.1. Preperation

A.1.1. Equipment

- Hot plate (set to $115°C$)
- 1x Borosilicate beaker $25ml$, (large enough to put sample flat on bottom)
- 1x borosilicate glass Petri dish, large enough to put under the beaker.
- 1x borosilicate glass petri dish, to put as a lid over the beaker.
- Spray bottle with acetone (Filled!)
- Plastic tweezers
- Laboratory timer
- Pipette glass $25ml$
- 1x Borosilicate beaker $100ml$, for cleaning the pipette
- 1x Borosilicate beaker $1000m$l, with $800ml$ water

A.1.2. Materials

- 100% fuming Nitric Acid (in a bucket in fume hood)
- Paper towels

A.1.3. Safety

- Purple nitrile gloves
- Lab coat
- Safety goggles
- Safety boots

A.2. Procedure

A.2.1. Start etching

1. Check hot plate temperature $115°C$
2. Check the acetone bottle, it must be filled to the rim. If not, fill up
3. Put the petri dish open side up on the hotplate.
4. Spread some paper towels on the table in the fume hood.
5. Put the beaker next to the bucket with the glass bottles fuming acid
6. Make sure there is a large empty heavy beaker inside the bucket
7. Place the pipette in the fume hood
8. Open the bottle, make sure to get the lid inside the cap also
9. Place the cap next to the bucket on a paper towel

10. Pick up the $25ml$ beaker with your left hand and the pipette in your right hand. Make sure you can close the pipette with your thumb.
11. Hold the beaker next to the opening of the bottle and with the pipette get some fuming acid out of the bottle in the beaker.
12. Get enough fuming acid by repeating the last step if necessary (depends on the level in the bottle)
13. Place the pipette in the heavy beaker in the bucket
14. Place the beaker at the back of the table in the fume hood, with the petri dish as the lid
15. Close the bottle with the cap!

A.2.2. Cleaning

1. Check if the bottle is closed with the cap
2. Put some acetone through the pipette, make sure there is a low level of acetone after this in the beaker in the bucket
3. Clean the outside of the pipette with a towel with acetone
4. Get the pipette out of the beaker and close it with a finger on the bottom side
5. Fill the pipette with acetone, until more than half the pipette is filled, or put the spray bottle on the end of the pipette, put in some acetone and pull the acetone up in the pipette
6. Close the top of the pipette with your thumb
7. Rotate the pipette to also clean the top
8. Empty the pipette in the beaker in the bucket
9. Clean the outside of the pipette with a towel with acetone
10. Blow some air through the pipette to clean the inside
11. Place the pipette in the drawer
12. Empty the beaker with acetone from the bucket in the waste jerrycan in the room

A.2.3. Place the chip in the fuming acid

1. Place the $25ml$ beaker with fuming acid in the petri dish on the hotplate, keep the lid on the beaker
2. Remove the lid
3. Put the chip in the breaker
4. Place the lid back
5. Wait for 75 minutes

A.2.4. Etch stop

1. Check your safety equipment, see A.1.3
2. Place a paper towel on the bench
3. Place the second beaker on the paper towel
4. Take the etch beaker from the hotplate and decant the acid in the second beaker. **Be careful** not to pour the die(s).
5. Cover the second beaker with the original lid (Petri dish) against accidental spraying acetone in this beaker,
6. At once put a lot of acetone from the spray bottle over the die(s) and other remains in the etch beaker. **Do this fast**, otherwise, an extreme exothermic reaction occurs, resulting in damage to the sample or even starting a fire.

A.2.5. Cleaning

1. Put the remaining nitric acid slowly (drop-wise) in the $800ml$ water.
2. The water can be put through the normal drain, while the tap is running.
3. Spray acetone on the out and inside of all glassware and clean with a paper towel drenched in acetone.
4. Leave the remaining acetone in the fume hood or put in the waste jerrycan in the room.

B

Summary processed samples

The table below shows the different samples used during this project and the processes they underwent. Some results are also mentioned but most tests were left out. It can be assumed that the samples worked until they tested negative since tests were carried out after each process.

Table B.1: All samples used in testing with short description

#Sample		Rename		Test after heatsink removal	Glue type and removal		Remaining bulk thickness	Further processes
Batch #1	1	-		OK	Hard	Attempt at removal with acetone		
	2	-		OK	Hard			
	3	-		OK	Hard			
	4	-		OK	Hard	Nitric acid etch for 4 seconds, damage to package and pads, glue still intact		
	5	-		OK	Hard			
	6	-		OK	Hard			
	7	-		OK	Hard			
	8	-		OK	Hard			
	9	-		OK	Hard	Attempt at removal with acetone		
	10	-		OK	Hard			
	11	-		OK	Hard			
	12	-		OK	Hard			
	13	Stub #3	1	OK	Soft	Removed with acetone	$\sim 4\mu m$	After polish too far, no bulk left
	14	Stub #3	2	OK	Soft	Removed with acetone	$\sim 10\mu m$	Resistance before was 3.55Ω, short
	15	Stub #1	2	OK	Soft	Polished with $30\mu m$ paste	$4.51 - 8.58\mu m$	Leveled on glue then polished resulting in damaged bulk. Afterwards thinned
	16	Stub #3	3	OK	Soft	Removed with acetone	$\sim 4\mu m$	Resistance before 130Ω. Verified OK, 2x hole to Vcc, ALD deposition
	17	Stub #3	4	OK	Soft	Removed with acetone	$5 - 15\mu m$	Resistance before 7Ω. Verified OK @ $0.25A$ short-circuit current. Used for cross-section Vcc
	18	Stub #3	5	OK	Soft	Removed with acetone	$5 - 20\mu m$	Resistance before 130Ω. Verified OK, 2x hole to Vcc with Pt and TEOS, ALD deposition

Continued on next page

Table B.1 – continued from previous page

#Sample		Rename		Test after heatsink removal	Glue type and removal		Remaining bulk thickness	Further processes
	19	Stub #3	6	OK	Soft	Removed with acetone	$\sim 5\mu m$	Resistance before 2.5Ω. Verified BAD, short
	20	Stub #1	1	OK	Soft	Removed with acetone	$17.7 - 21.4\mu m$	Applied carbon paint, unable to remove carbon paint
	21	Stub #1	3	OK	Soft	Removed with acetone	$6 - 13\mu m$	Applied carbon paint, unable to remove carbon paint
	22	-		OK	Soft	Removed with acetone	-	
	23	-		OK	Soft	Removed with acetone	-	Bad connection to pins
	24	Stub #2	1	OK	Soft	Removed with acetone	$\sim 80\mu m$	Verification OK, polish, Verification OK
	25	-		OK	Soft	Removed with acetone	-	
	26	Stub #1	6	OK	Soft	Polished with $15\mu m$ paste	-	Same polishing result as #15
	27	Stub #1	4	OK	Soft	Removed with acetone	$8.82 - 12.2\mu m$	Used for cross-section flash. Carbon paint applied. Used for testing IEE and TEOS
	28	Stub #1	5	OK	Soft	Removed with acetone	-	Milled too far, verification BAD, short
	29	Stub #2	2	OK	Soft	Removed with acetone	$10 - 24\mu m$	Verification BAD, short. Placed under IR heat camera
	30	Stub #3	7	OK	Soft	Removed with acetone	$\sim 15\mu m$	Bad connection to pins, resolved with scrating, verification OK. 2x holes to Vcc with Pt and TEOS, ICPECVD deposition
	31	Stub #2	3	OK	Soft	Removed with acetone	-	Milled too far, verification BAD, short
	32	Stub #2	4	OK	Soft	Removed with acetone	$5 - 10\mu m$	Verification BAD, short. Placed under IR heat camera
	33	Stub #1	7	OK	Soft	Removed with nitric acid	$3.87 - 11.5\mu m$	Sediment easy to remove but some damage to package and pads
	34	Stub #2	5	OK	Soft	Removed with acetone	-	Milled too far, verification BAD, sck error
	35	Stub #2	6	OK	Soft	Removed with acetone	$\sim 40\mu m$	Verification OK, polish, Verification OK. Create 2 holes to Vcc, ICPECVD deposition
	36	Stub #2	7	OK	Soft	Removed with acetone	-	
	37	-		-	-		-	Used for cross-section of entire chip
Isolation deposition	1	ALD	(#16)	-	-		-	Initial resistance 21Ω, after power connected 130Ω, verification OK
	2	ALD+	(#18)	-	-		-	Initial resistance 1Ω, after power connected 1Ω, verification BAD
	3	ICPECVD	(#35)	-	-		-	Initial resistance 11Ω, after power connected $1Omega$, verification BAD. Testing Pt adn W pads and traces, testing bond wires
	4	ICPECVD+	(#30)	-	-		-	Initial resistance 1Ω, after power connected 1Ω, verification BAD
Batch #2	40	Stub #1	1	OK	Soft	Removed with acetone	-	Too far, no bulk remaining
	41	Stub #1	2	OK	Soft	Removed with acetone	$\sim 8.44\mu m$	Initial resistance 0.68Ω, verified OK. 2 holes to Vcc, ICPECVD deposition

Continued on next page

Table B.1 – continued from previous page

#Sample		Rename		Test after heatsink removal	Glue type and removal		Remaining bulk thickness	Further processes
	42	Stub #1	3	OK	Soft	Removed with acetone	$\sim 4.79\mu m$	Initial resistance $2k\Omega$, verified BAD, short with resistance 2Ω
	43	Stub #1	4	OK	Soft	Removed with acetone	$\sim 5.05\mu m$	Initial resistance $1k\Omega$, verified BAD, short with resistance 2Ω
	44	Stub #2	1	OK	Soft	Removed with acetone	$\sim 5.19\mu m$	Initial resistance $0.67k\Omega$, verified BAD, short
	45	Stub #2	2	OK	Soft	Removed with acetone	$\sim 4.7\mu m$	Initial resistance 5.8Ω, verified BAD, short. 2 holes to Vcc, ALD deposition
	46	Stub #2	3	OK	Soft	Removed with acetone	$\sim 5.0\mu m$	Initial resistance 13.5Ω, verified BAD, short
	47	Stub #2	4	OK	Soft	Removed with acetone	$\sim 10.0\mu m$	Initial resistance 8.8Ω, verified BAD, short
	48	-		-	-		-	Pads in bad condition
	49	-		-	-		-	Pads in bad condition
	50	-		-	-		-	Pads in bad condition
	51	-		-	-		-	Pads in bad condition
	52	-		-	-		-	Pads in bad condition
	53	-		-	-		-	Pads in bad condition
	54	-		-	-		-	Pads in bad condition
	55	-		-	-		-	Pads in bad condition
	56	-		-	-		-	Pads in bad condition
	57	-		-	-		-	Pads in bad condition
	58	-		-	-		-	Pads in bad condition
	59	-		-	-		-	Pads in bad condition
Die extraction with nitric acid	1	-		-	-		-	For backside delayering
	2	-		-	-		-	For backside delayering
	3	-		-	-		-	For frontside delayering, delayered using Alpad
	4	-		-	-		-	For frontside delayering, delayered using RIE
	5	-		-	-		-	Reserve
	6	-		-	-		-	Reserve