THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

Préparée à l'École normale supérieure de Paris

Reverse Engineering Secure Systems Using Physical Attacks

École doctorale n°386
Sciences Mathématiques de Paris Centre

Spécialité Mathématiques

COMPOSITION DU JURY

Pr. Konstantinos MARKANTONAKIS,
Université de Londres, Royal Holloway,
Président du Jury

Dr. Sergei SKOROBOGATOV,
Université de Cambridge,
Examinateur

Dr. Markus KUHN,
Université de Cambridge,
Examinateur

Pr. Ingrid VERBAUWHEDE,
Université de Leuven, COSIC,
Examinateur

Général d'armée Marc WATIN-AUGOUARD,
Gendarmerie Nationale,
Examinateur

Pr. Keith MAYES,
Université de Londres, Royal Holloway,
Rapporteur

Dr. Damien SAUVERON,
XLIM, Université de Limoges,
Rapporteur

Commandant/Dr. Thomas SOUVIGNET,
Gendarmerie Nationale,
Expert Invité

Soutenue par Thibaut HECKMANN

le 18 Juin 2018, à l'École Militaire de Paris

Dirigée par Pr. David NACCACHE
École normale supérieure

ENS ÉCOLE NORMALE SUPÉRIEURE | PSL RESEARCH UNIVERSITY PARIS

Thèse de Doctorat

Rétro-Conception De Systèmes Sécurisés Par Attaques Physiques

Auteur: Thibaut HECKMANN

en vue de l'obtention du grade de

Docteur de l'École normale supérieure de Paris
spécialité Mathématiques

présentée et soutenue publiquement le 18 Juin 2018

devant le jury composé de:

Directeur de thèse:	David Naccache	*(École normale supérieure de Paris)*
Rapporteurs:	Keith Mayes	*(Royal Holloway, Université de Londres)*
	Damien Sauveron	*(XLIM, Université de Limoges)*
Examinateurs:	Konstantinos Markantonakis	*(Royal Holloway, Université de Londres)*
	Sergei Skorobogatov	*(Université de Cambridge)*
	Markus Kuhn	*(Université de Cambridge)*
	Ingrid Verbauwhede	*(COSIC, Université de Leuven)*
	Marc Watin-Augouard	*(Gendarmerie Nationale)*
Invité:	Thomas Souvignet	*(Gendarmerie Nationale)*

École doctorale 386: Sciences mathématiques de Paris Centre
Unité de recherche: UMR 8548 - Département d'Informatique de l'École normale supérieure

Doctorate Dissertation

Reverse Engineering Secure Systems Using Physical Attacks

Author: Thibaut HECKMANN

submitted in fulfillment of the requirements for the degree of

Doctor of the École normale supérieure of Paris
Mathematics Specialty

publicly defended and presented on June 18th, 2018

to the jury consisting of:

Supervisor:	David Naccache	*(École normale supérieure of Paris)*
Referees:	Keith Mayes	*(Royal Holloway, University of London)*
	Damien Sauveron	*(XLIM, University of Limoges)*
Examiners:	Konstantinos Markantonakis	*(Royal Holloway, University of London)*
	Sergei Skorobogatov	*(University of Cambridge)*
	Markus Kuhn	*(University of Cambridge)*
	Ingrid Verbauwhede	*(COSIC, University of Leuven)*
	Marc Watin-Augouard	*(National Gendarmerie)*
Guest member:	Thomas Souvignet	*(National Gendarmerie)*

Doctoral School 386: Mathematical Sciences – Paris Centre
Research unit: UMR 8548 - The École normale supérieure's Computer Science Department

Résumé

Avec l'arrivée des dernières générations de téléphones chiffrés (BlackBerry PGP, iPhone), l'extraction des données par les experts est une tâche de plus en plus complexe et devient un véritable défi notamment après une catastrophe aérienne ou une attaque terroriste. Dans cette thèse, nous avons développé des attaques physiques sur systèmes cryptographiques à des fins d'expertises judiciaires.

Une nouvelle technique de re-brasage à basse température des composants électroniques endommagés, utilisant un mélange eutectique 42Sn/58Bi, a été développée. Nous avons exploité les propriétés physico-chimiques de colles polymères et les avons utilisées dans l'extraction de données chiffrées. Une nouvelle technique a été développée pour faciliter l'injection et la modification à haute-fréquence des données. Le prototype permet des analyses en temps réel des échanges processeur-mémoire en attaque par le milieu. Ces deux techniques sont maintenant utilisées dans des dispositifs d'attaques plus complexes de systèmes cryptographiques.

Nos travaux nous ont mené à sensibiliser les colles polymères aux attaques laser par pigmentation. Ce processus permet des réparations complexes avec une précision laser de l'ordre de 15 micromètres. Cette technique est utilisable en réparations judiciaires avancées des crypto-processeurs et des mémoires.

Ainsi, les techniques développées, mises bout à bout et couplées avec des dispositifs physiques (tomographie 3D aux rayons X, MEB, laser, acide fumant) ont permis de réussir des transplantations judiciaires de systèmes chiffrés en conditions dégradées et appliquées pour la première fois avec succès sur les téléphones BlackBerry chiffrés à l'aide de PGP.

Mots Clés: Retro-Conception Matérielle, Extractions Physiques, Attaques par le Milieu, Tomographie 3D aux Rayons X, Interactions Laser-Matière et Electron-Matière, Attaques Chimiques, Transplantations Judiciaires.

Abstract

When considering the latest generation of encrypted mobile devices (BlackBerry's PGP, Apple's iPhone), data extraction by experts is an increasingly complex task. Forensic analyses even become a real challenge following an air crash or a terrorist attack. In this thesis, we have developed physical attacks on encrypted systems for the purpose of forensic analysis.

A new low-temperature re-soldering technique of damaged electronic components, using a 42Sn/58Bi eutectic mixture, has been developed. Then we have exploited the physico-chemical properties of polymer adhesives and have used them for the extraction of encrypted data. A new technique has been developed to facilitate injection and high-frequency data modification. By a man-in-the-middle attack, the prototype allows analysing, in real-time, the data exchanges between the processor and the memory. Both techniques are now used in more complex attacks of cryptographic systems.

Our research has led us to successfully sensitise polymer adhesives to laser attacks by pigmentation. This process allowed complex repairs with a laser with 15 micrometres precision and has been used in advanced forensic repair of crypto-processors and memory chips.

Finally, the techniques developed in this thesis, put end-to-end and coupled with physical devices (X-ray 3D tomography, laser, SEM, fuming acids), have made it possible to have successful forensic transplants of encrypted systems in degraded conditions. We have successfully applied them, for the first time, on a PGP-encrypted BlackBerry mobile phone.

Keywords: Hardware Reverse Engineering, Physical Extractions, Man-In-The-Middle Attacks, X-ray 3D Tomography, Laser-Matter and Electron-Matter Interactions, Chemical Attacks, Forensic Transplantations.

Acknowledgements

This thesis is the outcome of three years of work in the Computer Science Department's Information Security Group of the École normale supérieure (ENS) of Paris. This work would not have been possible without the support of many people that I would like to thank here.

First of all, I would like to express my gratitude to Prof. David Naccache, my thesis advisor, for his unfailing support. He had believed very early in the potential of our research. An altruistic scientist, a member of the Institut Universitaire of France, ENS's ISG director and forensic expert for the International Criminal Court of The Hague: I wish to express the honour and the unbounded privilege which have been mine to be able to conduct my research by his side.

I wish to thank Major Thomas Souvignet for introducing me to this thesis work, for his involvement in our publications and the attentive technical proofreading of this manuscript. I owe him a lot for all he has done for me.

I express my deepest respect to all my Gendarmerie commanders for having believed in me. I hugely thank General Daoust who managed to convince the Gendarmerie's highest echelons before the CEST commission. I thank my direct chain of command, General Touron, Director of the Forensic Sciences Institute of the French Gendarmerie (IRCGN), as well as my Divisional Director, Colonel Strebel, Major Rubens and Major Letrillard. I will never forget their contributions as I now understand that officers' decisions are required to support the projects we believe in.

I would like to widely thank my proofreader, Dame Natacha Laniado, English Lecturer at Panthéon Assas University and PhD pastoral care at ENS Paris, who has helped me to proofread and correct this thesis and our papers.

I would also like to thank the entire Royal Holloway, University of London (RHUL) Computer Science Department, the Information Security Group team and the Smart Card and IoT Security Centre Laboratory, that helped me throughout my thesis, and especially Prof. Keith Mayes, Head of School of Mathematics and Information Scurity Group (ISG), Prof. Konstantinos Markantonakis, Director of the ISG Smart Card and IoT Security Centre and Dr Akram Raja Naeem for their advice, thesis proofreading and precious help. I thank you for your good humour and high level of competence that you shared. I thank all the RHUL's Chemistry Department and School of Biological Sciences, especially Prof. James McEvoy for his help during chemical attacks on electronic components.

I would warmly like to thank the University of Cambridge Computer Laboratory, especially Dr Sergei Skorobogatov and Dr Markus Kuhn from the Tamper And Monitoring Protection Engineering Research (TAMPER), for all close technical exchanges, information sharing, and to have accepted to be members of my thesis jury.

I would like to thank Dr Damien Sauveron from XLIM, University of Limoges, for the careful technical review of this manuscript.

I thank Graham Houghton, volunteer at Strode's College, who has helped me to proofread and correct this thesis. He very generously offered me help that proved to be valuable for finalising the thesis.

I am grateful to everyone who has worked with me on various topics, and especially my team at the IRCGN, officers and NCOs. The diversity of their profiles and competences undeniably constitute the strength of this fantastic unit.

I would like to express my recognition to the referees and examiners for their interest in my work, for agreeing to be part of my thesis jury and for providing me with their support and advice.

On a more personal note, I wish to thank my close family who believed in me, especially my wife, my parents and my brother. Without them none of this would have been possible —my wife, my pride; my parents, eternal support; and my brother, my role model. A special thank you to those who are no longer here, especially my grandparents, who have made me who I am today.

London, UK, March 12th, 2018,
Thibaut HECKMANN

"Above all, do not tell me it is impossible." General Philippe LECLERC

To my wife, to my parents, to my brother, to my family.

Contents

List of Figures

List of Tables

1

Introduction

Nowadays, many civilians are living with a continuous threat of terrorism or mass accident, such as the Nice attack [BBC-News, 2016], the Paris attack [BBC-News, 2015b], the Germanwings crash [BBC-News, 2017], and the Puisseguin accident [BBC-News, 2015a]. Forensic investigators must frequently bypass the protection mechanisms of embedded systems to extract data, which constitutes evidence for the criminal court. Terrorism or mass accident, as we see, are current topics in our society, and the bypassing of security mechanisms must also be made possible so both legal questions (proof in court, understanding the disaster) and ethical challenges (mourning of the victims' families) can be answered.

During our research, we have worked on the interface of theoretical physics and computer security, and have brought some new physics approaches to forensic investigations. The main objective of this thesis has been the implementation of innovative methods of extraction, and the physical bypass of cryptographic mechanisms to assist investigators and experts in their investigations.

On December 2nd 2015, a terrorist attack was carried out on behalf of the Islamic State in San Bernardino, California, United States. Two heavily armed terrorists burst into a centre for unemployed people, and began shooting, causing 14 deaths and 22 wounded [Time, 2015]. The US authorities finally managed to kill the two suspects. As part of a judicial investigation operation conducted by the US authorities, an Apple iPhone 5C mobile phone, running iOS 9, belonging to one of the terrorists was found. The encrypted phone was unfortunately protected by an unknown lock code with a mechanism of physical erasure of the data after 10 attempts. Knowing that a four-digit passcode has 10,000 possibilities, in February 2016, the US Federal Bureau of Investigation (FBI) announced that it was unable to unlock the phone. Apple declared not to have the passcode, and the FBI asked Apple to code up a new version of iOS 9 without the 10-guess limit. The FBI actually wanted to brute-force the passcode, but Apple refused to do that. Therefore, FBI filed a complaint against Apple to force them to create a tool allowing the unlocking of the phone. However, the case did not go to court, because the FBI managed to find a solution, with the help of a third party, to unlock the phone. The

security point of phones (like Apple, BlackBerry or Samsung) is a selling point. What can be a stronger advertising message than saying that Apple itself and the police are unable to access the user's data?

This first example shows the authorities' new interest in collaborating with external laboratories or universities. Before the San Bernardino attack, for confidentiality reasons, investigators were often closed to communicating on their means and their advances. But the high level of cryptography and the rapid evolution of technology has forced investigators to open up to new collaborations, without which the judicial investigations would end up without a quick solution. It is in this context that this thesis took place, with a strong collaboration between the Forensic Sciences Institute of the French Gendarmerie (IRCGN) and the academic world: École Normale Supérieure (ENS) of Paris; the Smart Card and IoT Security Centre (part of the Information Security Group) of Royal Holloway, University of London; the School of Biological Sciences of Royal Holloway; and the University of Cambridge Computer Science Department.

At a press conference [Keizer, 2016], on March 24th 2016, the FBI Director, James Commey, told reporters that the technique call "NAND mirroring" would never work. However, just after this announcement, Dr Sergei Skorobogatov, from the Computer Science Department of the University of Cambridge, proved the opposite and showed it in his paper: *"The bumpy road towards iPhone 5c NAND mirroring"* [Skorobogatov, 2016]. It is from this moment that forensic investigators were convinced that universities could be allies in the search for judicial evidence.

In this context, in which manufacturers do not want to exchange information with state authorities, forensic investigators must resort to new methods. However, before going into the technical details of this thesis, the current context firstly leads us to question three points:

- What does the French law say about the use of cryptography (section 1.1.1)?
- What are the legal penalties for crimes commission using cryptography (section 1.1.2)?
- What is the investigator's legitimacy to carry out the reverse engineering of encrypted systems (cryptanalysis) before extracting and analysing the data (section 1.1.3)?

It is in this sense that we will present in section 1.1 the laws that protect forensic investigators. Indeed, without these laws, investigators could be prosecuted for hacking activity within the meaning of *Article 323-1 of the French Criminal Code*[1].

[1] Article 323-1 of the French Criminal Code, Legislative Section, Book 3, Title 2, Chapter 3.

A question beyond the scope of this thesis is the possible evolution of the French legal framework to force manufacturers to reveal a back-door or create software in the context of terrorist activities and criminal offences. For the time being, parliamentarians and lawyers see it as an infringement of individual freedom. They also see it as an unworkable legislation because it would aim to ban some mobile phone brands in French territory. This question was, and is still, valid when writing this thesis, and at the time of the production of the facial recognition decipher system (face ID) of the Apple iPhone X. Thus, in September 2017, a French deputy[2] proposed an amendment which prohibited the commercial sale and distribution in France of the iPhone X, in case of a terrorism threat. The question is far from being decided, and the investigators have to face the new technologies and must be inventive.

1.1 Forensic Investigators in the Face of Cryptography: French Law

Cryptography (protocols, symmetric or asymmetric ciphers) has been widely democratised in recent years, but it is regulated both for individuals and for experts mandated to extract data protected by cryptographic algorithms. The regulation on cryptography aims at preserving the internal or external security of the state. There is a constant tension between the defence of the state's interests (confidentiality) and the preservation of individual rights (individual privacy).

In addition, France must also face a context where European Union legislation aims to harmonise the legislation of its Member States. The legal framework for forensic experts is a pillar that the justice system seems to want to strengthen. However, the expansion of cryptographic technologies is faster than the legislative framework [Freyssinet, 2012].

1.1.1 The Birth of Cryptography in French Law

French law defined cryptology for the first time in *Act no. 90-1170 of December 29th 1990, on the regulation of telecommunications*[3]. The law provides that cryptological services are all services intended to transform information, or signals, into unintelligible information for third parties. The law also mentions the reverse operation, using means, hardware or software designed for this purpose.

Cryptography can be used in different sectors and may create safety problems, particularly at the commercial or state security level. Data security becomes a marketing

[2] http://www.lci.fr/politique/eric-ciotti-veut-encore-faire-interdire-les-produits-apple-en-france-2064289.html
[3] Official Journal of the French Republic n. 303, December 30th 1990, Page 16439, Paragraph 28.

argument for encrypted phones (Apple iPhone, BlackBerry PGP), encrypted messaging applications (Telegram messenger, WhatsApp messenger) and on-line purchase (Apple pay, Android pay). Criminals can divert the ethical and lawful use of cryptography to prepare, disseminate and communicate offences. In the area of forensics, we frequently find the illegal uses of cryptography in international narcotics trafficking, the preparation and commissioning of terrorist attacks, in industrial spying and assassination files. The possibilities of using cryptography for criminal purposes are numerous, and the list keeps on growing.

Cryptography is subsequently clearly defined in *Section 29 of the Act of June 21st 2004*[4], and in that sense, cryptological means any hardware or software designed or modified to transform data. It can be either information or signals that use secret conventions, or to perform the inverse operation with secret conventions. The main purpose of these cryptographic means (Figure 1.1) is to guarantee the security of the storage or the transmission of data, making it possible to ensure its confidentiality (information accessible only to authorised persons), its authentication (to be sure of author identity) or the control of its integrity (no message alteration).

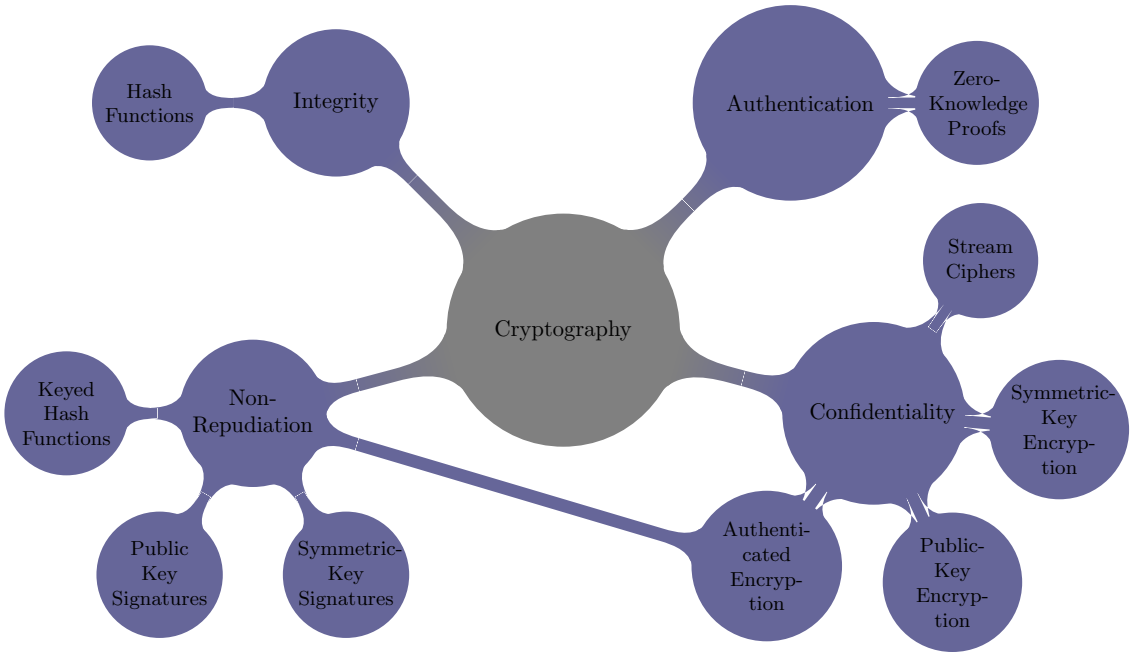


Figure 1.1: The Broad Fields of Cryptography [Menezes et al., 1997]

Cryptography within the legal framework aims at ensuring examination and authentication functions. In this sense, French laws have introduced some protocols for entering and for outputting cryptographic devices on national territory.

[4] Act n. 2004-575, June 21st 2004 for Confidence in the Digital Economy, LCEN.

If the software imported from a Member State of the European Union does not perform only control and authentication functions, an authorisation from the French Prime Minister will also be needed so the cryptographical means can be understood: source code, software and hardware. The software imported from other States will also need prior authorisation from the French Prime Minister. These applications are then sent to the French National Cybersecurity Agency[5] (ANSSI).

Decree No. 2007-663 of May 2nd 2007[6] and the *Order of January 29th 2015*[7] define the form and the content of the declaration files relating to the means of cryptology. **It should be noted that some categories of cryptological means remain free to import or export**. These categories concern, for example, the supply of equipment intended for the general public, such as radio or television reception, **mobile phones**, and on which most of the cryptographic capacity is not accessible to the user.

Through these French laws encrypted phones can freely and legally circulate on national territory. Both the forensic investigators and the users do not have access to the encryption mechanism operation. It will be seen in section 1.1.3 how French law allows forensic investigators to carry out the reverse engineering of encrypted systems (cryptanalysis) to access the phone's content.

1.1.2 The Criminal Sanctions Against the Illegal Use of Cryptography

When cryptology is used for illegal purposes, the law provides for a general mechanism for increasing sentences. Thus, *Article 132-79 of the French Criminal Code*[8] provides for the systematic increase of the penalty when criminals are using cryptology.

When a means of cryptology, within the meaning of *Article 29 of Act no. 2004-575 of June 21st 2004*, has been used for the preparation or commission of a crime or offence, the maximum penalty of deprivation of liberty shall increase as follows:

- It is liable to imprisonment for life when the offence is punishable by thirty years' imprisonment.
- It shall be increased to thirty years' imprisonment when the offence is punishable by twenty years' imprisonment.
- It shall be increased to twenty years' imprisonment when the offence is punishable by fifteen years' imprisonment.
- It shall be increased to fifteen years' imprisonment when the offence is punishable by ten years' imprisonment.

[5] http://www.ssi.gouv.fr/en/
[6] Decree n. 2007-663 adopted for the application of Articles 30, 31 and 36 of the LCEN.
[7] Order of January 29th 2015: https://www.legifrance.gouv.fr
[8] Legislative Section, Book 1, Title 3, Chapter 2, Section 3, Article 132-79.

- It is increased to ten years of imprisonment when the offence is punished by seven years of imprisonment.
- It shall be increased to seven years' imprisonment when the offense is punishable by five years' imprisonment.
- It shall be doubled when the offense Is punishable by a maximum of three years' imprisonment.

However, the provisions of this article do not apply to the perpetrator or accomplice of the offence if he has given a clear version of the encrypted messages and secret agreements necessary for decryption, as requested by the judicial authorities. This mechanism of aggravation is means of negotiation for investigators to increase the pressure on criminals. Unfortunately, if the investigator has no other means of proving the crime than the encrypted data, the article cannot be validated before the French courts. If no evidence is brought in before the courts, then there are no means of criminal punishment.

Article 434-15-2 of the Criminal Code[9] punishes anyone who has knowledge of the secret deciphering convention of a cryptology device, which may have been used to prepare, facilitate or commit a crime or offence, and refuses to hand over this convention to the judicial authorities, despite the authorities' requisitions. Such an act is punishable, by French Criminal Code, by three years of imprisonment and a fine of 45,000 euros.

In addition, as an amendment has been dropped by *Article 16 of the Act No. 2016-731 of June 3rd 2016*[10] which increases even more the level of penalties.

The law, therefore, provides for sanctions against perpetrators, accomplices of a crime, and against any other persons who are aware of an agreement of data deciphering used to commit or attempt to commit a crime.

1.1.3 Extracting Encrypted Data: the Experts' Work Legitimacy

The French Code of Criminal Procedure provides for two scenarios, depending on the nature of the infringement:

- The first one is when the infringement is a minor offence (consumption of narcotic drugs, theft, fraud, etc.).
- The second one is when the infringement is a criminal offence (terrorism, drug trafficking, assassination, child pornography, etc.).

The appointment of a judge, called "the opening of a judicial investigation", is decided by the public prosecutor. A judge is mandatory when the facts are criminal. Thus, at

[9] Criminal Code, Legislative Section, Book 1, Title 3, Chapter 1, Section 9, Article 434-15-2.

[10] Act n. 2016-731 of June 3rd 2016 "strengthening the fight against organized crime, terrorism and their financing and improving the efficiency and guarantees of criminal proceedings".

the opening of a judicial investigation, the judge is responsible for the investigations. Before the opening, the public prosecutor is responsible for the investigations. The Code of Criminal Procedure says that when it appears that data, seized or obtained during the investigation, has been encrypted, the authorities in charge of the case (public prosecutor or judge) may appoint an expert for a criminal offence or a qualified person for a minor offence to decipher the information: *Article 230-1 of the Code of Criminal Procedure, amended by Article 15 of the Act no. 2014-1353 of November 13th 2014.*

The law favours investigators when the data is protected by an authentication mechanism that prevents the access and understanding of the encrypted information contained in it. Thus, when the data entered was obtained during the investigation and has been transformed, the public prosecutor or the investigating judge may order the expert to carry out the technical operations of reverse engineering to be able to obtain access to the information. The objective is to obtain a clear version of the data, and in the case where means of cryptography has been used, to obtain the secret decryption agreement.

If the sentence incurred is equal to, or greater than, two years' imprisonment and if the investigations require specific cryptanalysis operations, the public prosecutor (or the judge) may prescribe recourse to the State's means subject to the secrecy of national defence[11]. The magistrate, who has control of the case, has the power to help the expert by allowing him to use military means to decipher the data (calculators). The aim is to give investigating magistrates the means to access the clear content of encrypted messages. Thus, judges may use the services of experts in cryptanalysis, as well as the means of the State covered by the secrecy of national defence.

Reverse engineering is therefore legal if the expert (or qualified person) follows the rules set by the Code of Criminal Procedure, Article 230-1: encrypted data for criminal offences (or minor offences) and a public prosecutor or judge's authorisation. The work of the expert is, therefore, authorised and regulated by French law.

1.2 Thesis Outline

With the increasing evolution of technologies and the generalisation of the use of cryptography, forensic investigators need to resort to new methods of data extraction (cryptanalysis). This forensic thesis deals with several topics related to bypassing the protection mechanisms and cryptographic implementations of embedded systems to help forensic investigators to extract and analyse data. The main contributions of this thesis are:

[11] Code of Criminal Procedure, Legislative Section, Book 1, Title 4, Chapter 1, Article 230-1.

- A new low-temperature soldering method minimising silicon thermal shock on electronic chips. The method reduces the risk of data modification/destruction during forensic desoldering operations.
- A new modus operandi of advanced repair of damaged electronic components after air crash, fire, submersion or terrorist attack by using various physico-chemical properties of adhesives.
- Development of a new method called "Chip Adhesives Method (CAM)" which is a Ball-Grid-Array Man-In-the-Middle (BGA-MIM) platform. The CAM allows investigators to provide advanced access to all exchanges in real time between CPU, memory and crypto-components. Thus, they can understand and modify the implemented security mechanisms inside secure mobile devices.
- Development of a new modus operandi that makes polymeric adhesives sensitive to laser decapsulation attacks while decreasing the laser power deposited during ablation.
- A new laser bond repair method with an accuracy of 15 micrometres.
- A new electrically conductive adhesive chemical composition using pigmentation properties.
- A new procedure called "PoP chip-off/TCA Technique" which allows forensic investigators to unsolder Package On Package (PoP) components without either damaging or destroying them.
- The full forensic transplantation of the BlackBerry PGP (Pretty Good Privacy).

Organisation. Chapter 2 introduces integrated circuit and logic design, from a transistor to hardware cryptochip. General semiconductors theory and fundamental operations are detailed in the first section. The second section shows the different types of memories on silicon chips and the constitution of their fundamental logic cells. The last section presents the manufacturing process of silicon memories from a piece of silicon to packaging memories.

Chapter 3 introduces the techniques used by forensic investigators to recover data on smartphones. The first section describes techniques used for undamaged mobile devices (manual extraction and logical extraction). The second section is devoted to the physical extraction and forensic hardware reverse engineering techniques applied to flash memory devices: chip-off, chip-on, attack by fuming nitric acid, ablation with laser cutter and nano-probing. The third section presents in detail techniques used for damaged and unsecured mobile devices. The last section shows an advanced technique developed during this thesis, called "forensic transplantation", used for damaged and secured mobile devices.

Chapter 4 details the fundamental study of the 42Sn/58Bi alloy and its forensic applications to the low-temperature re-balling and re-soldering of BGA components. The first section presents the limits of traditional unsoldering techniques. The second section consists of a theoretical study which presents the 42Sn/58Bi alloy fundamental properties and introduces two reverse engineering techniques: X-ray tomography and Scanning Electron Microscope. The last section is the practical part showing some PhD investigations' experimental results applied to 42Sn/58Bi alloy.

Chapter 5 describes polymer adhesive methods used in this thesis for advanced forensic applications. The first section presents the limits of traditional forensic methods through concrete cases. The second section introduces the material used and fundamental properties. The third section is devoted to developing new methods for solving the first section's limits. In the last section, we use the fundamental properties of adhesives to develop a memory man-in-the-middle platform. The platform is used for tracking signals between memory and CPU and to understand or modify the implemented security mechanisms inside secure mobile devices.

Chapter 6 presents a fundamental use of the laser cutter in the micrometric repair of an electronic component. Through concrete cases, the first section presents the limits of traditional forensic methods. The second section consists of a theoretical study of pigments and dyes. The third section presents some experimental results and the two industrial patents filed. The last section presents a direct forensic application case: the repair of a broken micrometric bonding wire.

Chapter 7 presents a method developed to transplant Package on Package (PoP) components and fundamental forensic reverse engineering methods. The first section presents traditional techniques and limits used by forensic experts. The second section describes the method developed to unsolder PoP components without destruction. The third section introduces, for the first time, the process developed during this thesis for the full transplantation of a damaged secured mobile phone using a hardware key on the processor of a BlackBerry 9900 PGP.

Chapter 8 concludes this thesis and sets out future work and advancements.

1.3 Publications

This thesis is primarily based on the publications described in this section. Among the developed work are 5 published papers, 2 patents and several talks in France, Canada, Belgium and the United Kingdom.

Low-Temperature Low-Cost 58 Bismuth—42 Tin Alloy Forensic Chip Re-Balling and Re-Soldering

with David Naccache, Thomas Souvignet and Sebastien Lepeer

Abstract. The re-soldering of electronic components is often necessary during forensic investigations. Such re-soldering usually occurs in two scenarios. In the first *in vivo* scenario, a component is extracted from the exhibit board and analysed (or unlocked) externally before being re-implanted in the exhibit board. In the second *in vitro* scenario, the extracted component is implanted in an external test board or an unlocked device of the same brand and model. We call such manipulations chip-off/chip-on procedures.

In some cases, data manipulation, performed during chip-off and chip-on, may also help to recover forensically significant data such as emails, text messages, contacts, photos or videos. Chip-off/chip-on techniques involve two risky steps during which forensic data can be irreversibly lost, given that a chip must be de-soldered and re-soldered again. During both operations, temperatures beyond the normal operating range are applied to a chip containing valuable information. In addition, for cost reasons, chips are usually not designed to withstand repeated soldering during their life-cycle.

A recent publication by the Netherlands Forensic Institute [Jongh, 2014] proposes to minimise thermal stress by using low-temperature alloys during the re-soldering phase. The necessary low soldering temperature 42Sn/58Bi 300 µm balls has a burdensome cost (several thousands of euros per ball jar) if a mass production is not considered. Given that the forensic analysis of mobile telephones has become a standard requirement in most criminal cases, unitary forensic analysis cost reduction is currently a necessity. Luckily, it is possible to find cheap soldering pastes composed of inhomogeneous balls of 42Sn/58Bi (25—45 µm) mixed in a solvent. How to use such pastes for forensic re-soldering proves to be a nontrivial laboratory exercise, on which we focus in this paper.

This work introduces a method called reballing, that will produce 300-micron beads from 25- to 45-micron balls. The proposed process is based on the use of a reballing stencil. We analyse the influence of the temperature descent curve during soldering and explore its effects on the final soldering quality. Finally, we will verify the compliance of our low-cost (less than 40 euros), low temperature (138°C), curve-optimised reworking process on micro-BGA components.

Note. This work is presented in detail in Chapter 4.

Electrically Conductive Adhesives, Thermally Conductive Adhesives and UV Adhesives in Data Extraction Forensics

with David Naccache and Thomas Souvignet

Abstract. Recent publications underline the interest of using polymers in microelectronics [Cui et al., 2014]. Polymers are the ideal interconnect alternative to solder materials containing lead. ECAs [Li and Wong, 2006], TCAs [Felba, 2011] and UVAs [Asif et al., 2005] mainly consist of a polymeric resin (epoxy, silicon, polyurethane or polyimide) that provides physical and mechanical properties such as adhesion and mechanical strength, while containing metal fillers (silver, gold, nickel or copper) that conduct electricity [Luo et al., 2016]. Currently, it is possible to find really cheap polymeric resin. Using these resins for digital forensic purposes is the focus of this paper, that we demonstrate in a hardware reverse engineering prototype case study.

When considering new mobile devices, such as secure phones, it is often necessary to spy on communications and perform numerous tests on the memory (e.g. by changing some bytes) to understand or modify the implemented security mechanisms (manipulate system time, locate password hashes, observe artefacts of implemented security algorithms, etc.). Traditional techniques use either laser attacks/probing (chip-on) or de-soldering/read/re-soldering (chip-off/on) [Heckmann et al., 2016] [Jongh, 2014]. These two techniques are unsuitable for repeated operations requiring many readings/changes/injections. This paper describes a concrete case study using adhesive properties complementary to chip-on and chip-off methods.

We present the steps using different properties of adhesives (ECA, TCA, UVA) that will lead to the realisation of a prototype particularly suitable for the repeating of the read phases/changes/injections necessary for reverse engineering secure mobile devices.

Note. This work is presented in detail in Chapter 5.

Decrease of energy deposited during laser decapsulation attacks by dyeing and pigmenting the ECA: application to the forensic micro-repair of wire bonding

with David Naccache and Thomas Souvignet

Abstract. Polymeric adhesives are of interest in the digital forensics domain. They can be used to perform more or less complex repairs or even to realise advanced man-in-the-middle attacks in order to carry out reverse engineering of secure systems.

The main aim of this paper is to develop a technique that makes polymeric adhesives sensitive to laser decapsulation attacks (decapping) while decreasing the laser power deposited during ablation. We will first introduce a theoretical part on the properties of laser radiation ablation and discuss the fundamental equations characterising laser–matter interactions.

In the practical part, we vary the absorbance of our target materials (ECA) by adding

either dyes or pigments at different concentrations, to evaluate the influence on sensitivity to laser decapping attacks. The addition of dyes or pigments will have an immediate and crucial impact on coefficients of the fundamental equation of heat that make the polymeric glues sensitive to laser decapping attacks.

Finally, to demonstrate the value of this work, a direct example of application is implemented for the micro-repair of broken bonding wires in areas where traditional techniques using the wire-bonder are not applicable or are likely to create additional damage to neighbouring bonding wires. This paper shows how to make conductive bonding, using ECA, with an accuracy of 15 micrometres.

Note. This work has served as the basis of 2 patents taken out jointly by the Ecole Normale Supérieure of Paris and the Forensic Sciences Institute of the French Gendarmerie (IRCGN). This work is presented in detail in Chapters 3 and 6.

Forensic Smartphone Analysis Using the Adhesives Transplantation of Package on Package Components

with David Naccache, Thomas Souvignet and Konstantinos Markantonakis

Abstract. Investigators routinely recover data from mobile devices. In many cases the target device is severely damaged. Events such as airplane crashes, accidents, terrorism or long submersion may bend or crack the device's main board and hence prevent using standard forensic tools. This paper shows how to salvage forensic information when the NAND memory, SoC or cryptographic chips are still intact. We do not make any assumptions about the state of the other components. In the usual forensic method, damaged phone components are analysed using a process called "forensic transplantation". This procedure consists of unsoldering (or lapping) chips, re-soldering them on a functionnal donor board and rebooting the phone.

Package on Package (PoP) component packaging is a new technique that allows stacking of two different silicon chips, e.g. memory, CPU or cryptographic processors. Currently, PoP is widely used by most device manufacturers, particularly by leading brands such as Apple, BlackBerry, Samsung, HTC and Huawei. Unfortunately, forensic transplantation destroys PoP components.

This work overcomes this difficulty by introducing a new chip-off analysis method based on High Temperature Thixotropic Thermal Conductive Adhesives (HTTTCA). The HTTTCA process allows the investigator to safely unsolder PoP components, which is a crucial step for transplantation. To demonstrate feasibility, we describe in detail an experimental forensic transplantation of a secure mobile phone PoP CPU.

Note. This work is presented in detail in Chapters 3 and 7.

Forensic problems of underfill epoxy on electronic components–a solution thanks to an acid attack: application to the chip-off of underfilled components

with David Naccache, Konstantinos Markantonakis, Raja Naeem Akram, and James McEvoy

Abstract. Acids and their use can be of interest in the digital forensics domain. Firstly, they can be used to de-capsulate the packaging of the electronic components before reading the chip (physical dump by chip-on), or to carry out reverse engineering of secure systems by micro-reading techniques of the silicon chip. Moreover, as we show in this paper, with the arrival of the latest generations of mobile phones, acid mixtures can be of interest in the legal transplantation of damaged phones using underfill epoxy.

In addition to traditional high temperature eutectic soldering, the use of underfill epoxy to glue the electronic components to the PCB (memory, CPU, cryptographic chips), has now become the norm amongst mobile phone manufacturers (Apple, BlackBerry, Samsung, etc.). Currently, this technique is the best solution to protect components against various mechanical stresses and improve reliability. Unfortunately, traditional techniques (chip-off or lapping) have become impossible to apply to underfilled components without destroying them or without moving peripheral electronic components. These component movements make the board unusable or require many hours of expensive repairs and specific hardware.

This work introduces a new method called "underfill acid attack". The proposed process is based on the use of different mixtures of acids heated to various temperatures. We quantitatively study the influencing factors on the efficiency of acid attacks on industrial underfill and present our results. Finally, we present our optimised process to unsolder electronic components which are glued together by an industrial high temperature underfill epoxy, without destroying the targeted electronic components and mobile phones' PCBs.

Note. This work is presented in detail in Chapters 3 and 7.

2

Theory

Silicon is the basis of transistor technology and forms the modern electronic cell. Silicon is found in the manufacture of electronic memories, processors, cryptoprocessors, biometric sensors such as fingerprint recognition (touch ID), iris or facial sensors (face ID) and cryptographic chips. Silicon is also the basis of SIM cards, SD cameras, internal GPS memories and more generally of all embedded systems: civil, military, aeronautical and space applications.

In this part we will see the fundamental properties of silicon explaining the generalisation of its use in modern electronics. We will also show the functioning of CMOS transistors and their uses in contemporary memory chips, and finally, we will explain the processes of silicon transformation from a raw element to its most complex form: the silicon nanometric chip.

2.1 General Semiconductor Theory and Fundamental Operations

2.1.1 Semiconductor Technology: Silicon

Quantum mechanical laws, that describe the behavior of matter at the nanoscale, were discovered at the beginning of the twentieth century [Gamow, 1985]. The silicon atom and its electron cloud, like any isolated atomic system, follows the laws of quantum mechanics. Thus, the energy of the silicon electrons can only possess discrete values well defined by the Schrödinger eigenvalues equation [Cohen-Tannoudji et al., 1973] (Equation 2.1):

$$H\Psi_{nlm}(r,\Theta,\Phi) = E_{nl}\Psi_{nlm}(r,\Theta,\Phi) \tag{2.1}$$

with (n) the main quantum number, (l) the orbital quantum number and (m) the magnetic quantum number.

The spherically symmetric potential $\Psi_{nlm}(r, \Theta, \Phi)$ is a function (Equation 2.2) of radial function $R_{nl}(r)$ and angular function $\gamma_{lm}(\Theta, \Phi)$ (spherical harmonic).

$$\Psi_{nlm}(r, \Theta, \Phi) = R_{nl}(r)\gamma_{lm}(\Theta, \Phi) \tag{2.2}$$

Thus, the distribution of silicon electrons in atomic orbitals obeys the Pauli's exclusion principle, the Klechkowski rule and the Hund rule (Equation 2.3, Figure 2.1). By complying with these three quantum rules, it is possible to precisely write the electronic structure of silicon:

$$(1s2)(2s2)(2p6)\ (3s2)(3p2) \sim [N_e](3s2)(3p2) \tag{2.3}$$

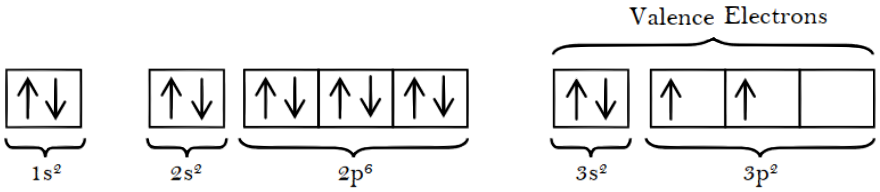


Figure 2.1: Silicon Electron Configuration

The laws of quantum physics [House, 2017] show that when a silicon atom is isolated, the energy levels of its valence electrons are discrete (E1 and E2, Figure 2.2).

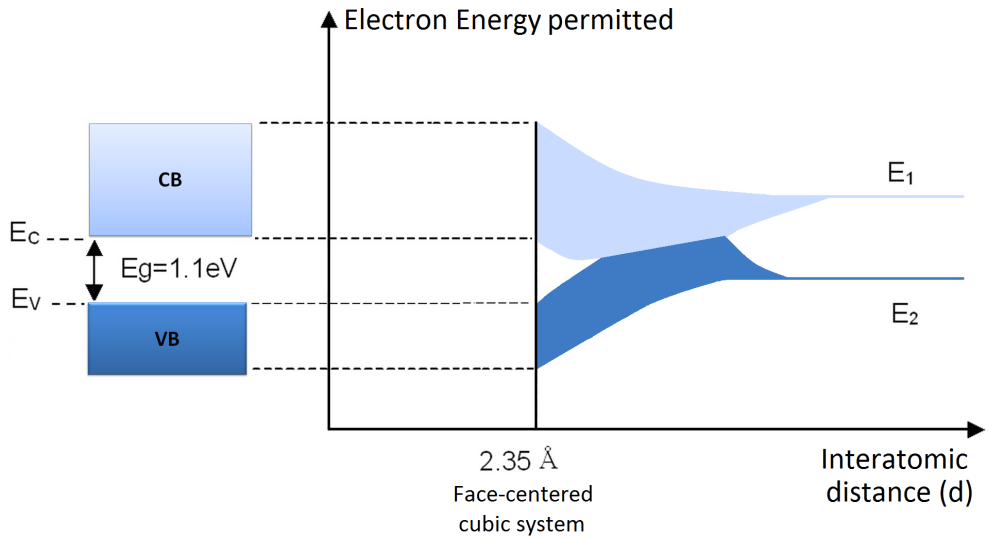


Figure 2.2: Si Atoms in Crystalline Meshes: Valence Electrons' Energy Bands

However, when we bring several silicon atoms closer together (decreasing the interatomic distance), energy levels are divided. We see the hybridisations of atomic orbitals

to give molecular orbitals. If the silicon atomic orbitals $3s^2$ and $3p^2$ are mixed to form molecular orbitals in sp^3 hybridisation [Guzmán-Verri and Lew Yan Voon, 2007], when approaching N silicon atoms, the electron's energy levels are split and being very close to each other, they form continuous energy bands (Figure 2.2).

To explain this phenomenon, it is necessary to express and take into account all the kinetic interactions and energies of the system: kinetic energy of electrons and nuclei, electron-electron interactions, electron–kernel interactions and kernel–kernel interactions. In order not to overload this thesis with lots of equations, we present in **Appendix A** the integral calculations. Thus, in the approximations of Appendix A, Schrödinger's equation for silicon crystal is written (Equation 2.4):

$$[-\frac{\hbar^2}{2m}\nabla^2 + V_c(\vec{r})]\Psi_n(\vec{k},\vec{r}) = E_n(\vec{k})\Psi_n(\vec{k},\vec{r}) \tag{2.4}$$

with $V_c(\vec{r})$ the potential seen by the electrons (crystalline meshes).

The resolution of this equation allows us to obtain the electronic states of the silicon crystal $E_n(\vec{k})$. In their work, [Guzmán-Verri and Lew Yan Voon, 2007], show the analytical solution for the electronic states $E_n(\vec{k})$ for the silicon Si(111) at the Γ point with first-nearest neighbour (1NN) sp3s* and first-nearest neighbour (2NN) sp3 (Figure 2.3):

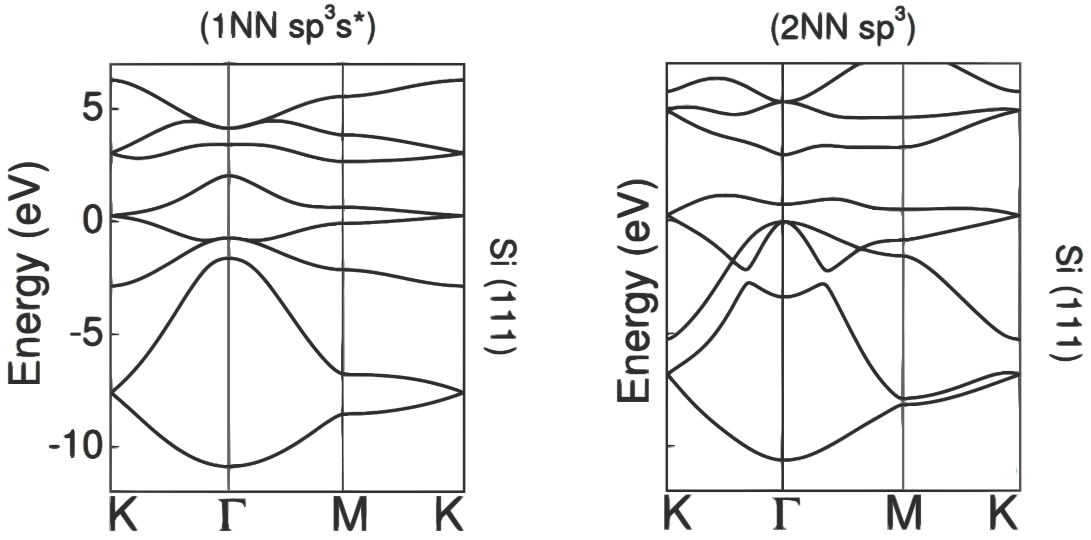


Figure 2.3: Band Structure of Si(111) Obtained from TB Models: First-Nearest Neighbour (1NN) sp3s* (Left), First-Nearest Neighbour (2NN) sp3 (Right)

The cubic system in which silicon is crystallised is a diamond structure (Figure 2.4), derived from the face-centered cubic structure (fcc), with a mesh parameter of 5.43 Å and interatomic distance of 2.35 Å [Henins, 1964].

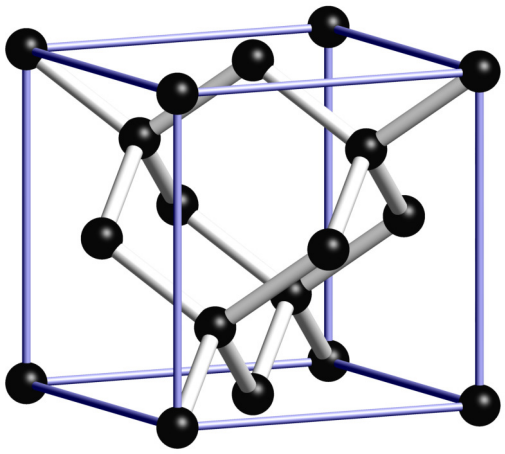


Figure 2.4: Silicon is Crystallised in a Diamond Structure

During the crystal's formation, the atom gains four electrons by forming covalent bonds which correspond to the common sharing of its peripheral electrons with the neighbouring atoms. Thus, a silicon atom that associates with four other atoms of silicon would have eight electrons in its outer fundamental electron layer.

The energy separation Eg between the conduction band and the valence band (Figures 2.2 and 2.3), called band gap energy, determines the electrical behaviour of the body (insulator, conductor, and semiconductor) [Seeger, 2013]:

- **An insulator** is a body whose valence band is saturated while the conduction band is completely empty. These two bands are separated by a band gap with a value $Eg \geq 6\,eV$ (Figure 2.5). Internal electric charges do not flow freely and the property that differentiates an insulator is its resistivity; insulators have higher resistivity than semiconductors or conductors.

- **In a conductor**, a free electron has sufficient kinetic energy to circulate freely in the crystal. The width of the forbidden band Eg is zero and there is then an overlap between the conduction band and the valence band. Thus, the conduction band is partially filled with free electrons that come from the valence band and allow the flow of an electrical current in one or more directions. Pure elemental silver, copper, gold, and aluminium are the best electrical conductors encountered in modern electronics.

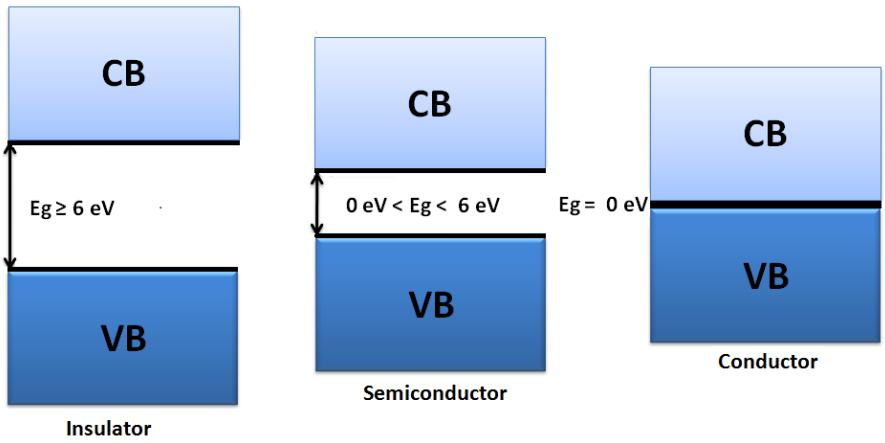


Figure 2.5: Insulator, Semiconductor and Conductor Gap Energies

- Materials having a band gap between $0 < Eg < 6\,eV$, are referred to as intrinsic **semiconductors**. [Rodríguez et al., 2009] shows that the size of the silicon crystal's band gap is dependent on pressure and temperature. As $E_{g,Si}(0,0) = 1.1\,eV$ at 0 K, silicon is a semiconductor (Equation 2.5):

$$E_{g,Si}(P,T) = E_{g,Si}(0,0) - \frac{kT^2}{T+c} = 1.1(eV) - \frac{kT^2}{T+c} \tag{2.5}$$

with k and c are temperature coefficients.

The electrons are half-integral spin particles and taking into account the exclusion principle of Pauli, the electrons obey the Fermi–Dirac statistic [Cohen-Tannoudji et al., 1973]. Thus, $f_n(E)$, the probability of occupying a level of energy E by the n^{th} electron at the thermodynamic equilibrium is Equation 2.6:

$$f_n(E) = \frac{1}{1 + e^{\frac{E-E_F}{k_\beta T}}} \tag{2.6}$$

with E_F the Fermi energy level and k_β the Boltzmann constant.

Silicon is a semiconductor that can be either pure, in which case it is said to be an **intrinsic semiconductor**, or doped by impurities (which makes it possible to control its resistivity), in which case it is said to be an **extrinsic semiconductor**.

The energy state density *N(E)* depends on the electron energy E and corresponds to the available space for the electrons in the conduction band $N_c(E)$ (Equation 2.7) and the space available for the holes in the valence band $N_v(E)$ (Equation 2.8):

$$N_c(E) = \frac{1}{2\pi^2}(\frac{2m_c}{\hbar^2})^{\frac{3}{2}}\sqrt{E - E_c} \tag{2.7}$$

$$N_v(E) = \frac{1}{2\pi^2}(\frac{2m_v}{\hbar^2})^{\frac{3}{2}}\sqrt{E_v - E} \tag{2.8}$$

where m_c and m_v are the effective density mass of states, $\hbar$ the reduced Planck constant, E_c the energy of the conduction band and E_v the energy of the valence band.

The energy state density (conduction or valence band) is the number of states per interval of energy at each energy level available to be occupied.

2.1.1.1 Silicon Without an Electric Field $\overrightarrow{E}$: Thermodynamic Equilibrium

When no electric current is applied to intrinsic or extrinsic silicon, it is said to be in **thermodynamic equilibrium**. In this case, outside the electric field, the free carriers have a Brownian motion, and their average displacement is zero. There is no charge carrier displacement and therefore the silicon is insulating.

The electron density n $[cm^{-3}]$ in the conduction band is then obtained by adding, over the entire energy range covered by this band, the electrons's energy state density weighted by the probability of finding an electron at this same level of energy (Equation 2.9):

$$n = \int_{E_c}^{+\infty} N_c(E).f_n(E)dE \tag{2.9}$$

In the same way we obtain the hole density p (Equation 2.10):

$$p = \int_{-\infty}^{E_v} N_v(E).(1 - f_n(E))dE \tag{2.10}$$

In the case of intrinsic silicon without an applied electric field, the expressions of the electron density n in the conduction band is expressed via the Boltzmann equation (Equation 2.11):

$$n = N_c e^{-\frac{E_c - E_F}{k_\beta T}} \iff N_c = \int_{E_c}^{+\infty} N_c(E).e^{-\frac{E - E_c}{k_\beta T}}dE \tag{2.11}$$

In the same way, the density of holes p in the valence band is (Equation 2.12):

$$p = N_v e^{-\frac{E_v - E_F}{k_\beta T}} \iff N_v = \int_{-\infty}^{E_v} N_v(E).e^{\frac{E - E_v}{k_\beta T}}dE \tag{2.12}$$

For an intrinsic semiconductor $n = p = n_i$ we can deduce the expression of the intrinsic

density of the carriers (Equation 2.13):

$$n_i = N_c e^{-\frac{E_c - E_F}{k_\beta T}} = N_v e^{-\frac{E_v - E_F}{k_\beta T}} \iff n_i = \sqrt{N_c N_v} e^{-\frac{E_c - E_v}{2k_\beta T}} \tag{2.13}$$

We find the position of the intrinsic Fermi level (E_{Fi}) (Equations 2.14, 2.15, and 2.16):

$$\frac{n}{p} = 1 \iff \frac{N_c}{N_v} e^{-\frac{E_c - E_v - 2E_{Fi}}{k_\beta T}} = 1 \tag{2.14}$$

$$\frac{n}{p} = 1 \iff \frac{-E_c + E_v + 2E_{Fi}}{k_\beta T} = \ln \frac{N_v}{N_c} \tag{2.15}$$

$$\frac{n}{p} = 1 \iff E_{Fi}(T) = \frac{E_c - E_v}{2} + \frac{k_\beta T}{2} \ln \frac{N_v}{N_c} \tag{2.16}$$

We notice that at 0 K(Equation 2.17 and Figure 2.6):

$$E_{Fi}(0) = \frac{E_c - E_v}{2} \tag{2.17}$$

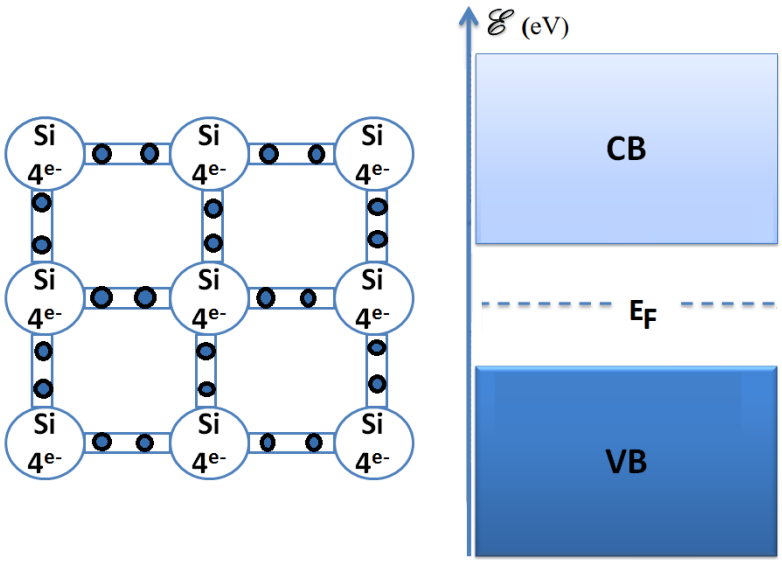


Figure 2.6: Intrinsic Silicon at T = 0 K

The intrinsic silicon may be doped with an electron donor or an electron accepting elements. In this case, the silicon crystal is called an extrinsic semiconductor. A **p-type silicon crystal** (Figure 2.7) is an intrinsic semiconductor in which acceptable impurities have been introduced (boron, indium). These impurities are so-called because they accept an electron from the conduction band to make a bond within the silicon crystal.

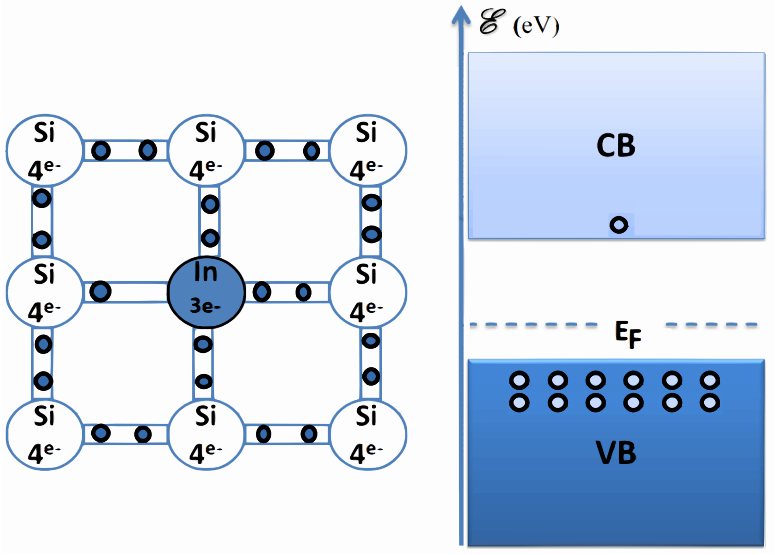


Figure 2.7: P-Type Silicon Crystal Doped With Indium Impurities

The doping of silicon by indium (group III of the Mendeleev table) of concentration N_D allows us to write $N_D \gg n_i$. Thus the density of electrons is equal to the density of donors $n = N_D$ (Equations 2.18 and 2.19):

$$n = N_D = N_c e^{-\frac{E_c - E_F}{k_\beta T}} \iff N_D = n_i e^{-\frac{E_{Fi} - E_F}{k_\beta T}} \tag{2.18}$$

$$n = N_D = N_c e^{-\frac{E_c - E_F}{k_\beta T}} \iff E_F(T) = E_{Fi} - k_\beta T \ln \frac{N_D}{n_i} \tag{2.19}$$

Thus, a p-doped silicon crystal has a lower n electron density and a higher p hole density than the silicon taken in its intrinsic configuration. It is said that the electrons are the minority carriers and that the holes are the majority carriers.

By contrast, an **n-type silicon crystal** (Figure 2.8) is an intrinsic silicon crystal into which donor impurities have been introduced (phosphorus, arsenic). These impurities are so-called because they give an electron to the conduction band to make a bond within the silicon crystal. The doping of silicon by arsenic (group V of the Mendeleev table) of concentration N_A allows us to write the Fermi level as (Equation 2.20):

$$E_F(T) = E_{Fi} + k_\beta T \ln \frac{N_A}{n_i} \tag{2.20}$$

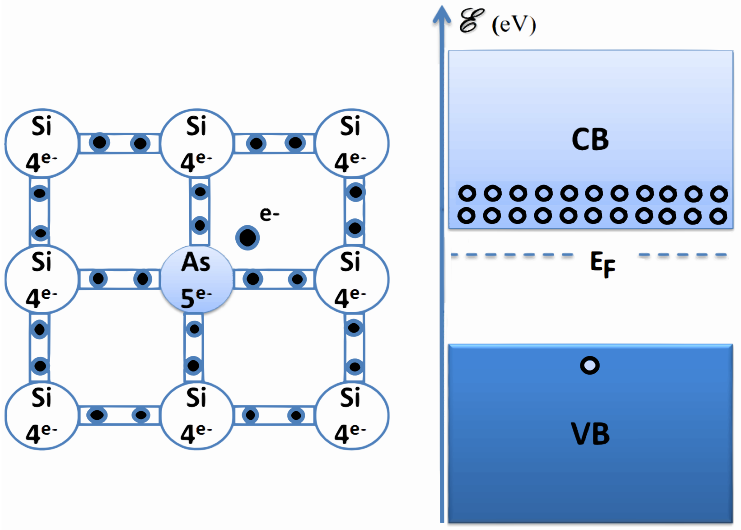


Figure 2.8: N-Type Silicon Crystal Doped With Arsenic Impurities

We see that according to the doping and the concentration of the doping, manufacturers can play on the fundamental properties of silicon to meet their needs. The Fermi level can thus be varied and charge carrier reserves (electrons or holes) can be created.

2.1.1.2 Silicon in an Electric Field $\overrightarrow{E}$: Out Of Equilibrium

As silicon is a semiconductor, the application of an external electric field can move electrons and holes and create a particle flow (electrical current). Intrinsic and extrinsic silicon are said to be **out of equilibrium**; they are electric conductors. When an electric field $\overrightarrow{E}$ is applied to a semiconductor, each charge carrier (electron and hole) undergoes an electrostatic force $\overrightarrow{F} = {}^{\pm}q\overrightarrow{E}$ (+ for the holes and – for the electrons) and a viscous-type friction force $-f_r\vec{v}_d$ which describes the effect of collisions.

The equation of dynamics allows us to write (Equation 2.21):

$$m\frac{d\vec{v}_d}{dt} = q\overrightarrow{E} - f_r\vec{v}_d \iff \frac{d\vec{v}_d}{dt} = \frac{q\overrightarrow{E}}{m} - \frac{d\vec{v}_d}{\tau_r} \tag{2.21}$$

with τ_r the relaxation time of collisions.

Thus, in steady state (Equation 2.22):

$$\frac{d\vec{v}_d}{dt} = 0 \iff \vec{v}_d = \frac{q\tau_r\overrightarrow{E}}{m} = \mu\overrightarrow{E} \tag{2.22}$$

with μ is the mobility of the charge carriers.

The velocities of electrons and holes are written (Equation 2.23):

$$\vec{v}_{electron} = \frac{-q\tau_r \overrightarrow{E}}{m} = -\mu_n . \overrightarrow{E} \quad and \quad \vec{v}_{hole} = \frac{+q\tau_r \overrightarrow{E}}{m} = \mu_p . \overrightarrow{E} \tag{2.23}$$

with μ_n the electrons mobility, μ_p the holes mobility.

The displacement of the charges corresponds to a current whose density is defined as the quantity of charge *dq* which passes through the *ds* surface unit during a *dt* time and for each type of carrier (Equations 2.24 and 2.25):

$$\overrightarrow{J_{electron}} = -n.q.\vec{v}_{electron} = n.q.\mu_n \overrightarrow{E} \tag{2.24}$$

$$\overrightarrow{J_{hole}} = p.q.\vec{v}_{hole} = p.q.\mu_p \overrightarrow{E} \tag{2.25}$$

The conduction current, resulting from the displacement of the electrons and the holes under the action of the electric field, is written (Equation 2.26):

$$\overrightarrow{J_c} = \overrightarrow{J_{hole}} + \overrightarrow{J_{electron}} = \sigma \overrightarrow{E} \;\; with \;\; \sigma = n.q.\mu_n + p.q.\mu_p \tag{2.26}$$

with σ is the conductivity of the materials.

Therefore, to summarise the fundamental properties of silicon :

- If the silicon is intrinsic and with (or without) an external electric field it is insulating.
- If the silicon is p-type and if we apply an external electric field, the silicon is conductive and constitutes a reserve of holes.
- If the silicon is n-type and if we apply an external electric field, the silicon is conductive and constitutes a reserve of electrons.

The introduction of impurities in a silicon crystal makes it possible to modify the number of free carriers, to choose the type of conduction (by electrons or by holes) and to control the conductivity.

We will now see with the juxtaposition of p-type silicon and n-type silicon, insulating materials and conductive materials allow us to obtain the complex electronic structures that are the basis of modern electronics.

2.1.2 Transistor Technologies

As shown in the fundamental equations of section 2.1.1, silicon is a semiconductor which, depending on the treatment it receives, can either drive or block an electrical flux. It is the ideal support to accommodate the millions of transistors that make up the modern chip.

There are two classes of transistors:

- Bipolar transistors [Liu, 1998].
- Metal Oxide Semiconductor (MOS transistors) [Pelgrom et al., 1989].

Each class comprises two types of transistors: the bipolar transistors may be of the NPN or PNP type, and the MOS transistors may be of type N or P. These two types are complementary, i.e. they are controlled by means of electrical quantities of opposite sign:

- The NPN bipolar transistors are controlled by an incoming current (positive)[Ning et al., 1981].
- The PNP bipolar transistors are controlled by an output current (negative) [Ma and Kan, 2017].
- The N-type MOS transistors are controlled by a positive voltage [Brian, 1990].
- The P-type MOS transistors are controlled by a negative voltage [Terao et al., 1991].

In order not to overload this thesis we will only present CMOS technology, which is mainly used in memory components and processors of mobile phones.

2.1.2.1 The CMOS Transistor

The Metal Oxide Semiconductor Field Effect Transistor (MOSFET) is shown as a switch in electronics (Figure 2.9). The electrode, called the gate (G) [Lundstrom, 1997], controls the current (i_D) flowing through a conduction channel established between the source electrode (S) and the drain electrode (D). Thus, the gate acts as a communication electrode from the blocked state (0) to the on state (1) and vice versa.

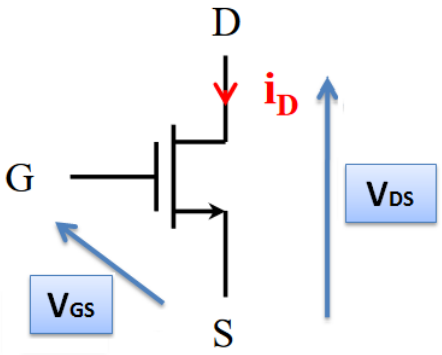


Figure 2.9: CMOS Switch

The conduction channel is created below the gate oxide between the two charge tanks represented by the source and drain regions (Figure 2.10).

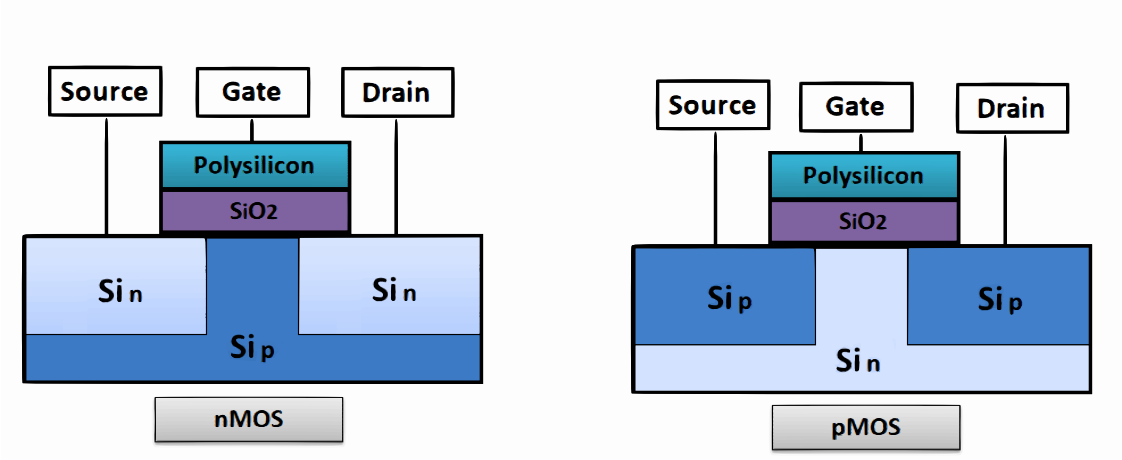


Figure 2.10: nMOS (Left) and pMOS (Right)

It is created by a vertical electric field crossing the semiconductor through the MOS capacitance. Depending on the type of charge carriers forming the conduction channel, the MOSFET is divided into two categories (Figure 2.10):

- The MOSFET is nMOS if the conduction channel is formed of electron.
- The MOSFET is pMOS when the conduction channel is formed of holes.

Applying a voltage to the gate contact has a direct bearing on the nature of the carriers on the semiconductor surface. We can vary the voltage v_{GS} and v_{DS} to change the operating mode of the CMOS. So there are 3 MOFSET modes [Ward and Dutton, 1978]: cutoff, non saturation, and saturation (Figure 2.11).

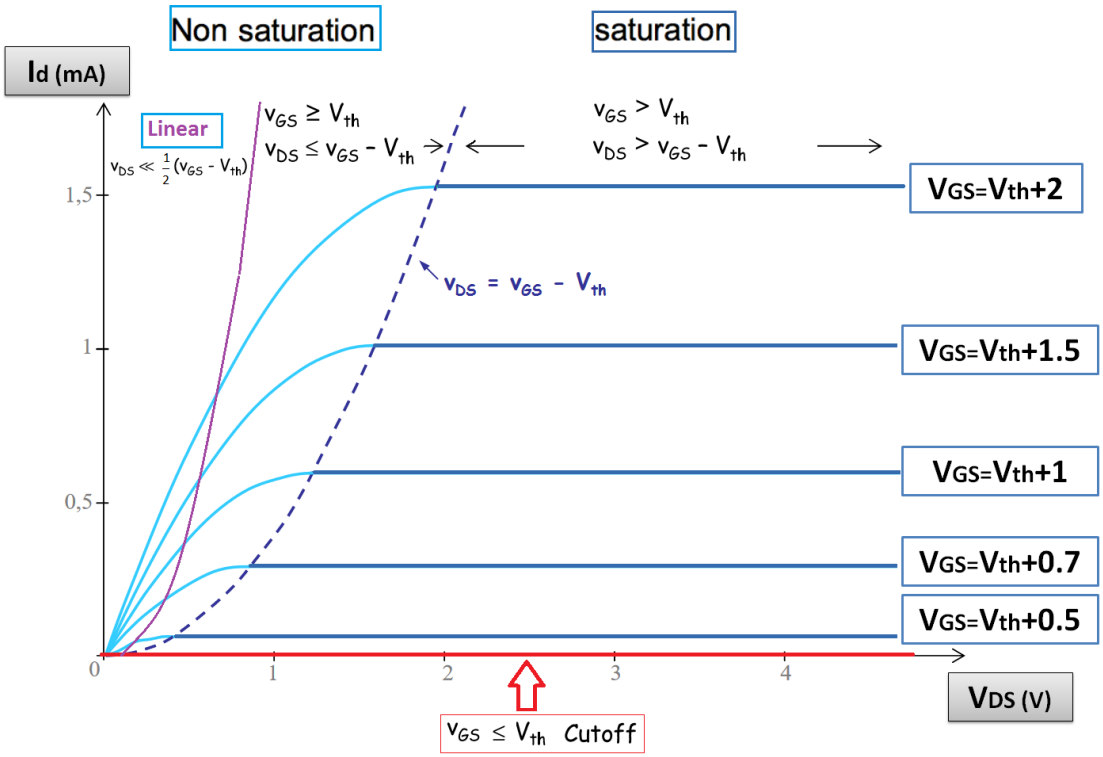


Figure 2.11: Circuit nMOS: $I = f(V)$

We can write [Mathieu and Fanet, 2009] the drain current i_D (Equation 2.27) in terms of v_{GS} and v_{DS} for nMOS transistor:

$$i_D = \mu_n W(-Q_I)\frac{dV_s}{dx} + \mu_n W \phi_t \frac{dQ_I}{dx} \tag{2.27}$$

with Q_I the electron charge of the channel per unit area, (W) the width of the device.

Depending on the threshold voltage of the MOS (v_{th}) we can define a first zone called *cutoff* in which $v_{GS} \leq v_{th}$.

If $v_{GS} = 0$ the cutoff regime is found in *accumulation regime*. The holes, which are the majority carriers of the substrate, are accumulated at the interface. The potential energy of the channel is greater than that of the source and constitutes a potential barrier preventing the flow of charges: $i_S = 0$ the transistor is blocking.

If $0 < v_{GS} \leq v_{th}$ the cutoff regime is found in *depletion regime*. The holes are still present in the interface but less numerous than in the volume of the substrate. The voltage applied to the gate is negative but lower than the threshold voltage. The electrons under the grid are pushed back leaving, only fixed charges in the channel. Thus, a zone of depletion (zone empty of any mobile load) forms under the grid and close to the source and drain zones. As a result, the potential barrier height between the source and the channel decreases but does not release the current as: $i_S = 0$ is being blocked by the

transistor.

If $v_{GS} \geq v_{th}$ *and* $v_{DS} \leq v_{GS} - v_{th}$ the transistor is found in *non saturation regime*. A conduction channel is established and the electric current flows between the source and the drain. The channel behaves first as a resistor and the current varies proportionally with the drain voltage. We call this zone *the linear zone of the non saturation regime* in which $v_{DS} \leq \frac{1}{2}(v_{GS} - v_{th})$. The intensity of the electric current is expressed in the form (Equation 2.28) with L the channel length and k_n a function of the oxide thickness, dielectric constant, and mobility of carriers in the channel:

$$i_D = k_n \frac{W}{L}(v_{GS} - V_{th})v_{DS} \tag{2.28}$$

By increasing the v_{DS} voltage, the drain current deviates ohmically and stabilises due to the reduction of loads in the drain region (*non-linear regime*). The intensity of the electric current is expressed in the form (Equation 2.29):

$$i_D = k_n \frac{W}{L}[(v_{GS} - V_{th})v_{DS} - \frac{v_{DS}^2}{2}] \tag{2.29}$$

If $v_{GS} > v_{tn}$ and $v_{DS} > v_{GS} - v_{th}$ the transistor regime is found in saturation. The increase of v_{DS} has the effect of reducing this excess voltage in the vicinity of the drain, to the point where the voltage vanishes $v_{DS} = v_{GS} - v_{th}$. The channel, which then has a zero thickness, it is said to be pinched. When the transistor is saturated, the current i_D remains constant despite any subsequent increase in v_{DS}. Thus, the intensity of the electric current is expressed in the form (Equation 2.30):

$$i_D = \frac{1}{2} k_n \frac{W}{L}(v_{GS} - V_{th})^2 \tag{2.30}$$

Complementary Metal Oxide Semiconductor technology is the elemental cell with much more complex electronic structures. We will see now how the assembly of CMOS structures allows us to realise an infinity of elementary logical structures.

2.1.2.2 The Inverter in CMOS Technology

It is implemented with a nMOS transistor and a pMOS transistor (Figure 2.12), with V_{DD} the Voltage Drain Drain and V_{SS} the Voltage Source Source. Only one of the transistors is conductive at a time [Weste et al., 2005].

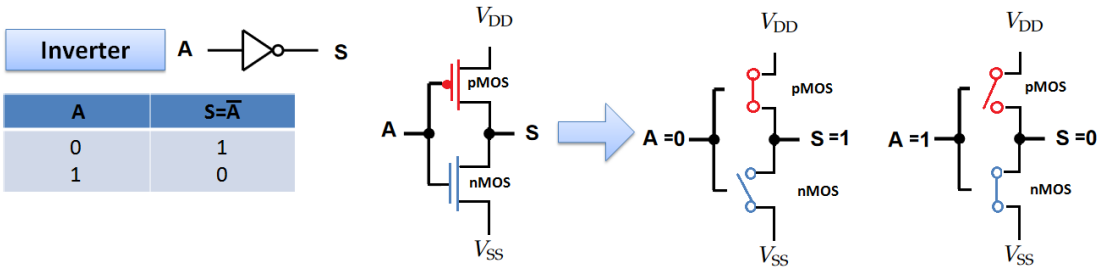


Figure 2.12: Inverter in CMOS Technology

When the input is in a low state, only the pMOS transistor is on. The voltage applied to its source, a high voltage level, is found on the drain which is also the output of the inverter element. When the input In of the inverter is in the logic state 0 (0 V), we have:

- VGS = 0 V for the nMOS; it is blocked.
- VGS = -5 V for the pMOS; it is conductive.

For 5v technology, the output of the inverter is at 5 V which corresponds to logic 1.

When the input is in the high state, only the nMOS transistor is on. The voltage applied to its source, a low voltage level, is found on the drain which is also the output of the inverter element. When the input In of the inverter is in the logic 1 (5 V) state, we have:

- VGS = 5 V for the nMOS; it is conductive.
- VGS = 0 V for the pMOS; it is blocked.

The output of the inverter is at 0 V which corresponds to the logic state 0.

2.1.2.3 The NAND Gate in CMOS Technology

The Complementary Metal Oxide Semiconductor NAND gate [Chen, 1990] contains two series nMOS transistors between the output and V_{SS} and two parallel pMOS transistors between the output and V_{DD} (Figure 2.13). The purpose of this logic function is to output 0 when both inputs are at 1 and output 1 in all other cases.

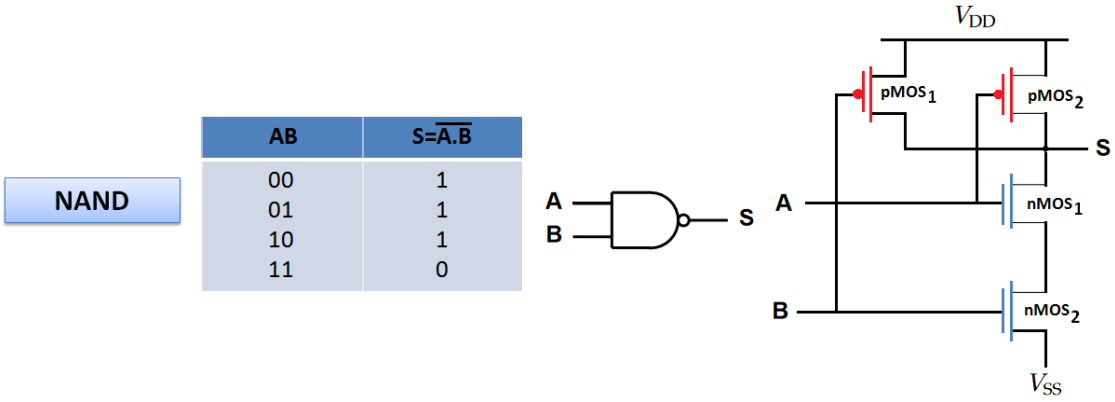


Figure 2.13: NAND Gate in CMOS Technology

When V_A is low state and V_B is low state, both pMOS are conductive and both nMOS are blocked. Thus, the electronic flow at the output Vs has two possible paths, via the pMOS transistors, to connect to V_{DD} and create a movement of the charge carriers (electric current). The output S will be loaded at the V_{DD} level. The output line will not get any path to the GND because bith the nMOS are disabled.

Therefore, there is no possible passage for the electrons by which the output line can discharge. The output line maintains the voltage level at V_{DD} and thus Vs is in a high state. The output logic state is 1.

When V_A is in a low state and V_B in a high state (or similar when V_A is in a high state and V_B in a low state), we have a pMOS1 conductor, nMOS1 blocked, pMOS2 blocked and a nMOS2 conductor. pMOS1 and pMOS2 are two parallel transistors.

Thus, even with blocked pMOS2, the electron flow will get a channel via pMOS1 to connect to V_{DD}. nMOS1 and nMOS2 are two transistors in series. As nMOS1 is blocked, V_S will not be able to find a path to GND to unload its electron stream. This consequently entails that the V_S is maintained at the level of V_{DD}. The output logic state is 1.

Finally, when V_A is low and V_B is high, pMOS1 is blocked; nMOS1 is conductive, pMOS2 is blocked and nMOS2 is conductive. Both pMOS are blocked.

Thus, V_S will not find any way to create a displacement of its electron flow and create a connection to V_{DD}. Since nMOS is on, the connected nMOS series will create a movement from its V_S to GND electron flow. The output logic state is 0.

2.1.2.4 The NOR Gate in CMOS Technology

The CMOS NOR gate [Chen, 1990] contains two nMOS parallel transistors between the output S and V_{SS} and two pMOS series transistors between the output and V_{DD} (Figure 2.14).

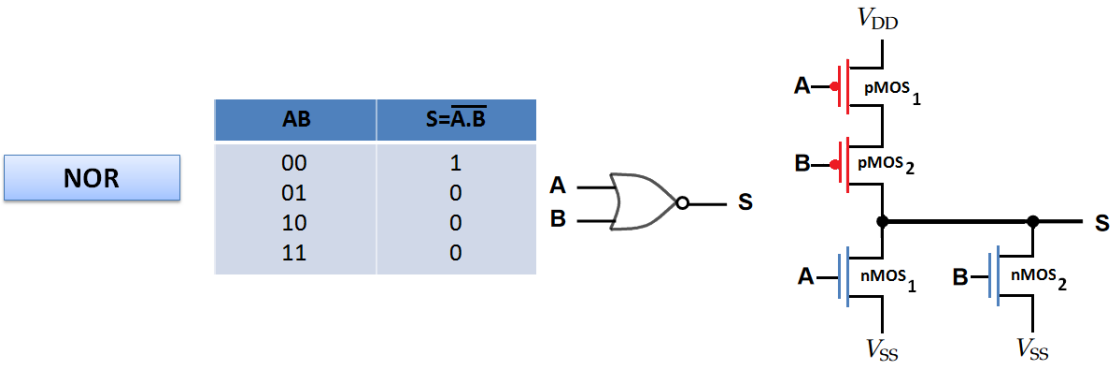


Figure 2.14: NOR Gate in CMOS Technology

The same reasoning as for the NAND gate can be carried out:

- When V_A is in a low state and V_B is in a low state: pMOS1 is conductive, nMOS1 is blocked, pMOS2 is conductive and nMOS2 is blocked. The output logic state is 1.
- When V_A is low and V_B high: pMOS1 is conductive, nMOS1 is blocked, pMOS2 is blocked and nMOS2 is conductive. The output logic state is 0.
- When V_A is high and V_B low: pMOS1 is blocked, nMOS1 is conductive, pMOS2 is conductive and nMOS2 is blocked. The output logic state is 0.
- When V_A is high and V_B high: pMOS1 is blocked, nMOS1 is conductive, pMOS2 is blocked and nMOS2 is conductive. The output logic state is 0.

2.1.2.5 Complex Electronic Circuits

We have just discussed three elementary logical cells (inverter, NOR gate and NAND gate) produced from CMOS transistors. The reader is invited to consult [Nielsen and Girgis, 2000] and [Chen, 1990] for a complete description of the other elementary logic cells obtained from CMOS transistors. Building complex electronic circuits require the sharing of thousands of CMOS cells in series and in parallel.

These interconnections are a long process to wait for the execution of a particular function (Figure 2.15).

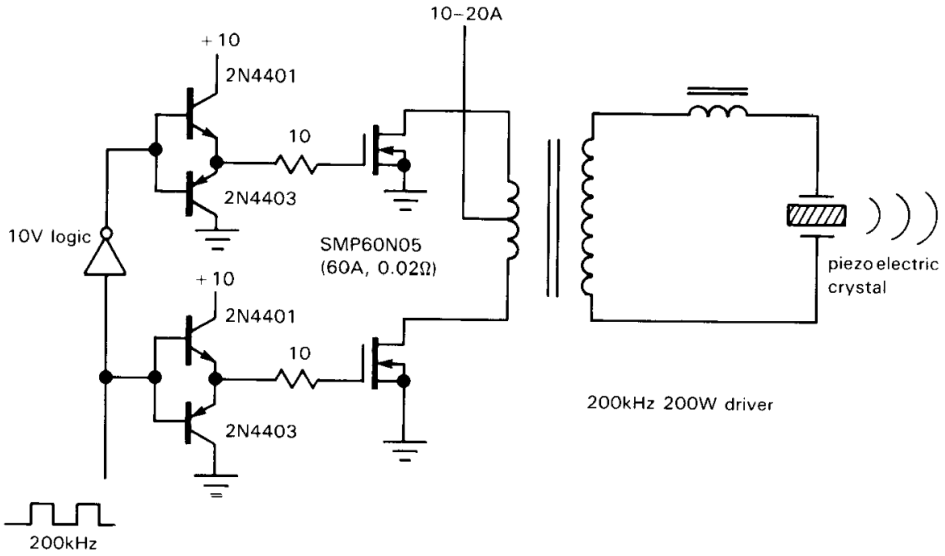


Figure 2.15: Bipolar Power Piezo Driver [Horowitz and Hill, 1989]

In order to make complex circuits, designers use HDL hardware description language such as Verilog [Thomas and Moorby, 2008], VHDL [Ashenden, 2010] or SystemC to represent the behaviour and architecture of a digital electronic system [Bushnell and Agrawal, 2004]. The design tools allow the direct transition from a functional description in HDL to a schematic of several thousand logic gates that will be used and optimised by electronic circuit manufacturers.

We have just seen that the CMOS transistor is the fundamental element on which a basic and complex logical structure is based. We will show how the arrangement of these elementary cells makes it possible to realise different types of memory in the silicon chip.

2.2 The Various Types of Silicon Memories

Silicon memories fall into two main categories: volatile memories and non-volatile memories [Bez et al., 2003] (Figure 2.16).

- Volatile memories lose their information as soon as they are no longer powered.
- Non-volatile memories retain information stored independently of the external power supply.

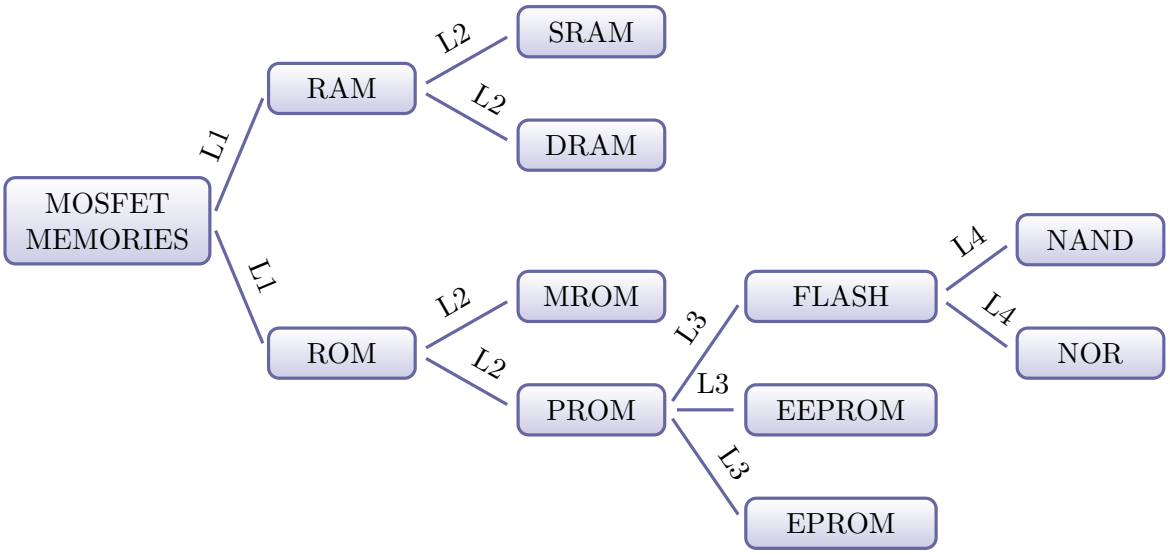


Figure 2.16: MOS-Memories Tree

2.2.1 Volatile Memories

Volatile memories are commonly referred to as Random Access Memory (RAM). Two types of RAM memory exist: Static Random Access Memory (SRAM) and Dynamic Random Access Memory (DRAM).

The SRAM [Bhavnagarwala et al., 2001] can retain the information stored on it as long as a voltage is applied across its terminals. Each bit is stored using flip-flops made up of 6 transistors (Figure 2.17).

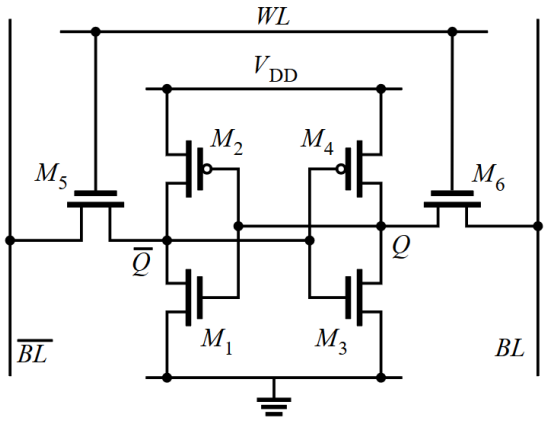


Figure 2.17: SRAM Logic Cell

The area occupied by these memories is therefore large. However, it is not necessary to refresh it regularly, so it consumes very little current. Its very fast reading and writing

makes it a wise choice for microprocessor cache memories.

The DRAM, or dynamic memory [Adler et al., 1995], must be regularly refreshed to keep the information stored on it. It therefore consumes more current than the SRAM. However, it consists only of an access transistor and of a capacitor storing a load (Figure 2.18), which makes it possible to obtain high integration densities.

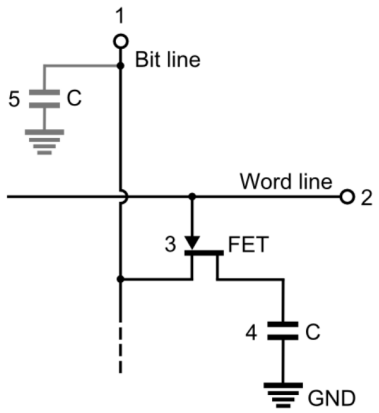


Figure 2.18: DRAM Logic Cell

This reduces the cost of DRAMs and makes them widely used as the main memory of computers. Examples include Samsung, SK Hynix, Microns, Elpida, but also the Corsair Dominator Platinum 16GB DDR4, the Ballistix Tactical 16GB DDR4 CL16 or the Kingston ValueRAM 16GB DDR4.

2.2.2 Non-Volatile Memories

Non-volatile memories are used to store information in the absence of power. The ROM (Read Only Memory) was the first non-volatile silicon-only memory invented. The data is written during manufacture, by etching, and cannot be subsequently modified. It is useful in microcontrollers. Two types of ROM memory exist: the MROM (Mask ROM) and the PROM (Programmable ROM). The PROM is an evolution of the ROM in which the memory can only be programmed once. Its operation corresponds to a fuse array whose elements are blown according to the information to be stored. The Erasable PROM (EPROM) [Shiner et al., 1980] is an improvement of the PROM where data can be erased by ultraviolet (UV) exposure. It is the first device to use the principle of storage of charges in a floating gate. The memory cell consists of a single transistor. The EEPROM (Electrically EPROM) [Harari et al., 1994] differs from the EPROM in that it can be electrically erased without UV exposure. This is done at the expense of the addition of a selection transistor for each storage transistor, which increases the area

occupied by these memories.

Finally, Flash memory is similar to the EEPROM, without the selection transistor, whose role is now played by the storage transistor. The surface gain is therefore consistent for the same functionalities.

There are two types of flash memory architectures: NOR and NAND (Figure 2.19).

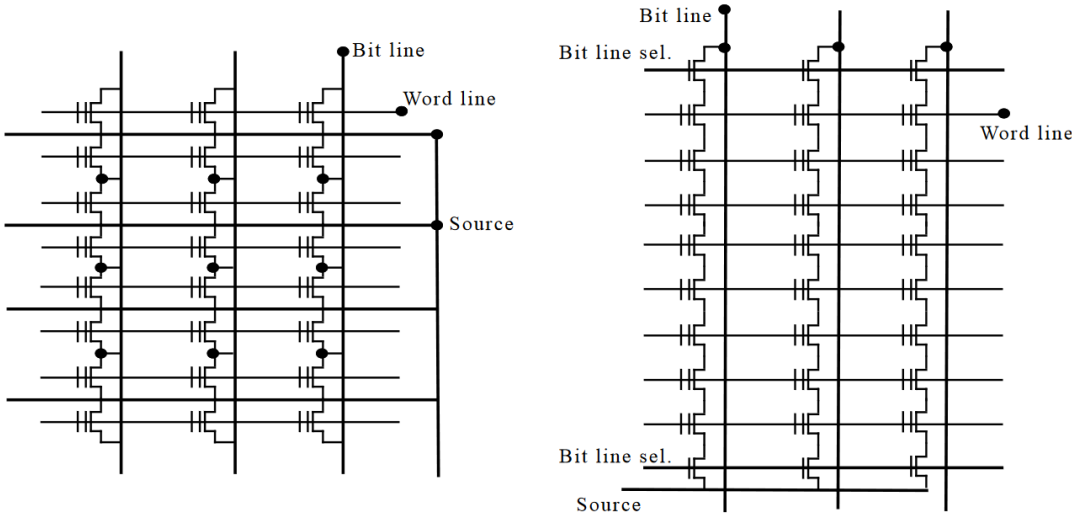


Figure 2.19: NOR and NAND [Campardo et al., 2010]

In NOR flash memory [Chen et al., 2009], memory cells are connected in parallel with a common ground node. Bitlines are formed by memory cells sharing the same drain contact and wordlines are formed by flash cells sharing gate contact.

NAND flash memory employs a different array organisation [Ueno and Uchiumi, 2010]. In NAND flash, several memory cells are connected in series between bit line and ground, thus increasing the density compared to NOR flash. Although series connection of memory cells increase density in NAND flash, it also reduces the current for read operation. Reading a single memory cell requires being able to read other cells in the same bit line, therefore a NAND flash memory cannot provide fast random access and is usually employed as a serial memory. The elementary structure of the NAND cell (or floating TOR for NOR flash cell) is based on that of a Complementary Metal Oxide Semiconductor transistor (Figure 2.20) but is not CMOS. The difference is that a trapping layer, called a floating gate, is added to the grid stack.

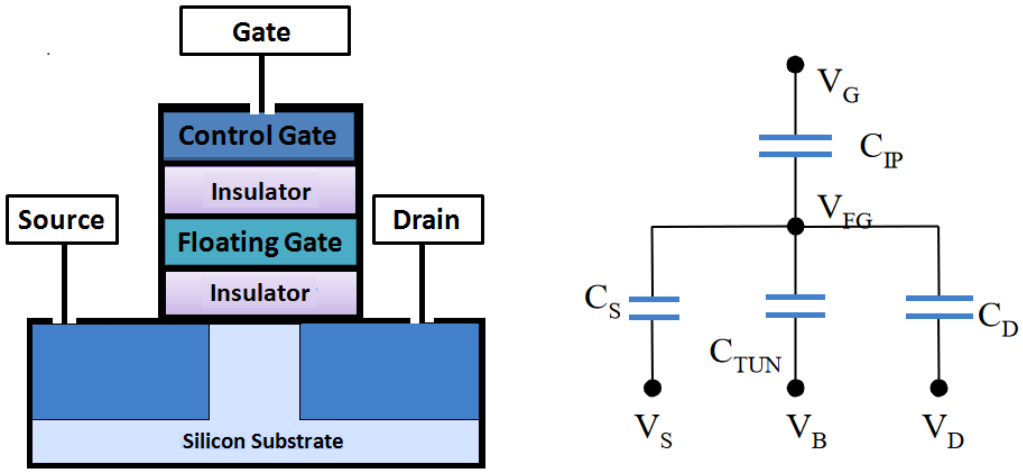


Figure 2.20: NAND Flash Logic Cell

This floating gate is isolated from the control gate and the substrate by two oxides: the tunnel oxide and the interpoly oxide, also known as control oxide. The floating gate is charged with electrons when a sufficient electric field is applied between the substrate and the floating gate. The application of the Gauss theorem to the floating gate makes it possible to establish the floating gate voltage (Equation 2.31), [Morancho, 2004]:

$$\begin{aligned} V_{FG} = &\frac{Q_{FG}}{C_{IP}+C_{TUN}+C_D+C_S} + \frac{C_{IP}}{C_{IP}+C_{TUN}+C_D+C_S}V_G \\ &+ \frac{C_D}{C_{IP}+C_{TUN}+C_D+C_S}V_D + \frac{C_S}{C_{IP}+C_{TUN}+C_D+C_S}V_S \\ &+ \frac{C_{TUN}}{C_{IP}+C_{TUN}+C_D+C_S}V_B = \frac{Q_{FG}+C_{IP}V_G+C_DV_D+C_SV_S+C_{TUN}V_B}{C_{IP}+C_{TUN}+C_D+C_S} \end{aligned} \tag{2.31}$$

with Q_{FG} the load stored in the floating gate and C_{IP} the capacity between the floating gate and the control gate.

This electric charge interacts with the potential applied to the control gate during reading and causes the threshold voltage to shift to positive voltages (Figure 2.21, [Bez et al., 2003], Equation 2.32):

$$V_{Th} = -\frac{Q_{FG}}{C_{IP}} \tag{2.32}$$

To read a cell, a voltage is applied to the control gate and the current flow from the source is probed. If there is no current, this means that the floating gate is loaded (0). If there is a current, the floating gate is not loaded (1).

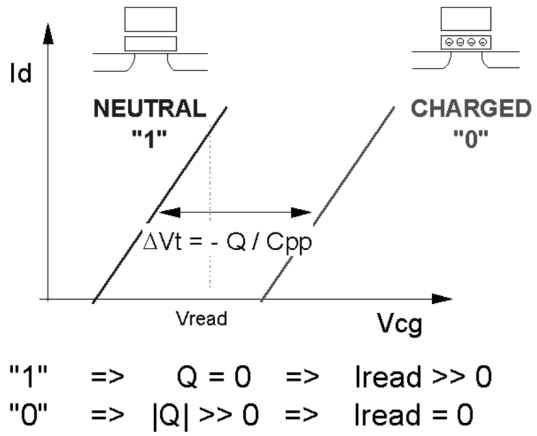


Figure 2.21: NAND: Floating-Gate MOSFET Reading Operation

To write a cell, a high voltage is applied to the control gate and electrons move from the silicon substrate to the floating gate. This process is called tunneling as the electrons "tunnel" through the oxide insulator to reach the floating gate.

To erase a NAND cell, a high voltage is applied to the silicon substrate and the electrons move from the floating gate to the silicon substrate. This uses the same tunneling process as the writing process. We have just seen that the elementary cell of the NAND can handle two voltage levels (0 and 1). We note that manufacturers are now able to store several bits in the same cell. [Bez et al., 2003] (Figure 2.22).

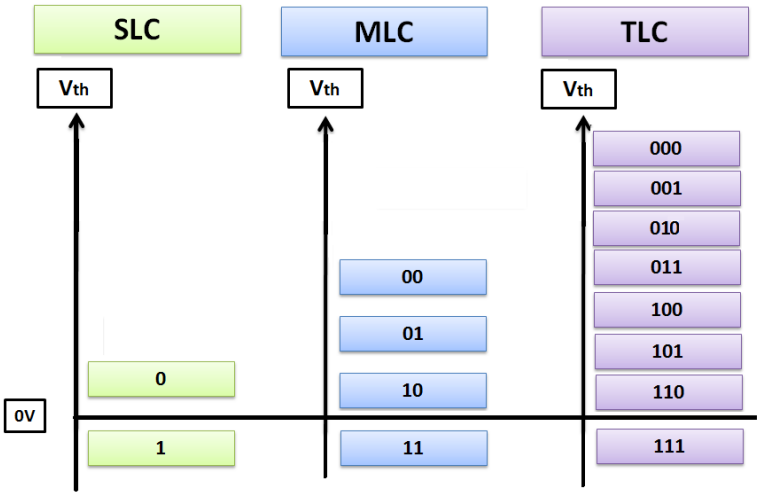


Figure 2.22: SLC, MLC and TLC Memories Levels

They are able to manipulate the load stored in the Q_{FG} floating gate and thus modify the Vth voltage on several levels.

Thus, the MLC memories [Kanda et al., 2008] support several voltage levels, which

makes it possible to store four different values (8 values in the case of the TLC memory [Ho, 2005]). TLC memory was used on Apple 64 GB iPhone 6 and 128 GB iPhone 6 Plus models.

We have just seen that the nMOS/pMOS transistors are the fundamental element on which the basic electronic structure is based. Elemental logic structures are then used to produce more complex electronic circuits. We will now present the manufacturing process of microchips and show how millions of transistors can be miniaturised enough to be implanted in modern microchips.

2.3 Chip Memory Manufacturing Process

The starting point for the manufacture of microchips is a pure piece of silicon extracted from sand by reduction. To obtain pure silicon with a perfect monocrystalline structure, the piece (polysilicon) obtained by reduction is heated to 1420 °C in an argon-gas-purged hermetic furnace in order to remove any residual traces of air ([Feigelson, 2004]). The molten silicon is then processed in a crucible using the Czochralski process ([Czochralski, 1918]). The block is then cut (wire saw) into 1 mm slices, called wafers, with a diameter up to 300 mm (Figure 2.23).

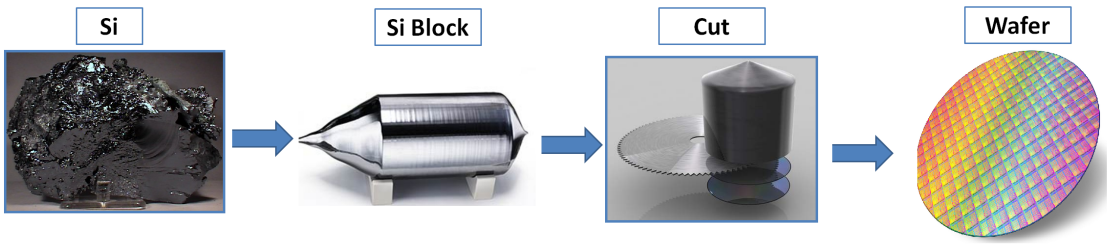


Figure 2.23: From a Piece of Silicon to Wafer

Since the wafer must have a perfectly flat structure, the micro-marks left during cutting are reduced by a mechanical and then chemical lapping process. The entire manufacturing process of an electronic component is carried out in a clean room with a strictly controlled environment (humidity, light, dust, temperature) in order to avoid any particle and light contamination that would seriously harm the components. In the clean room the air is constantly renewed and filtered. The room contains 2 million times less dust than outside air. At the chip scale, a tiny grain of dust represents a potential barrier that blocks the paths dedicated to the circulation of the electrons that transmit the signal and the information.

Once the blank wafer is obtained, in order to give it the function of an electronic

circuit, doped areas, oxide zones and contact zones (electrical conductor) must be added to the wafer to create the elementary cells, transistors (Section 2.1.2) and thus complex electronic circuits (Section 2.1.2.5). The general manufacturing process (transistors) of electronic chips on silicon is called photolithography. The different regions are etched directly on the base silicon support, adding dopants, depositing metal (aluminium or copper) and oxide (insulator).

The lithography (Figure 2.24) consists of the miniaturisation of an electronic circuit and its micrometric printing on the wafer of silicon. The wafer is coated with a photo-sensitive chemical varnish which hardens when in contact with ultraviolet light (coating).

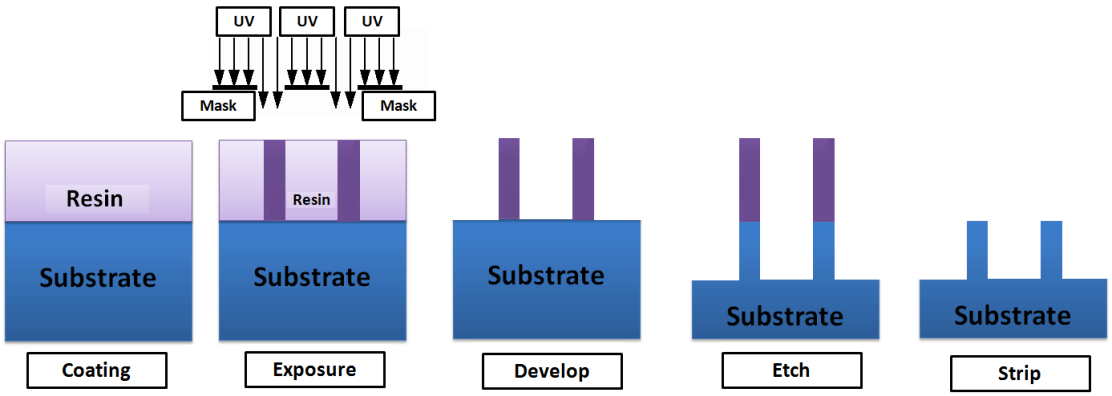


Figure 2.24: Principle of Lithography

In a darkroom, a light ray (UV for photolithography) is focused on the image of the circuits before passing through a miniaturisation lens before reaching the varnished wafer. Electron-beam (e-beam) lithography writes directly without mask involved. This step is called "exposure.

The exposure to light causes a chemical change that eliminates the photoresist with a special solution called "developer". Positive photoresist, the most common type, becomes soluble in the developer when exposed; with negative photoresist, the unexposed regions are soluble in the developer.

Then, in the etching process, the chemical agent removes the top layer of the substrate in the areas that are not protected by photoresist. Once a photoresist is no longer needed, it must be removed from the substrate. This usually requires a liquid "resist stripper", which chemically alters the resistance so that it no longer adheres to the substrate.

The wafer preserves, on its first level, the drawing of the circuit exactly as a developed photograph. The successive layers are then produced by stacking an electronic circuit

pattern (Figure 2.25). Each design consists of an assembly of insulators, doped semiconductors and conductive parts. In total, there are more than 2,000 operations involved in creating a microchip.

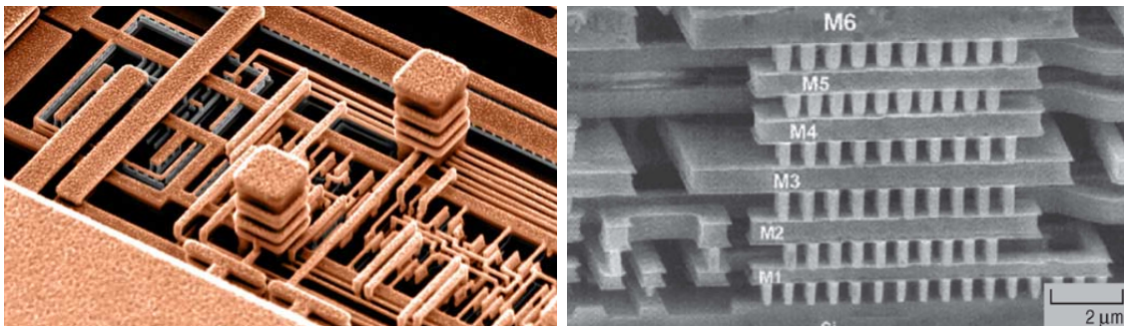


Figure 2.25: Metalisation

The resolution R of the photolithography is given by the Rayleigh equation (Equation 2.33):

$$R = -k_1 \frac{\lambda}{N_A} \tag{2.33}$$

where λ is the wavelength of the exposure light, k1 is a constant determined by the parameters of the optical elements used, the mask type, and N_A is the numerical aperture of the lens.

Thus, for UV sources using KrF (krypton fluoride laser) the spacial output is about 250 nm. For sources using ArF (argon fluoride) the resolution is about 190 nm. Researchers [Stulen and Sweeney, 1999] also have developed Extreme Ultra Violet (EUV). The EUV light is not emitted by the laser, but rather by a tin or xenon plasma which is excited by an excimer laser (allowing a 13.5 nm EUV source). This technique allowed the development of the technological node of 28 nm [Planes et al., 2012], 14 nm [Liu et al., 2013], 10 nm [Boeuf, 2017] and 7 nm in 2016 [Ha et al., 2017]. In June 2016, IBM announced the development of this new technique on its latest generation chips [1].

Electron beam lithography is a maskless lithography technique (MLL) that operates with an electron beam [Carley, 1973]. The patterns are drawn directly using the electron beam that scans the resin deposited on the wafer. This lithography, however, is a slow technique and low yield, but very flexible compared to the geometry of the exposed patterns. Due to its low efficiency, electron beam lithography is hardly used in industrial production but common in university research or for fixing errors.

The wavelength associated with electrons depends on the energy at which they are

[1] https://www.ibm.com/blogs/research/2016/12/advancing-toward-7nm/

accelerated. For an electron acceleration at 50 kV, $\lambda = 0.0248$ nm and at 100 kV, $\lambda = 0.0124$ nm. As a result of the electron–matter interactions, the practical resolution of electron beam lithography is not determined by the size of the electron beam, but by the widening of the beam in the electron-sensitive resin.

The finished piece of silicon is then assembled and connected via gold wires called bonding wires (Figure 2.26).

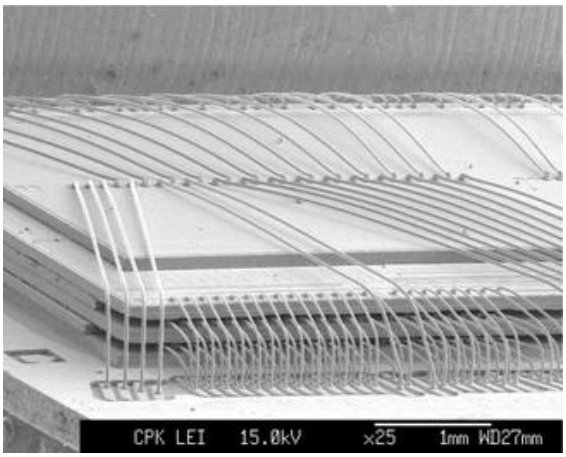


Figure 2.26: Internal Chip Package

Thus, after bonding of the silicon to the contact pins, the component is encapsulated in an insulating epoxy resin.

There are several types of packages (Figure 2.27). Among the main ones found in the forensic area we have: Ball Grid Array (BGA), Land Grid Array (LGA), Small Outline Package (SOP), Thin Small Outline Package (TSOP), Dual Inline Package (DIP), Thin Quad Flat Package (TQFP), Small-Outline Integrated Circuit (SOIC), Pin Grid Array Package (PGA), Plastic Small-Outline Package (PSOP), Quad-Flat No-leads (QFN), Micro-Lead Frame (MLF), and Plastic Leaded Chip Carrier (PLCC).

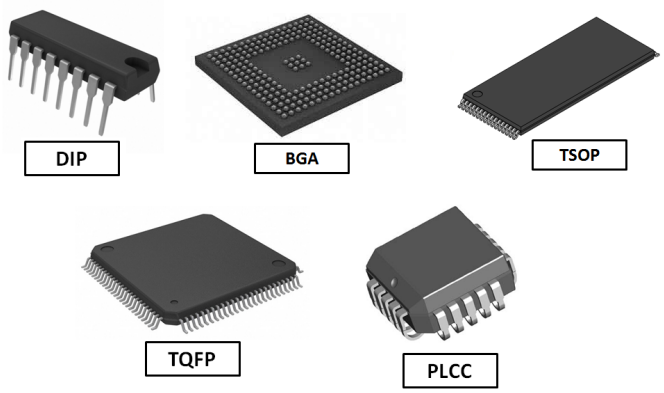


Figure 2.27: External Chip Packages

Once the component is packaged, the memory (or CPU) chip's manufacturing process is complete. The component is finally soldered to the Printed Circuit Board (PCB). It is the assembly of several types of components on multi-layer PCBs that allows major technological developments in the field of modern electronics (mobile phone, GPS, etc.). Among the main components can be found: memories, processors, capacitors and resistors. There are also specific components such as crypto-components, baseband processor, NFC chip, MP3 chip, barometric pressure sensor chip, display power chip, etc. If the example of the iPhone 7 is taken (Figure 2.28), this phone is an assembly of several hundred components that all have a special role.

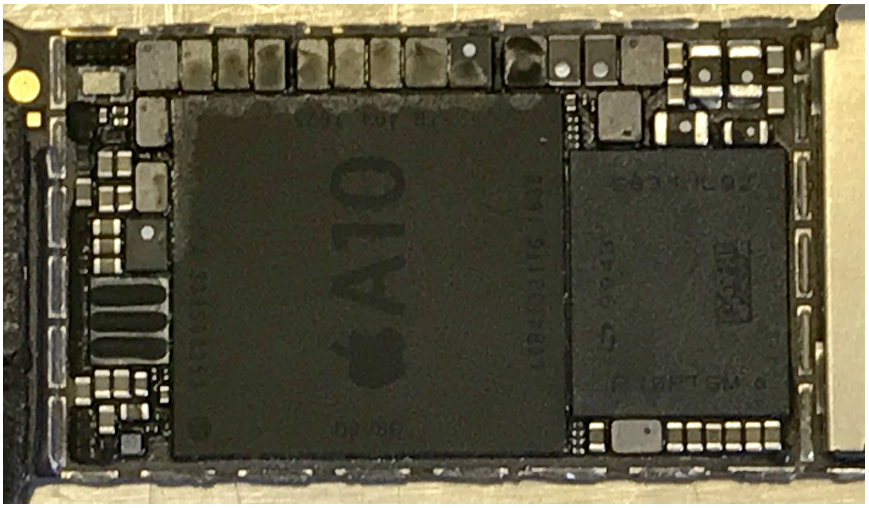


Figure 2.28: Assembly of Electronic Components (A10 CPU, baseband, and capacitors) on the Multi-Layer PCB of the iPhone 7

Most of the components are millimetric and assembled on the PCB by a specific process for each component. The malfunction of one of these components often leads to a fatal error of the major functions of the phone.

The miniaturisation of these components is the key point of current performance, with the assembly of a large number of components on a single electronic card. It will be seen in the thesis that this miniaturisation and the precision of the components' assembly on the PCB make the investigator's work more and more challenging.

3

Forensic Investigators Facing Data Recovery From Mobile Phones

When a forensic investigator wants to analyse a mobile phone, he must first, even before handling the phone, consider what information he needs for his investigations: call log, deleted data, GPS locations, emails, email headers, file system, text messages (SMS), phonebook, pictures, EXIF data on images, Wi-Fi networks, hidden files, multimedia content, multimedia messaging service, web history, space analysis allocated or not allocated, etc.

Thus, based on their needs, their knowledge of the mobile phone and the equipment they have, investigators can carry out different types of extractions that give him access to different levels of information [Brothers, 2009].

3.1 Traditional Forensic Techniques Used to Extract Data from Undamaged Mobile Devices

Forensic investigators consider [Ayers et al., 2014] three types of extraction (manual, logical and physical) corresponding to five abstraction levels (Figure 3.1):

- Manual extraction as level 1.
- Logical extraction as level 2.
- Physical extraction as levels from 3 to 5:
  - Physical extraction using hex-dumping/JTAG as level 3.
  - Physical extraction using chip-off/chip-on and memory read as level 4.
  - Physical extraction using micro-read as level 5 [Brothers, 2009].

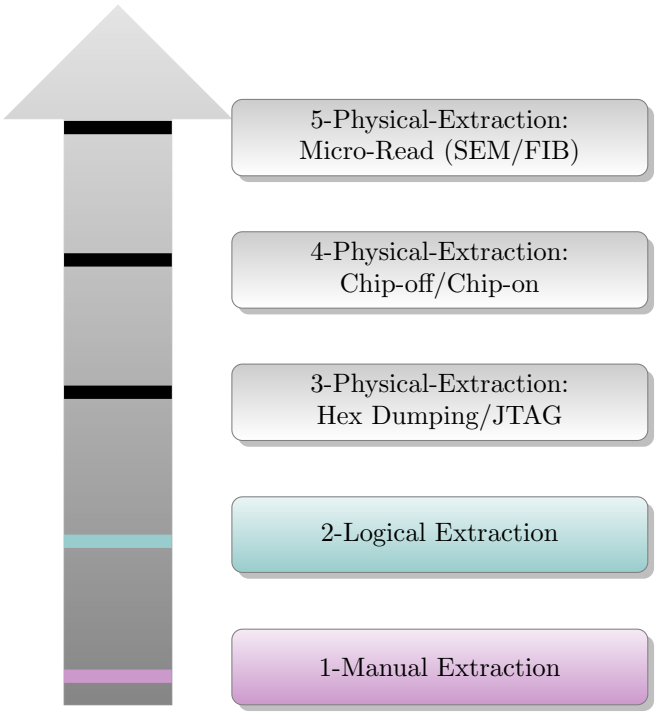


Figure 3.1: Mobile Phone Extraction Levels: Ranked by Difficulty Levels

It should be added that the higher the level of extraction, the higher the cost of the materials required for the extraction process, the more invasive the technique will be, and the more specialised the forensic investigator will need to be. The choice of extraction level is part of the investigator's expertise. He will have to adjust his action according to the information he has to extract, the time he has, and the financial cost of the intervention.

3.1.1 Manual Extraction

In many cases encountered by investigators, manual extraction is the only possible extraction technique. This technique consists of making copies of the phone screen when investigators manually investigate the phone (using the keyboard or touchscreen). The extraction level 1 allows forensic investigators to capture only images and videos of the mobile phone screen or manual copying of phone information. They often use manual devices (digital camera) or hardware devices specially developed by private companies: Eclipse[1] (Figure 3.2), ZRT[2], UFED cam[3], XRY cam[4], etc.

[1] http://www.teeltech.com/mobile-device-forensic-tools/eclipse-3-pro-kit/
[2] http://www.fernico.com/ZRT3.aspx
[3] http://www.cellebrite.com/Pages/ufed-camera
[4] https://www.msab.com/products/xry/xry-camera/

Figure 3.2: Eclipse 3 Pro Kit

This method can, however, be time consuming if the investigators have to photograph each SMS, each email, each picture, each contact, etc.

This technique is also difficult to apply when the phone screen is damaged (visualisation, touchscreen) or when certain keyboard keys are not functioning or defective. This level of extraction is, however, very often used in surveys conducted on the latest generations of secure mobile phones (BlackBerry, Apple, Samsung, etc.). The mobile phone's internal security policy (Figure 3.3) does not allow the investigator to connect the phone to an external device or scan with a computer.

Figure 3.3: BlackBerry PGP Security Policy: Backup Unauthorised

On this kind of secured phone, by changing any of the security settings, an overall memory erase will occur. The investigator then has no other alternative but to perform a manual extraction. In some cases, forensic investigators must also act in emergency

conditions in order to retrieve the mobile phone's latest data in cases of a missing person, a terrorist attack, or an air crash. In complex cases, such as a crime scene with Chemical, Biological, Radiological and Nuclear (CBRN) risks, the investigators cannot remove the electronic object from a restricted perimeter because it is considered contaminated (Figure 3.4).

Figure 3.4: IRCGN's Investigators Act in CBRN Emergency Conditions

This is why investigators often use manual extraction directly at the crime scene with their CBRN protective equipment. Thus, except in the cases mentioned above, the manual extraction technique is also time-consuming and does not recover the phone's deleted data, locked devices (unknown password, locking path, fingerprint lock, locking by iris or facial recognition), or data from a damaged phone. However, manual extraction is often used to verify and validate other types of extraction (automatic extractions, etc.).

3.1.2 Logical Extraction

Logical extraction [Kim et al., 2007] allows investigators to find visible elements via the file system: call logs, short message services, contacts, application data, the phone's International Mobile Equipment Identity (IMEI), images, multimedia messaging services, videos and audio files.

The investigator needs to know the phone model, as well as the Operating System (OS) version, because this information directly determines the correct connection to be used for the logical extraction: Universal Serial Bus (USB) connection, serial protocols,

Bluetooth protocols, Application Programming Interface (API), proprietary commands, etc.

Logical extraction usually consists of an extraction performed by API. The investigator therefore uses logical protocols between the mobile phone and the analysis computer. Thus, by sending API commands or client/server architecture commands, the investigator communicates directly with the mobile device's operating system. However, he is only able to request the data extraction that is accessible through the operating system (Figure 3.5). Forensic investigators can also set the mobile phone in "diagnostic mode" and have direct access to the file system (example: Embedded File System, EFS). Data are finally recovered using the specific manufacturer's protocols.

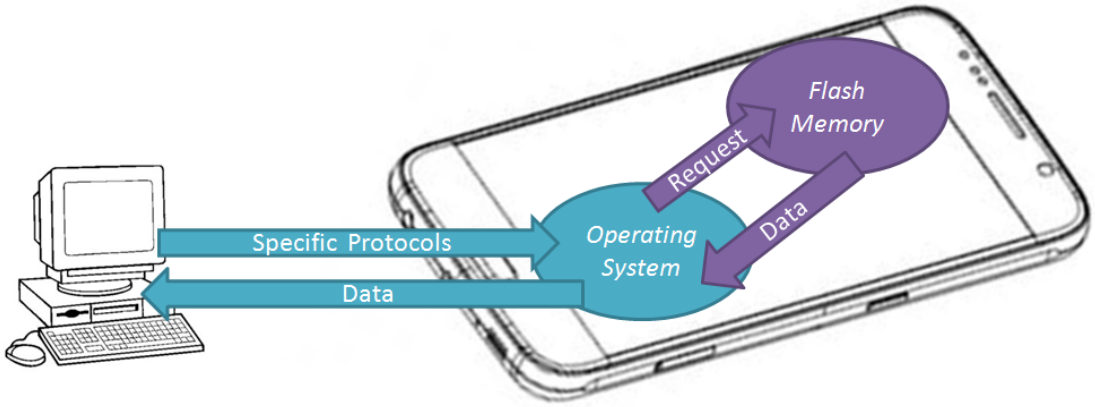


Figure 3.5: Logical Extraction Communication Concept

Communication via the operating system is not able to extract erased files, which a have a lower level of information. As is often the case when a criminal has consulted child-pornographic or terrorist websites, the investigator just needs to prove the presence of files (photos, videos, website consultation history) on the mobile phone. In this case, the logical extraction is sufficient when the investigator finds proof during the logical analysis. If he does not succeed in finding a visible file, then he must use physical extraction to see if deleted files are present. Moreover, we notice here the importance of a communication channel with the file system. Thus, when this communication is not possible, like damaged mobile phones (memory, CPU, PCB or output ports), the logical extraction is not possible.

3.2 Physical Extraction

Physical extraction is a very low-level copy of all the binary data physically present on the mobile phone's silicon flash memory. As discussed in chapter 2, transistors use

elementary cells to store binary information. Physical extraction is an electronic reading of the state of all the elementary memory cells. Thus, unlike a logical extraction, the physical extraction allows the investigator to collect all the information still physically present in the flash memory. It is precisely for this reason that this extraction mode is suited to recover the user's deleted data. Indeed, this deleted data is still physically present in the flash memory unallocated space, but was deleted by the system file. This is the crucial difference with logical extraction that is focused on the allocated space.

As long as the system does not decide to rewrite some data to the location of this deleted file (section 3.2.1), the data is always physically present and recoverable during physical extraction. Data carving can be applied directly to the physical dump. Therefore, to search for a deleted file, the investigator needs to create a script, in python or other programming language, to search all headers and footers corresponding to the file he wants to retrieve (**Appendix B**). To find a JPEG picture by carving it, he just needs to look for the header 0xffd8 and the footer 0xffd9 (Figure 3.6).

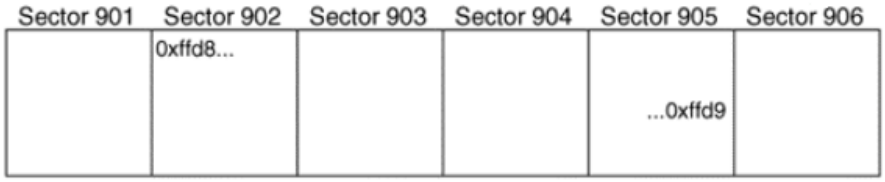


Figure 3.6: Blocks of Raw Data That Can Be Carved to Find a JPEG Picture in Sectors 902 to 905 [Brian, 2005]

Physical extraction is suitable to find SMS, contacts, call logs, application data, files, hidden files, deleted data, and to recover the lock code. To analyse a mobile phone in good working order [Van Der Knijff, 2002], the forensic investigators first create a physical forensic memory dump and then extract data.

In most cases, mobile phones are connected to the private companies' devices (Cellebrite, Micro Systemation, etc.) via a USB cable connection [Hoog, 2011], which communicates with the phone to physically extract its data [Ayers et al., 2007] bit by bit. For this reason, forensic investigators often use devices (software suites and hardware) specially developed by private companies, like Cellebrite UFED[5], Micro Systemation XRY[6], and Oxygen Forensic Suite[7], etc. In the case where manufactured equipment does not support the mobile phone, forensic investigators may:

- Develop their own programs or communication protocols (requires development, extremely time-consuming process).

[5] http://www.cellebrite.com/Mobile-Forensics/Applications/ufed-physical-analyzer
[6] https://www.msab.com/products/xry/xry-physical/
[7] https://www.oxygen-forensic.com/en/

- Physically extract data using flasher-boxes/boot-loader/JTAG (level 3).
- Physically extract data using chip-off/lapping and memory reading (level 4).
- Physically extract data using chip-on/micro-probing and memory reading (level 4).
- Physically extract data using micro-reading (level 5).

Before going into the details of the different types of physical extraction, it is necessary to understand the read, write and erasing mechanism operations present in flash memories. Understanding these mechanisms allows the investigator to select the most appropriate extraction method based on the type and overall condition of the phone "sealed" as evidence.

3.2.1 Flash Memory Management Mechanisms

From an elementary point of view (Figure 3.7), NAND flash is organised in blocks. The block is the smallest erase unit. Each block is itself divided into pages which is the smallest unit of writing and reading.

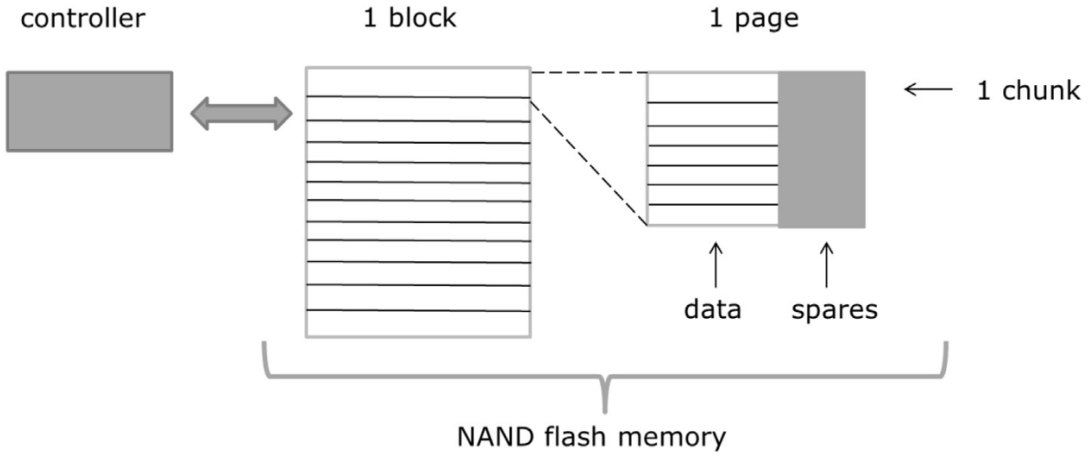


Figure 3.7: Elementary Organisation of a Flash Memory

During the reading or writing on the transistors (section 2.2.1), some bits may be physically erroneous. To counterbalance this, the memories integrate a verification/check algorithm that produces a sort of ECC value for each accessed page (Figure 3.8): the value is then stored in the spare area.

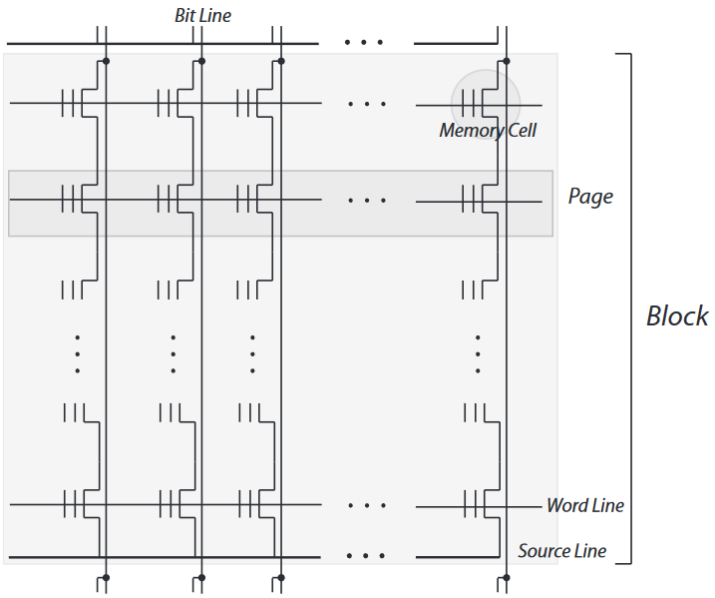


Figure 3.8: Transistor Organisation of a Flash Memory

This is actually a data control space that is stored on a page extension. The paper of [Fiorillo, 2009] describes the composition of this space. This control data allows the management of errors and their correction. This system of identification, control, and correction of errors makes for the strength of management mechanisms on flash memories. This system makes flash memories a robust technology with the ability to identify and counterbalance transistors' errors in read and write modes.

Flash memories use another algorithm called wear leveling (Figure 3.1), enabling the controller to erase as little as possible from each block, writing the new data in priority over the empty blocks that are least used [Regnery and Souvignet, 2013]. Thus, wear leveling is a technique for prolonging the memories' period of use and there are several wear leveling mechanisms that offer different levels of longevity improvement in such flash memory.

The wear leveling is based on the number of erases of each block with an incremented counter mechanism. In the example presented in Table 3.1, five blocks are filled with data. Blocks A and B that we want to delete are actually marked as invalid by the system.

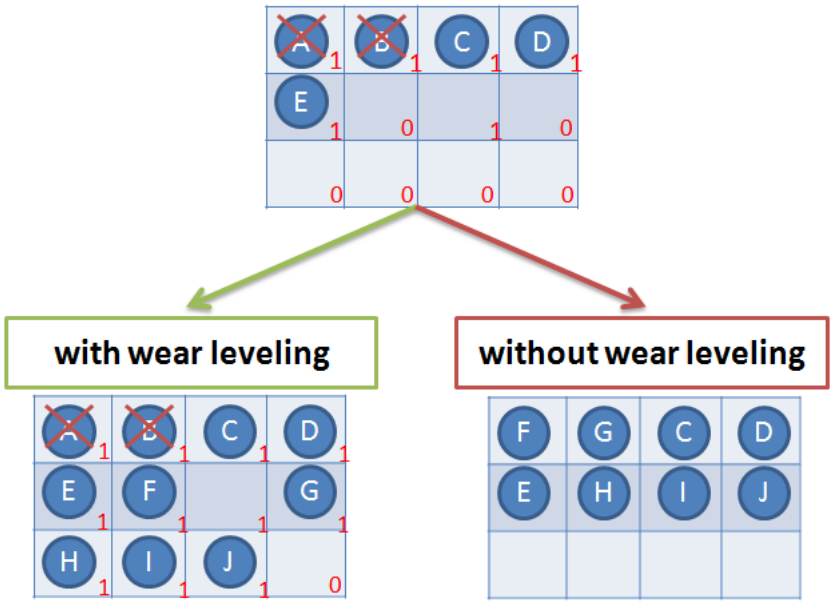


Table 3.1: Wear Leveling in Flash Memories

The blocks are not systematically released in order to preserve the life of the flash memories. Thus, without wear leveling, the new pages (F and G) are written in the first available blocks, in place of the old pages (A and B). But with wear leveling, the first two blocks are marked as having been used once already. Therefore, the new pages are written in the first blocks that have the lowest usage counter.

In the presence of wear leveling, when a block needs to be modified, the controller copies it with the data modified elsewhere in the memory and connects this new physical address to the logical address; the previous block is marked as invalid.

The physical deletion of a block, which is marked as invalid, does not systematically preserve the flash memory's life. The data changes on phones is constant for every use and manufactures have therefore introduced new mathematical mechanisms to increase the life of flash memories. Without this kind of preservation mechanism, the lifetime of memories would be much shorter and would lose all advantage in their use in mobile phones, GPS, and computers.

Thus, in addition to the mechanism of wear leveling in flash memories technology, a mechanism called **garbage collector** is responsible for managing the free space (Table 3.2, [Regnery and Souvignet, 2013]). Its role is to be able to determine which space can no longer be used by the program and then recover the space used.

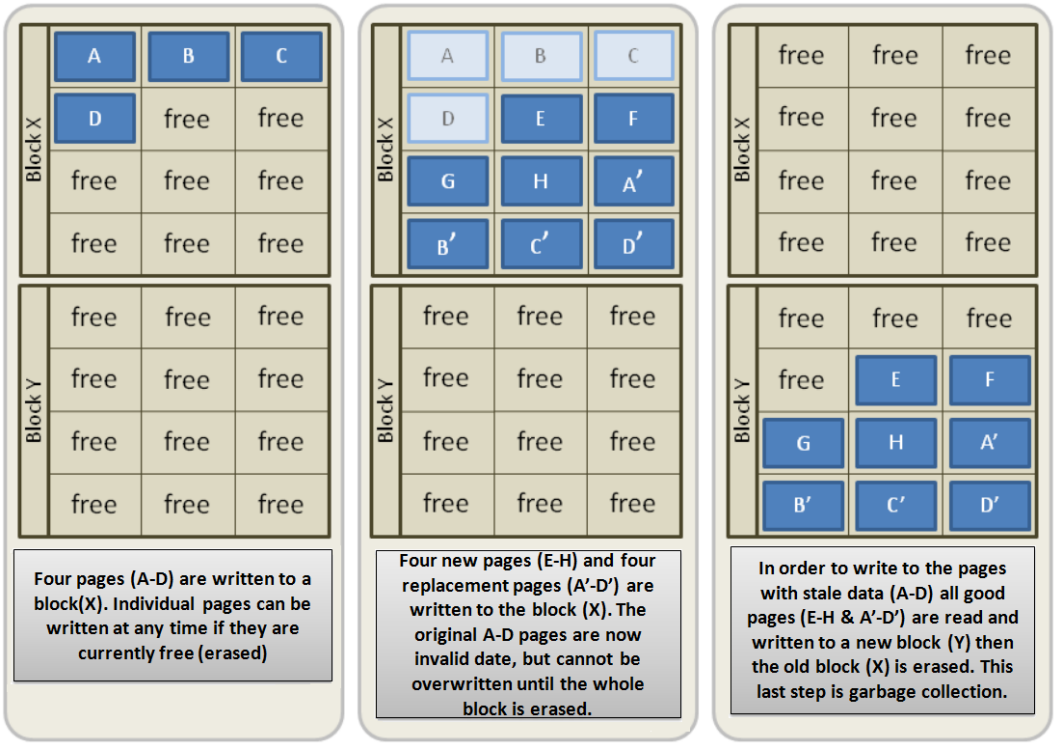


Table 3.2: Garbage Collector in Flash Memories

If a block contains pages which are marked as invalid, the algorithm moves the valid pages to another block, and then clears the first one. It uses an accessibility criterion to determine if an object can potentially be used. Thus, an accessible object can be obtained by following a chain of pointers or references. This mechanism is very complex and must find a fair balance so the block is not erased as soon as a page is invalid. Previously, we have seen that to delete data physically, the released blocks must be marked as invalid, so that the garbage collector actually performs the erase operation.

But when the user deletes data via its file system, this mechanism is not automatically triggered because only the entries in the allocation tables are modified. This is the substantial difference between a logical extraction and a physical extraction. The investigator is able to recover data still physically present on the silicon chip. When the flash translation layer is under the care of the host file system, the logic is external to the NAND, and the flash is said to be a raw NAND (Figure 3.9). On the other hand, when the flash translation layer (FTL) logic and management mechanisms (Error Correction Code (ECC), Garbage Collector, and Wear Leveling) are embedded, then the flash is said to be a managed NAND.

We show in chapter 5 a particular type of managed NAND memory whose management mechanism is managed by an internal controller (eMMC). This kind of memory is found in most modern mobile phones. We cannot be exhaustive, but the main ones can be

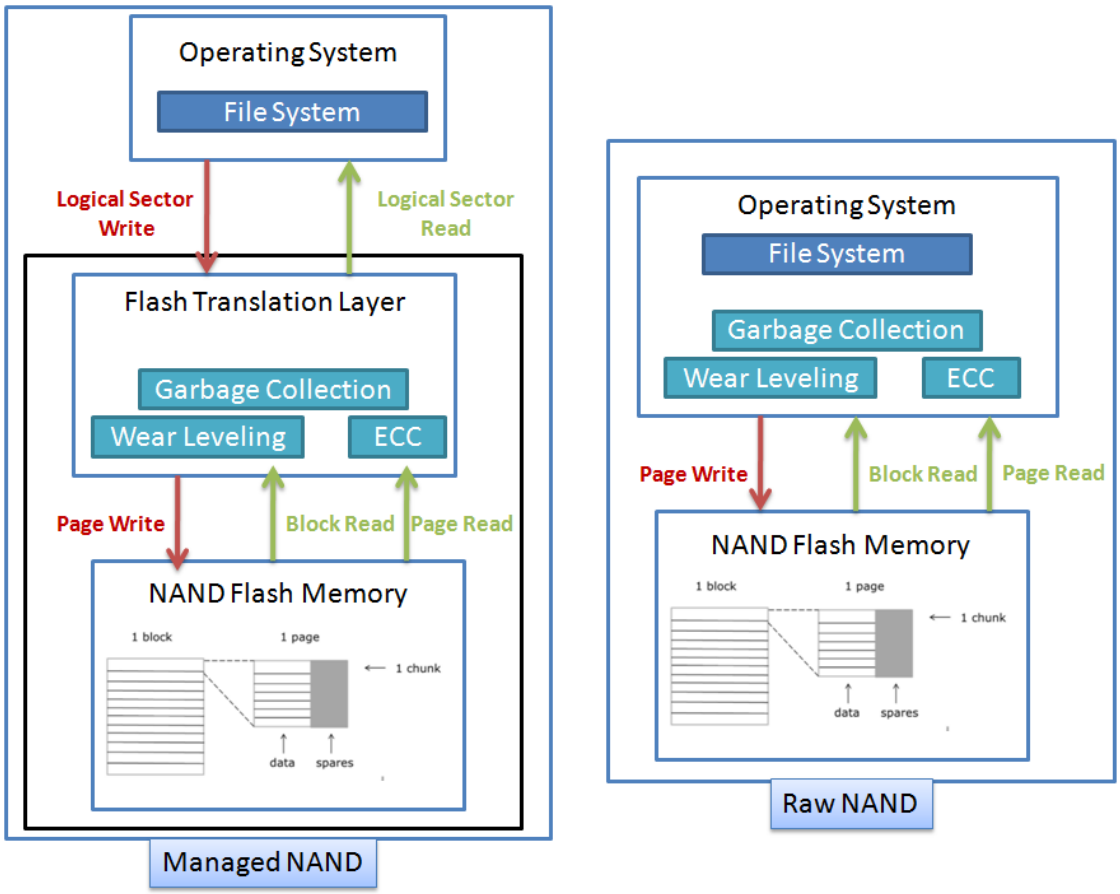


Figure 3.9: Managed NAND vs Raw NAND

found in the following models: BlackBerry (Q10, Z10, Z30, 9900, 9790, 9720), iPhone (5/5c/5s/6/6Plus/6s/6sPlus), Samsung (Galaxy S5 SM-G900f, Core prime), LG Nexus 5, Sony Xperia, Motorola GXT1032, HTC (Desire C, One SV, Windows Phone 8S), etc.

3.2.2 Physical Extraction Using Flasher-Boxes/Boot-Loader/JTAG

Physical extraction by flasher boxes, boot loader and JTAG is an area that is of great interest to private companies because it constitutes a real technical challenge that allows highlighting the companies' technical capabilities.

3.2.2.1 Physical Extraction Using Mobile Phone Flasher Boxes

Mobile phone manufacturers first used flasher boxes for software updates and diagnostics. But they can also be illegally used, for example to change the IMEI number of a mobile phone devices.

In this thesis, we use them to acquire a full physical copy of the mobile phone's memory [Breeuwsma et al., 2007, Jonkers, 2010]. Therefore, they can be used to update the internal software that is stored in the phone's memory (firmware). The flasher box is mostly connected, on the one hand, to a PC via a USB connection, and on the other hand, to the mobile phone via a special cable specially designed for the mobile phone model.

The cables are generally pins that contact the mobile phone's service ports through the Joint Test Action Group (JTAG) connection, the Mbus/Fbus connections or Ethernet RJ45.

The flasher box will, therefore, put the phone's operating system in a special diagnostic mode. Then it sends a request to the mobile phone memory address. And finally, the mobile phone responds by giving the binary data, called the physical memory dump (Figure 3.10).

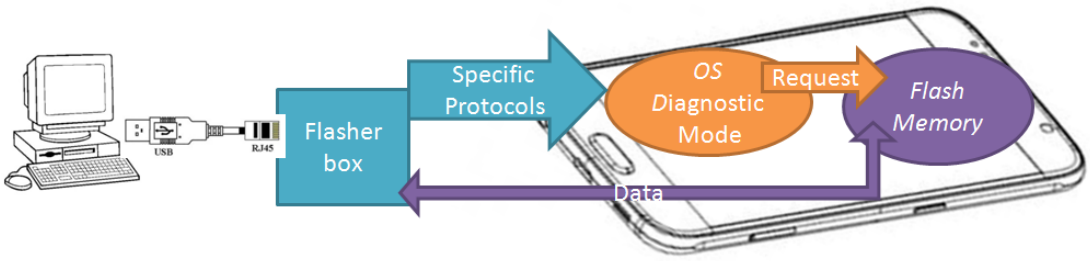


Figure 3.10: Flasher Box Communication

The main problem with flasher boxes is that the forensic investigator has no guarantee of the good forensic conditions of use (change of memory data, data written to the phone's memory chip, etc.)

3.2.2.2 Physical Extraction Using a Boot Loader

Physical acquisition techniques using a boot loader must execute a code on the target mobile phone [Guido et al., 2016]. The process consists of pushing a boot loader into the mobile phone, then dump its memory and finally analysing the physical memory dump.

With commercial toolkits, data is typically sent through a USB interface that is connected to the investigator's computer. Unfortunately, this uploaded program can be written in the memory where data are stored. As a result, the data may change during the download process.

Manufacturers promote the forensic use of their equipment. However, the investigator should be suspicious and always keep in mind that modifying the data in the phone's memory is a possible step when it injects the boot loader.

3.2.2.3 Physical Extraction Using Joint Test Action Group (JTAG)

The JTAG protocol and interfaces are normally used to test and debug embedded systems. They can also be used to access flash memory [Breeuwsma, 2006]. When manufacturing and assembling components on the PCB, the manufacturers use the JTAG communication to verify the correct process at each assembly step of the functioning of electronic components and soldering quality.

Investigators use the JTAG to communicate directly with the processor and get an entire physical image of the flash memory. The IEEE 1149.1 standard defines the JTAG protocol: "Standard Test Access Port and Boundary-Scan Architecture"[Maunder, 1993].

Most JTAG enabled processors to offer a debug mode, and the forensic investigator can use specialised flashing tools (RIFF Box[8], Octoplus[9], Medusa Box[10], etc.) designed for phone repair. Such flashing tools use the JTAG interface for their work and send instructions to the flash memory to perform a physical dump of all the data present on the flash.

The JTAG specification 1149.1 requires five signals: TCK for Test Clock, TMS for Test Mode Select, TDI for Test Data-In, TDO for Test Data-Out and an optional TRST for Test Reset.

These four signals are collectively known as the Test Access Port (TAP). Therefore, the IEEE standard defines a 16-state machine (Table 3.3) called the TAP controller that controls several actions.

[8] http://www.riffbox.org/
[9] http://octoplusbox.com/
[10] http://medusabox.com/

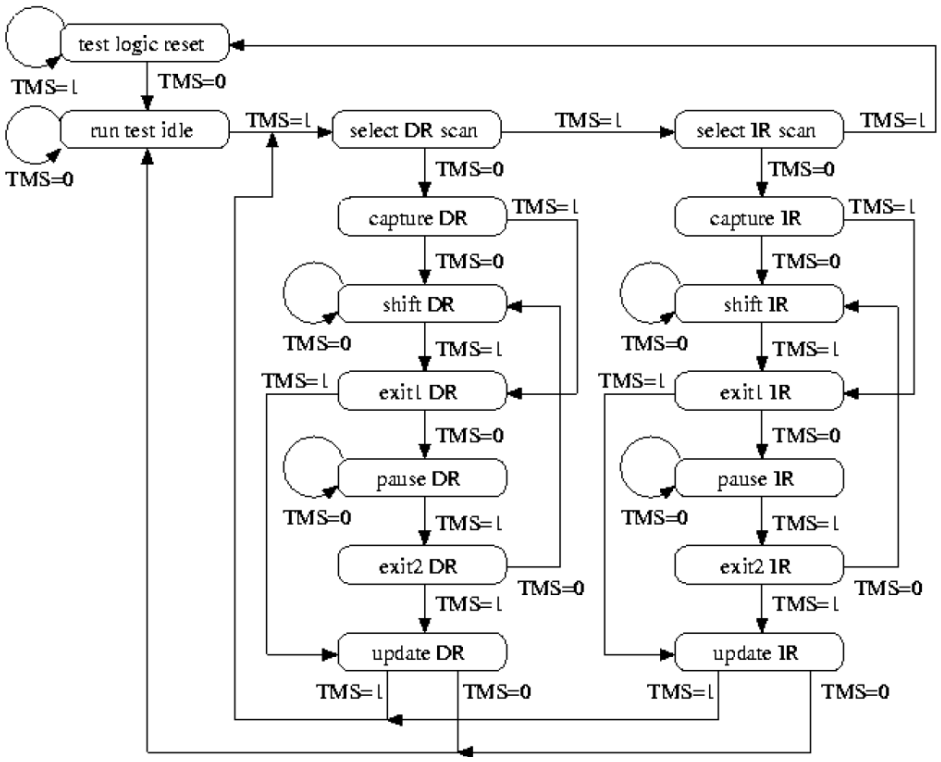


Table 3.3: 16-State Machine: JTAG IEEE [Maunder, 1993]

The limitation of this technique is that the TAPs of the JTAG are often deactivated at the end of the assembly process, directly by the manufacturer and investigators are not able to reconnect them. The JTAG then becomes an unusable method, and the investigator must then desolder the flash memory.

3.2.3 Physical Extraction Using Chip-off Methods: Unsoldering/Lapping and Memory Reading

The chip-off can refer to two complementary techniques:

- The first consists of extracting the components by unsoldering and then a physical reading.
- The second involves the extraction of the components by the lapping technique and then their physical reading.

3.2.3.1 Unsoldering Technique

If extraction by the JTAG interface is not possible, the Breeuwsma flash-desoldering and reading technique must be used [Breeuwsma et al., 2007]. The investigator uses a

reworking station to perform this technical operation (Figure 3.11).

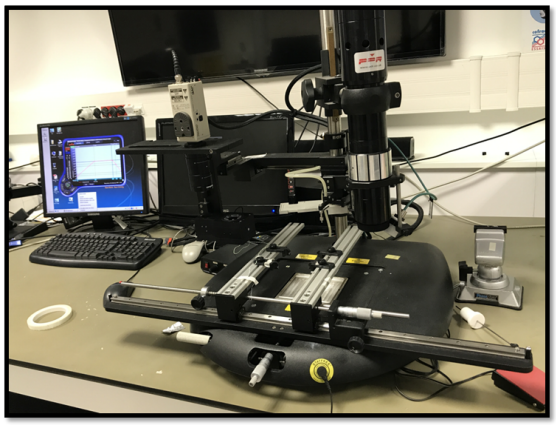


Figure 3.11: IRCGN's Reworking Station

An infrared beam or a hot air gun heats the top of the component to be desoldered. The temperature is controlled at the same time by an infrared thermometer. Simultaneously, the PCB is also heated by a joule effect grid. The temperature of the PCB is followed in real time by another thermocouple (Figure 3.12).

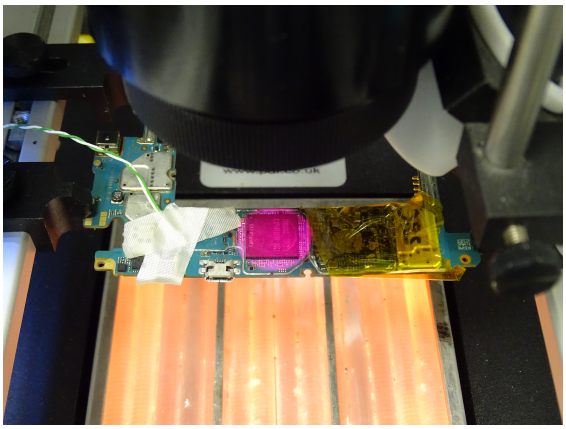


Figure 3.12: Memory Being Heated Before Being Unsoldered

This system of double heating makes it possible to minimise the too-sudden temperature gradient, which would result in fracturing the electronic component and thus destroying it. Thus, the infrared beam makes it possible to add a small temperature gradient that is sufficient to bring the solder balls to the liquefaction temperature. A vacuum pump then creates a uniform mechanical motion on the component, which terminates the desoldering process. As we will see in chapter 4, the curves of the rise in temperature are not a random process and must be optimised. Thus the entire desoldering process is computer-assisted and follows a process well defined by the investigator.

Unfortunately, the desoldering of the memory component is often a delicate step because traditional components use high-temperature soldering balls and underfill mate-

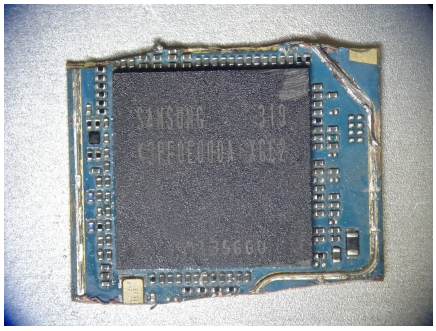
rials. This applies a significant thermal stress to the electronic component. Thus, in the latest generation of mobile phones, this technique quickly reached its limits. This thermal stress can lead to data changes, deletion, or even the destruction of the mobile phone memory. The liquefaction temperatures of the soldering depend to a large extent on the alloys used to solder the component. Thus, as we will see in chapter 4, bismuth-based alloys have a liquefaction temperature of 138 °C, compared to 212 °C for 91.84Sn–3.33Ag–4.83Bi alloy, 205 °C for 86.4Sn–5.1Ag–8.5Au alloy and 183 °C for 63Sn–37Pb alloy [Vianco et al., 2004].

In the new generation of mobile phone or GPS, the manufacturers add, in addition to the soldering, a thermal glue which further increases the desoldering temperature to 300 °C for the low-temperature underfilling and to 400 °C for the high-temperature underfilling. We will study in more detail in chapters 5 and 7 this type of thermal adhesive and develop innovative solutions. Thus, the desoldering technique is limited when the memory is bonded with thermal adhesive. Indeed, the memory component would have to be brought to 400 °C in order to be desoldered, which is a risk that the investigators must be aware of.

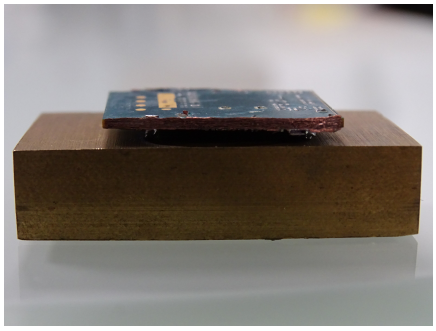
3.2.3.2 Lapping Technique

The techniques using the reworking station cannot be used for components that do not withstand the temperature rise. This is generally the case for components already weakened, such as after an air crash or an explosion (terrorist attack). Lapping techniques are therefore a complementary technique to desoldering, but take longer to implement.

The lapping technique (Figure 3.13) is a technique in which the PCB is first cut mechanically around the component to be extracted (Figure 3.13a). Next, it is glued with an adhesive onto a perfectly flat metallic support (Figure 3.13b). The different layers constituting the PCB are then sanded manually (Figure 3.13c) and removed one by one (Figures 3.13d to 3.13g). Finally, once the last layer is removed, the component is cleaned using a soldering iron and flux remover (Figure 3.13h).

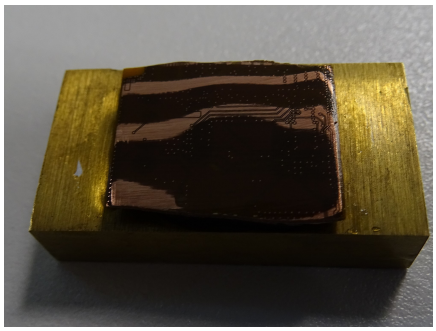


(a) BlackBerry PCB cutting

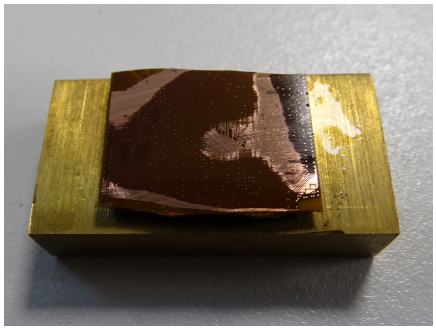


(b) CPU before lapping

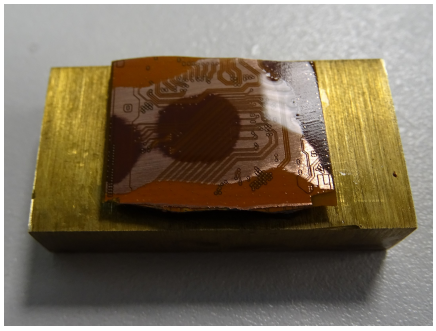
(c) Manual lapping



(d) 3 minutes' lapping

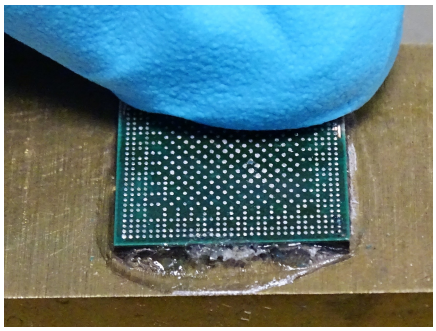


(e) 6 minutes' lapping



(f) 9 minutes' lapping

(g) 12 minutes' lapping



(h) CPU cleaning after 15 minutes

Figure 3.13: Lapping Process Steps

Unfortunately, with some encrypted mobile phones, the investigator has to re-solder the components (for example after password extraction and brute force attacks) and to turn on the phone to decrypt the memory by the processor's hardware key.

Thus, he must use the processor encryption key to decrypt the memory's data by soldering memory and processors on a donor board (Section 3.4). Unfortunately, on the new mobile phones, these two elements generally face each other, which prevents the use of the lapping technique because it would require destroying one or the other components (Figure 3.14).

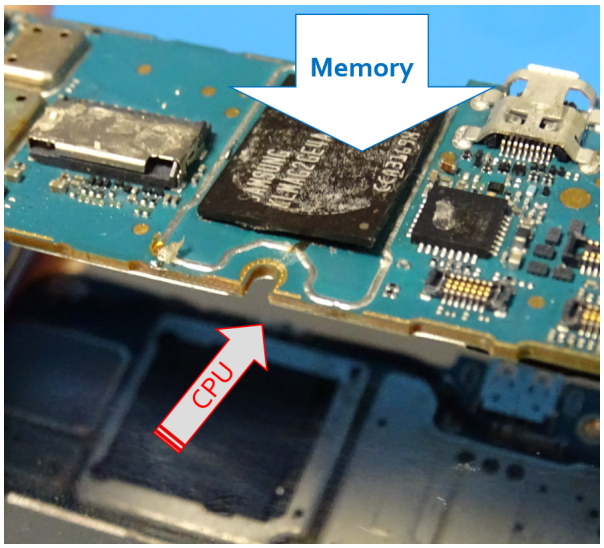


Figure 3.14: BlackBerry Z10: Lapping Method Unusable

This principle will be presented in more details in chapter 5 and the techniques developed during this thesis to solve this problem will be explained.

3.2.3.3 Memory Reader

We have just presented two memory extraction techniques: desoldering with a reworking station and lapping. Once the memory component is removed, the investigator uses a specialised component reader to extract the data physically (Figure 3.15), or a forensic reader (Memory Toolkit MTK II). For this forensic operation to be carried out under these conditions, the investigator uses a forensic bridge that is located between the analysis computer and the electronic component reader. Thus, the analysis computer leaves no trace on the memory to be analysed which guarantees data integrity on a phone which is "sealed" for investigation.

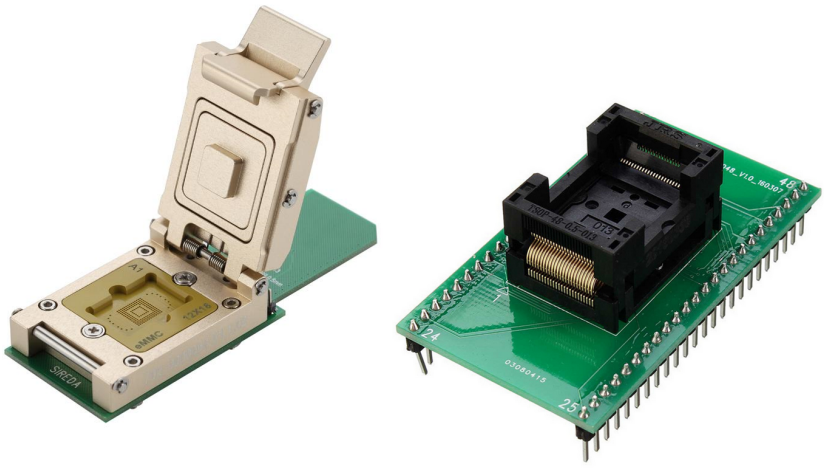


Figure 3.15: eMMC and TSOP Memory Adapters

In some cases, investigators do not have readers that can directly read the proprietary memory communication protocol. Nonetheless, they can use a homemade reading mode "wire-to-wire", in which they instantly create wire junctions on the BGA balls of the communication protocol (Figure 3.16).

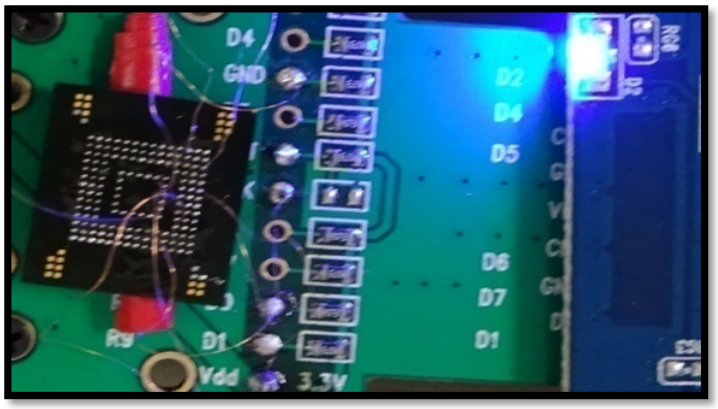


Figure 3.16: Wire-to-Wire Method on eMMC Controller Via Secure Digital Protocol

After making the wire-to-wire connection, we used in this thesis the memory's specific protocol to retrieve binary data. If we take the example of memory using an internal eMMC controller, then we can read the 8 bits of the data bus by sending reading instructions via the Secure Digital protocol (Figure 3.17). We will study this type of communication protocol in more detail in chapter 5.

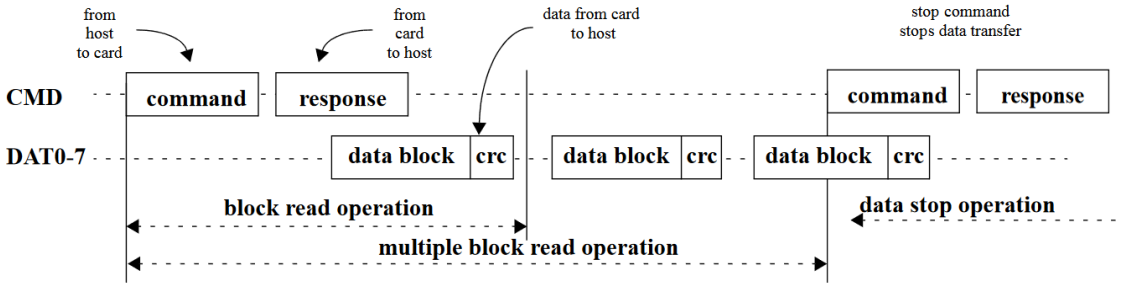


Figure 3.17: eMMC 8 Bits Read Operation Using Wire to Wire Method [JEDEC, 2010]

The forensic investigators can also use an FPGA to read the memory in the proprietary communication protocol. The FPGA therefore constitutes a universal component reader programmable by the investigators (Figure 3.18).

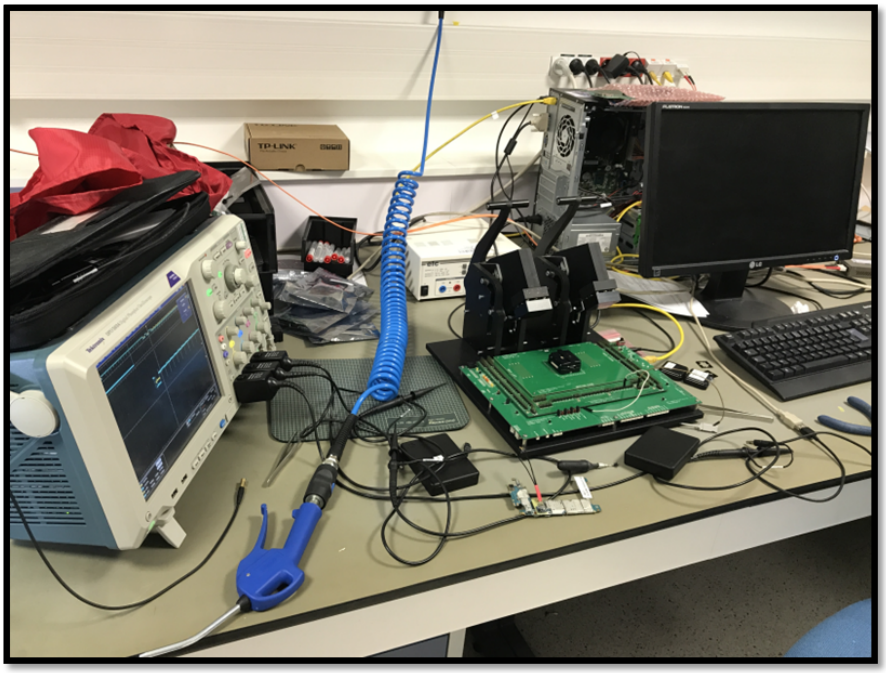


Figure 3.18: IRCGN's FPGA Reader: NFI Memory Toolkit II

This technique is a method for coupling a hardware part and a software part. The first is the junction between the electric microwires and the communication BGA balls or using the proprietary connections. The second one constitutes the programming part of the FPGA (Verilog, VHDL) using the memory's communication protocols (Figure 3.19[11]).

[11] https://www.forensischinstituut.nl/

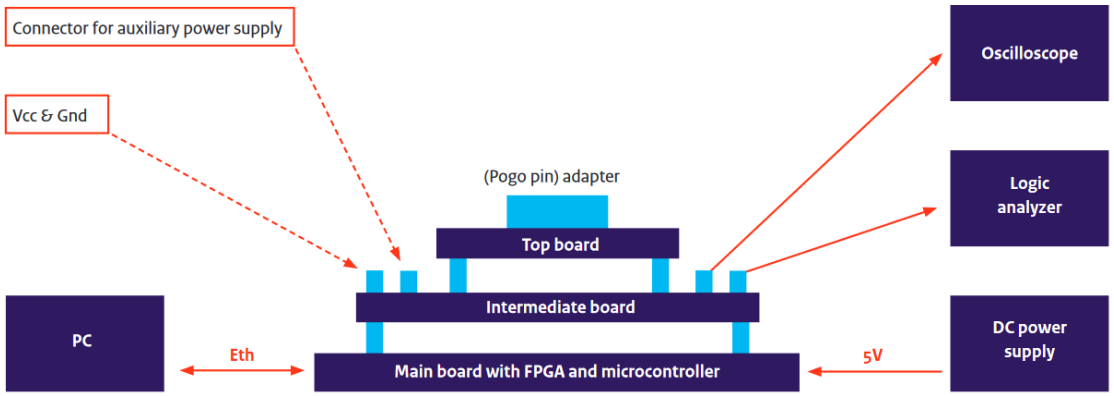


Figure 3.19: Schematic of the NFI Memory Toolkit II

In chapter 5 we will add a platform for man-in-the-middle attacks using the oscilloscope and logic analyser for spying.

3.2.4 Physical Extraction Using Chip-On: Acid/Laser Attack and Memory Reading by Micro-Probing

3.2.4.1 Chip-On Method: Acid Attack and Memory Reading by Micro-Probing

Just like laser–matter interactions, which will be presented in the next section, acid attacks for forensic applications aim at removing the encapsulation box from the chip to reveal the silicon chip's memory (Figure 3.20, [Kerisit et al., 2014b]).

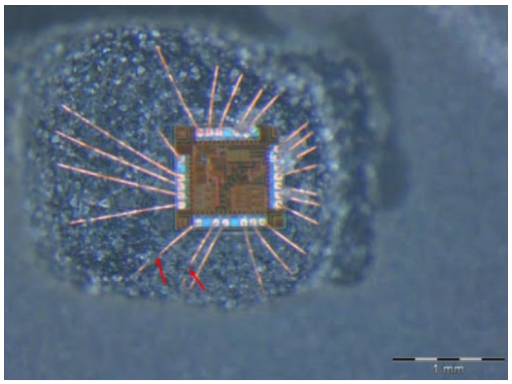


Figure 3.20: Silicon Chip After Acid Attack

The chip's insulating box is composed primarily of epoxy resin and other elements, which are mainly glass beads, alumina beads (Al_2O_3), boron nitride (BN), or aluminium nitride (AlN). The acids that we used in this thesis had to remove the insulating box without damaging the rest of the chip (the silicon chip's memory, bonding wires, etc.).

The different acid mixtures allowed us to be selective about the elements to be removed. Thus, we see in section 3.2.5 that some specific acid mixtures permit us to carry out nano-lapping on the silicon's elements. Thus, via a Scanning Electron Microscope (SEM), the acids make it possible to read the primary states of the transistors of the memory.

For now, we are interested in acid mixtures that eliminate the insulation box without damaging the rest of the chip. In this thesis, we open the insulating boxes with a mixture of fuming nitric acid with a 100% concentration (3 volumes) and sulphuric acid with a 100% concentration (1 volume). The whole is heated to 90 °C with an attack time that depends on the amount of resin to be removed (Figure 3.21).

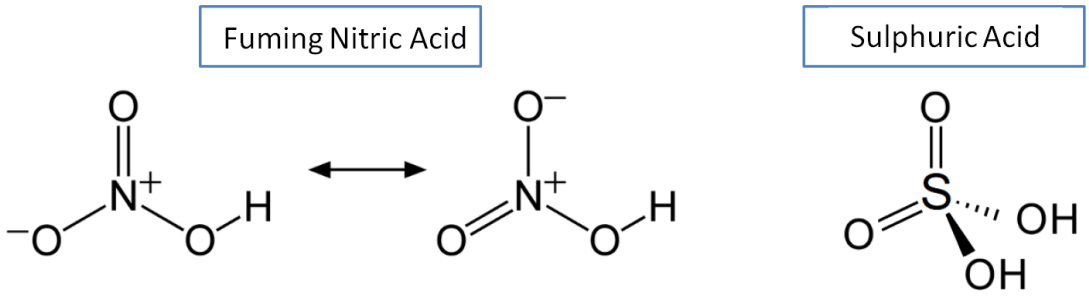


Figure 3.21: Nitric Acid and Sulphuric Acid Molecular Patterns

For an eMMC packaging, an attack of 3 times 1 minute is carried out on the insulation box. A study of the fundamental chemical reactions shows that fuming nitric acid corrodes copper by:

$$4HNO_3 + Cu \longrightarrow Cu(NO_3)_2 + 2H_2O + 2NO_2$$

and sulphuric acid attacks aluminum by:

$$3H_2SO_4 + 2Al \longrightarrow Al_2(SO_4) + 2H_2$$

Nitric acid reacts differently from sulphuric acid with metals because of the oxidising properties of the NO_3 radical. Thus, HNO_3 reacting with a metal will never give rise to hydrogen H_2. Conversely, nitric acid creates a passivation layer on aluminium:

$$HNO_3 + Al \longrightarrow Al(NO_3)_3 + H_2O + NH_4NO_3$$

This equation shows the creation of an aluminium nitrate protective layer. In the same way, sulphuric acid protects copper with a layer of copper sulphate:

$$Cu + 2H_2SO_4 \longrightarrow CuSO_4 + SO_2 + 2H_2O$$

By mixing the two acids, it is then possible to take advantage of both by removing the resin while protecting the copper and aluminium which are present in the components (Figure 3.22).

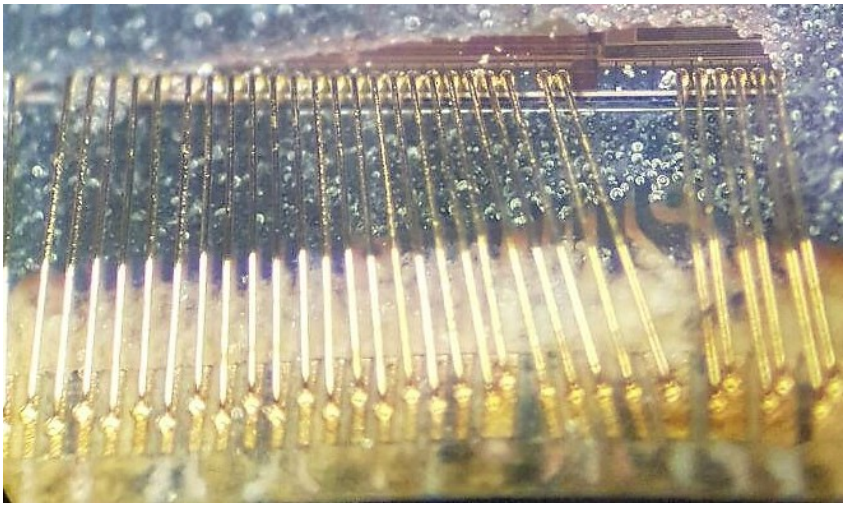


Figure 3.22: Memory Chip After Acid Attack

3.2.4.2 Chip-On Method: Ablation Laser and Memory Reading by Micro-Probing

We start this section by recalling the theory of the spatio-temporal laser radiation ablation, and then discuss the fundamental equations characterising laser–matter interactions.

Laser–Matter Interaction

We now present the theoretical behaviour of the laser–matter interaction and show how the energy deposited on the material is dependent on the target material.

Following the work of Albert Einstein and Louis de Broglie, scientific theories gave to all objects a double nature of wave and corpuscle: light is both an undulating phenomenon represented by an electromagnetic wave with a wavelength, and a corpuscular phenomenon, as evidenced by the elementary particles of photons of quantum physics [Ngô, 2005].

The ablation of materials by photons in the 1064 nm infrared domain is a thermal process in which absorption of the photon by the target material results in temperature rise and heat diffusion (Figure 3.23).

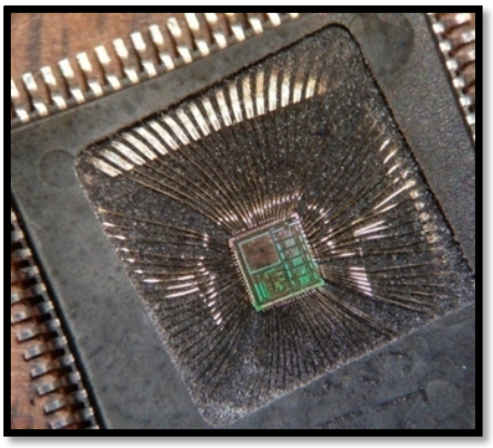


Figure 3.23: Silicon Memory Chip After Laser Attack

Some of the material is ejected as vapour or fragments of matter. After the laser pulse, the material begins to cool and the material recombines. The spatial and temporal evolution of temperature is described by the equation of heat (Equation 3.1).

$$\rho(T)C_p(T)\frac{\partial T(x,y,z,t)}{\partial t} - \nabla(\kappa(T)\nabla(T(x,y,z,t)) = Q(x,y,z,t) \tag{3.1}$$

ρ is the material's density, C_p is the specific heat capacity of the material, κ is the thermal conductivity. T and Q change as a function of the time t during laser irradiation, with the coefficient of the heat source (Equation 3.2).

$$Q(x,y,z,t) = (1-R)I(r,t)M(z) \tag{3.2}$$

R is the reflection coefficient at the air/material interface, $I(r,t)$ is the spatial and temporal profile of the laser pulse propagating along the z axis (Equation 3.4). $M(z)$ is the normalised energy deposition such that (Equation 3.3):

$$\int_0^\infty M(z)\mathrm{d}z = 1 \tag{3.3}$$

With $w(z)$ the radius of the laser beam along the z axis and P the laser's power, the spatial intensity of the beam (Equation 3.4, Figures 3.24[12] and 3.25) is equal to:

$$I(r,z) = \frac{P}{\frac{\pi w(z)^2}{2}} e^{\left(\frac{-2.r^2}{w(z)^2}\right)} \tag{3.4}$$

[12] http://www-lpl.univ-paris13.fr/pon/lumen/documents/Optique%20des%20lasers_et%20faisceaux%20gaussiens.pdf

[13] http://vlab.amrita.edu/?sub=1&brch=189&sim=342&cnt=1

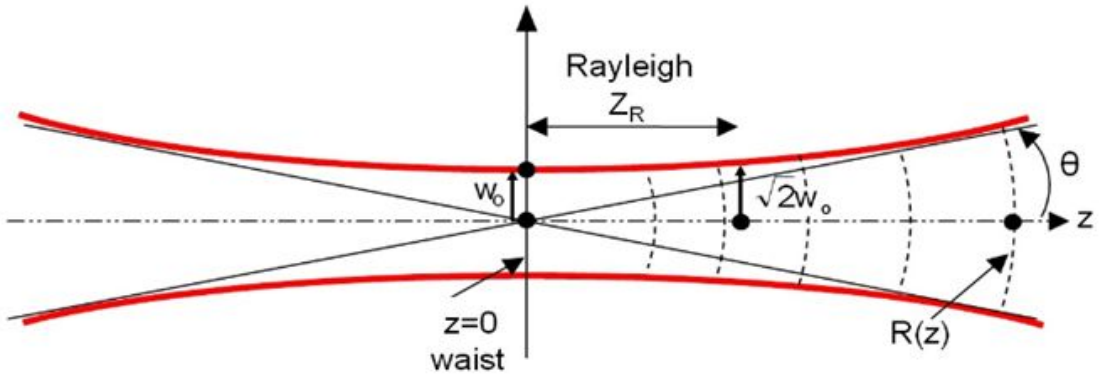


Figure 3.24: The Spatial Intensity of the Beam Along the Axis z of Propagation, With w_0 the Minimum Radius of the Laser Beam, at the Waist

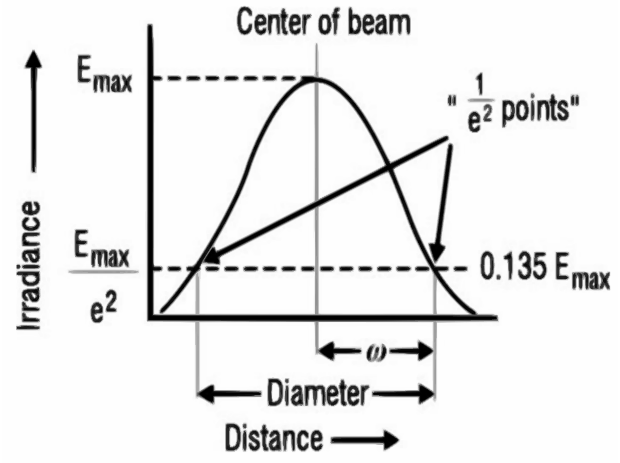


Figure 3.25: The Spatial Intensity of the Beam Along the Plane (x, y) Perpendicular to the Axis of Propagation[13]

In the case of a deposition of energy by a laser beam which follows the Beer–Lambert law we obtain (Equation 3.5),

$$Q(x, y, z, t) = (1 - R)I(x, y, t)\alpha e^{-\alpha z} \tag{3.5}$$

With α the absorption coefficient of the material. The absorption coefficient (Equation 3.6) is defined by the ratio between the absorbance A and the optical path L traveled by electromagnetic radiation in a given medium,

$$\alpha = \frac{A}{L} \tag{3.6}$$

According to the Beer–Lambert law, the absorbance is additive (Equation 3.7). Thus, for a solution containing several absorbing substances, the overall absorbance is the sum of their absorbances. The Equation 3.7 is the key point of the technique developed in

chapter 6. For n absorbent substances:

$$A = \sum_{i=1}^{n} A_i(\varepsilon_\lambda, i, l = 1\text{cm}, c_i) \tag{3.7}$$

Laser IC Decapsulation

Theoretically we find that the maximum intensity is reached at the waist and at the centre of the laser spot as shown in Figure 3.24 and Figure 3.25. Note that the theoretical power given by the constructor, $\text{P}_{\text{constructor}}$, corresponds to the intensity (I) integrated on the surface of the laser spot at the waist (S_{waist}).

$$P_{constructor} = \int_0^\infty I.\text{d}S_{waist} \tag{3.8}$$

Some studies [Kor et al., 2014] have shown the advantages and disadvantages of laser de-focusing during decapsulation. The tests were performed by focusing at the local waist. The laser we used[14] in this thesis was a Nd:YAG laser, power 30 W adjustable ($\text{P}_{\text{constructor}}$= 30 W) that operates at the 1064 nm fundamental wavelength (Figures 3.26 and 3.27).

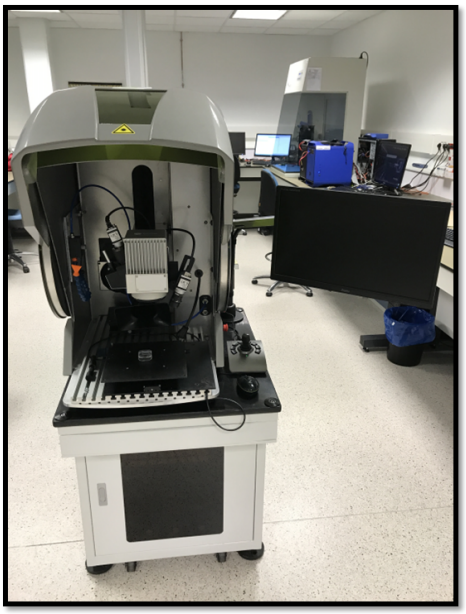


Figure 3.26: IRCGN's 1064 nm Laser

Note that there are other types of laser[15,16] that use the 532 nm second harmonic of the Nd:YAG laser. We will see in this thesis why the use of this type of laser is interesting. Indeed, depending on the application, these two types of lasers can be complementary.

[14] https://www.controllaser.com/laser-machines/laser-semiconductor-decapsulation/
[15] https://www.controllaser.com/lasers/laser-semiconductor-decapsulation/falit-duo/
[16] http://www.digit-concept.com/equipment/sesame-laser/

There are also suppliers who promote lasers which can emit on the first and second harmonics[17].

Figure 3.27: Ablation of Chip's Package by 1064 nm Laser

Note that the laser attack is a rather selective technique that will attack only the elements that absorb the wavelength of the laser. All other materials are insensitive. Since gold and silicon consume very little of this wavelength and the materials constituting the insulation layer absorb this wavelength a lot, the laser is an indispensable tool for investigators (Figure 3.28).

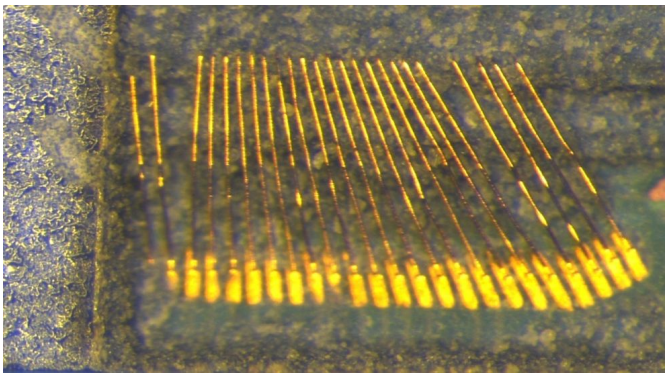


Figure 3.28: Memory Chip's Bonding Wires Access After 1064 nm Laser Decapsulation

As we will see in chapter 6, the theoretical study of laser–matter interaction is very significant in the development of an innovative method of advanced repair of an electronic part (memories, processors and crypto-components).

Micro-Probing

[17] https://www.rofin.com/en/products/lasers-for-marking/end-pumped-lasers/

Once the open window is made (by acid or by laser), we can directly read the information on the silicon. Micro-probing is used to create electrical access to a point on the silicon matrix or the connecting bonding wires. To establish this micrometric electrical contact, we use in this thesis a probing station (Figure 3.29):

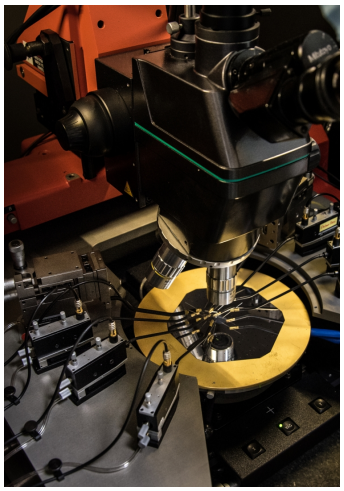


Figure 3.29: Probing Station

Electrical contact is made by dropping micrometric needle probes directly onto the silicon (the area on which the wire is soldered) (Figure 3.30) or on the bonding wires.

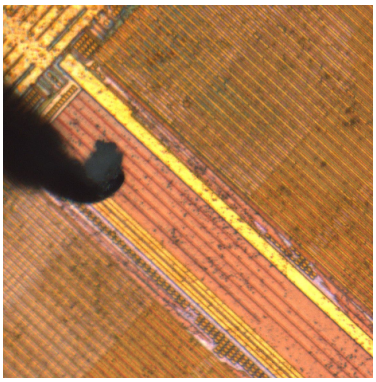


Figure 3.30: Needle Micro-Probe in Contact With Silicon Memory

In order for the needle probe to be precisely placed on the silicon matrix, each needle is held by a micro-manipulator controlled by the investigator. All operations are performed on an air cushion table to minimise vibration. The measurements of the voltage and the current are then carried out by the electrical measuring instruments connected to the needle probe by the micro-manipulator (oscilloscope, logic analyser, etc.), or to the investigators' homemade reader (Figure 3.31).

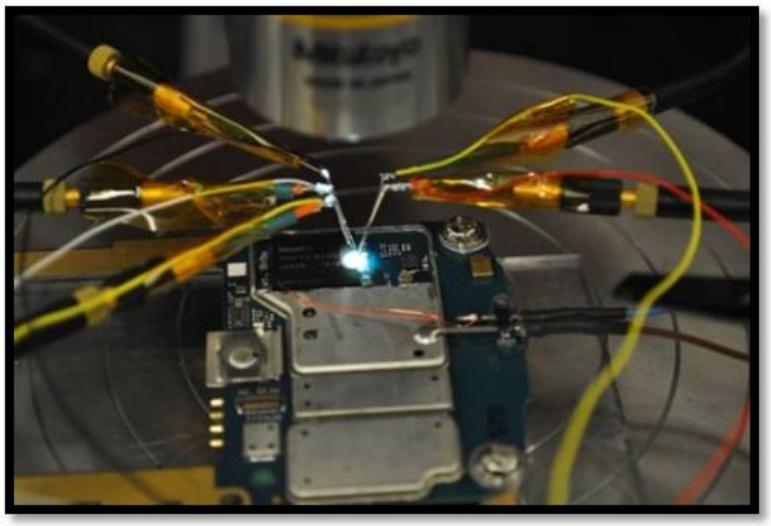


Figure 3.31: SD Bus Reading on eMMC's Silicon Via Probing

3.2.5 Physical Extraction Using Micro-Read

The main idea of this method lies in 3 steps. The first step is **Sample Preparation** that can either be carried out using a Focused Ion Beam (FIB), or using a polisher or acids. The second step is **Image Acquisition** by using a Scanning Electron Microscope (SEM) to view the memory's state. Finally, the last step is **Image Processing** which aims to transform the transistors' images obtained with the SEM into binary data 0 and 1.

3.2.5.1 Sample Preparation by Focused Ion Beam

The Focused Ion Beam (FIB) performs ion etching by attacking a very localised area of silicon (a few tens of micrometers) by means of a focused beam of ions. The ions, usually gallium, which constitute the beam, are accelerated to an energy of 1 to 150 keV and focused on the sample by electrostatic lenses (Figure 3.32).

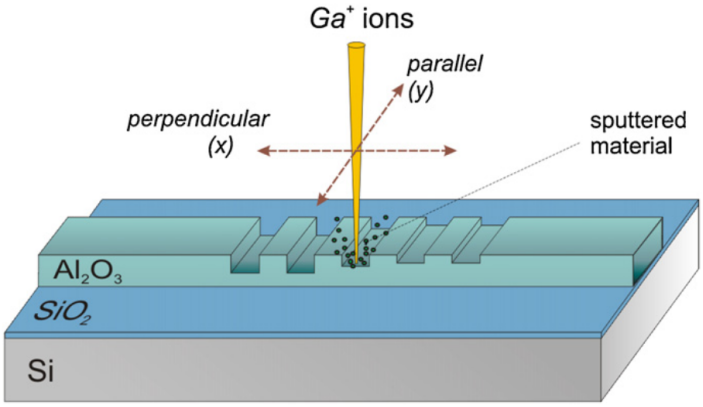


Figure 3.32: FIB–Matter Interactions [Ay et al., 2012]

When the energetic gallium ions reach the material, they eject atoms from the surface of the material, thus removing atomic layers. The FIB can also produce localised deposits of materials. By combining the etching and the deposition of material, it is then possible to carry out circuit modifications [Volinsky et al., 2004]. This modification is likely to modify an electronic structure by connecting or disconnecting specific tracks of the chip. This specificity allows the detaching of the protective mechanisms on the silicon, such as protective gratings.

3.2.5.2 Sample Preparation by Micro-Lapping and Reverse Engineering

To remove the different layers of the silicon chip for micro-reading [Brothers, 2009] and reverse engineering, investigators can also use the micro-lapping method. The paper from the University of Cambridge [Courbon et al., 2016] illustrates this principle and the steps followed here (Figure 3.33):

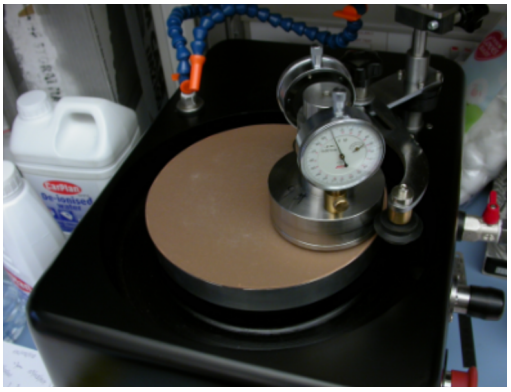


Figure 3.33: Cambridge Lapping Machine [Courbon et al., 2016]

3.2.5.3 Sample Preparation by Acid Attack

Finally, it is also possible to use different mixtures of acids. The chip is a juxtaposition of metals (Al, Cu, Au), dielectrics (SiO_2, Si_3N_4) and the silicon substrate. To remove the silicon oxide (dielectric), in this thesis we use hydrofluoric acid (HF) at 48% concentration or diluted in ammonium fluoride (NH_4F):

$$SiO_2 + 6HF \longrightarrow H_2(SiF_6) + 2H_2O$$

The acids needed to remove the metallisation layer depend on the metal. To remove the aluminium, we use a mixture of HCl + H_2SO_4 or a mixture of HNO_3 (to oxidise the aluminium) and phosphoric acid (to dissolve Al_2O_3). To remove gold, we use KI/I_2 because gold and iodine form gold iodide via:

$$2Au + I_2 \longrightarrow 2AuI$$

The solubility of AuI is improved by adding KI to the solution. Therefore, to remove the copper, we use fuming HNO_3:

$$4HNO_3 + Cu \longrightarrow Cu(NO_3)_2 + 2H_2O + 2NO_2$$

Finally, it may be advantageous to remove the silicon substrate to carry out a backside attack.

3.2.5.4 Micro-Read Process

After the Sample Preparation step, the Image Acquisition step is performed by a scanning electron microscope (Figure 3.34).

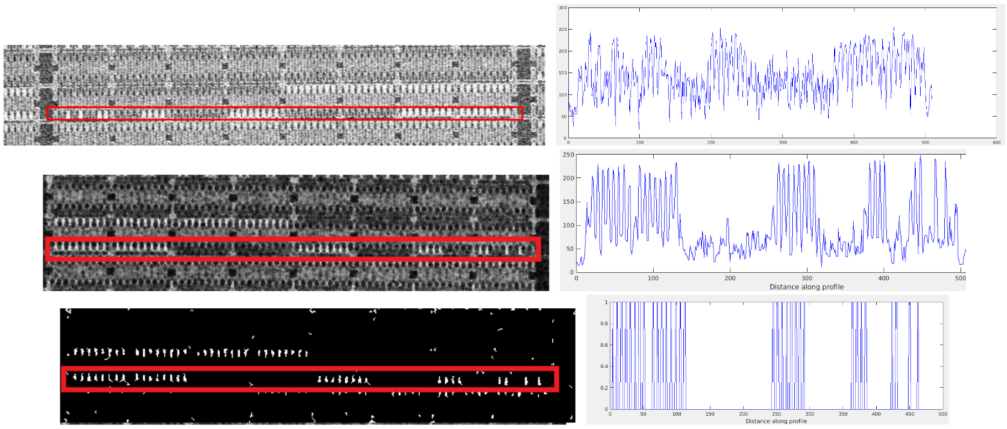


Figure 3.34: Cambridge 0s and 1s Memory Extraction Using Intensity Values Variation [Courbon et al., 2016]

The advantage of this device is its high precision and the ability to observe the memory states. Then, it is possible to read each transistor one by one (Image Processing) and use an algorithm to automate the reading process.

The layers of the memory chip are removed one by one with chemistry (or lapping or FIB), and then the transistors are read using the SEM images. The step continues until all memory cells are read or totally reverse engineered.

However, this technique is rarely used by investigators because it is destructive. It does not allow for any errors and the memory reading can only be performed once. Therefore, it does not allow for any counter-expertise.

This technique is very time-consuming and requires the use of FIB, SEM and complex chemistry techniques. Thus, currently, this technique is used only on very rare occasions and cannot be generalised to all judicial electronic investigations.

3.3 Traditional Forensic Techniques Used to Extract Data from Damaged Mobile Devices

The mobile phone can be damaged in several ways, and the identification of the level of damage is part of the expert's work. In this thesis, we have set up a decision-diagram to help the investigator to select the right extraction method (Figure 3.35).

His diagnosis is crucial for the recovery of information and for choosing the right method that could recover the data. A diagnostic error could cause the permanent destruction of the data. As shown in the diagram (Figure 3.35), depending on the type of diagnosis, some extraction methods are preferable to others.

We will discuss in chapters 4, 5, 6 and 7 methods to push back the current limitations of conventional extraction methods. But for now, we present our choices that led us to realise this decision-diagram and the selection of extraction methods. In all cases, a diagnosis of the memory component must be made before any technical intervention, in particular by 3D computed tomography.

Thus, we present in more detail in chapter 4 the theoretical functioning of micro-tomography and 3D reconstruction algorithms.

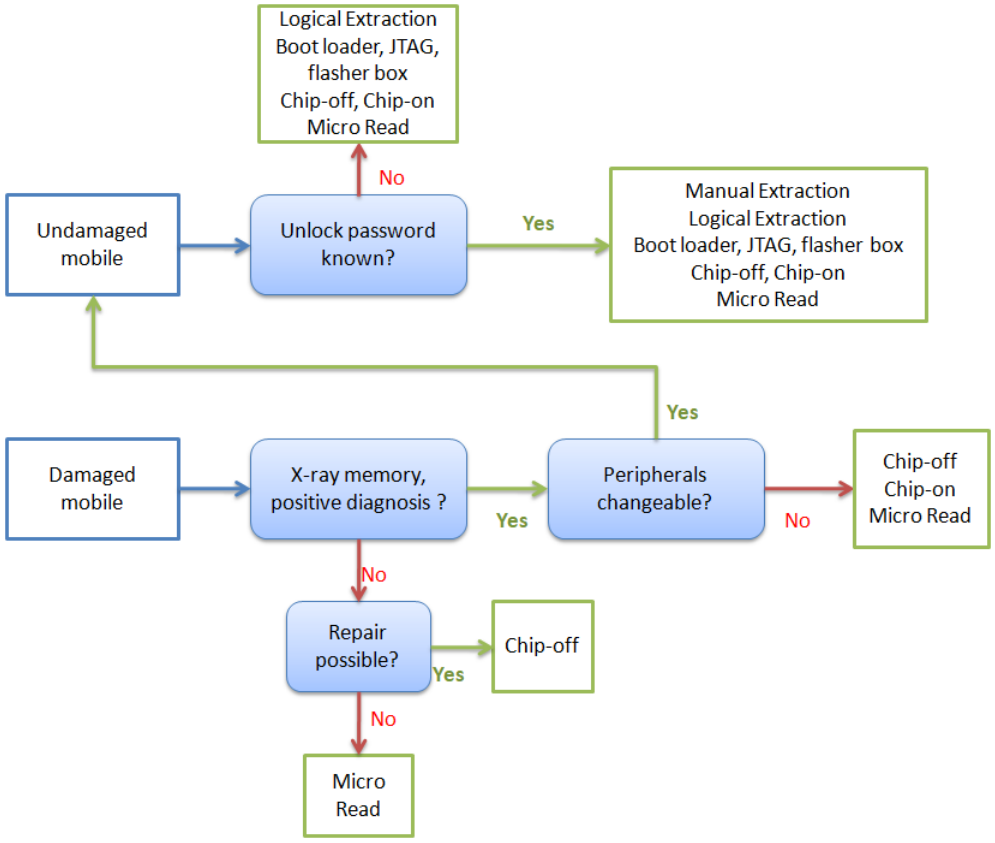


Figure 3.35: Damaged and Undamaged Mobile Devices Decision Diagram

The first level that the expert must identify is the destruction of the external devices (screen, keyboard, battery, palmar unlock sensor, facial identification sensor), with no impact on the PCB or the components like memory, CPU and crypto-chips (Figure 3.36).

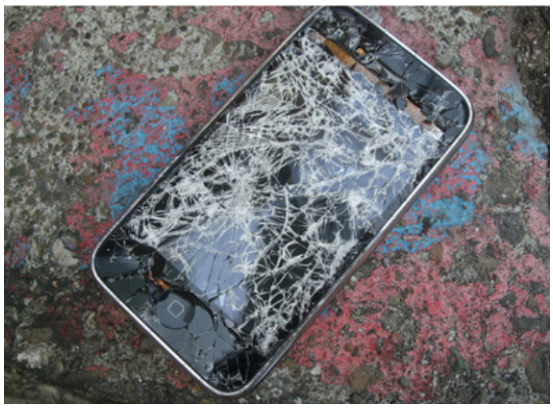


Figure 3.36: iPhone With External Device Destruction Following a Shock

In this case, although apparently the phone seems unworkable, all the extraction levels previously presented are possible. The investigator may then choose to perform a man-

ual extraction, provided that the peripheral devices are replaced. The logical analysis can also be chosen insofar as the logical extraction device is not affected (or after its replacement). In the same way, physical extraction by JTAG, boot loader, and flasher boxes can be performed if the logical extraction device is not touched. Finally, the other types of removal (chip-off, chip-on and micro-read) can be implemented without difficulty depending on the investigator's needs, even if the peripherals are not reparable.

The second level is the identification of the mobile phone that has been possibly immersed. This requires special attention, although the mobile phone does not seem damaged. For example, a phone "sealed" for forensic analysis that has been immersed needs to undergo a special PCB drying process in an oven before being desoldered or lapping.

The drying time is generally 24 hours at 60 °C. The phone (Figure 3.37) should never be turned on, at the risk of destroying the data. Thus, chip-off, chip-on and micro-reader techniques are the only techniques that could be used in this case.

The chip-off technique must be carried out at the end of the drying process; otherwise, the moisture between the component and the PCB would destroy the BGA component (CPU, Memory, crypto-component) by a popcorn effect.

Figure 3.37: Mobile Device Damaged by Immersion

As is it usually the case after an air crash (Figure 3.38) or an explosion following an attack, if the phone is more severly damaged (broken PCB), the levels from 3 to 5 are potentially possible.

Figure 3.38: Mobile Phone After Air Crash

We will discuss in chapter 5 the repair methods of damaged electronic components that we have implemented during this thesis. Therefore, after the component repair, the investigator is back to the situation of an undamaged mobile phone extraction, and he must apply the previous specific extraction procedures. During the investigator's diagnosis, if the radiography shows a silicon memory destruction, the only method to be implemented is a micro-read. However, currently, this method has no chance of success if the memory binary data of the chip is damaged.

Finally, the last case is when the damaged mobile phone uses encryption with an encryption key inside the CPU and/or crypto-processor and/or crypto-components. In this precise case, the previous extraction methods are not possible because it is not possible to decrypt the memory's data. The only solution is to develop an advanced repair technique called forensic transplantation.

3.4 Transplantation: Advanced Forensic Techniques Used to Extract Data from Damaged and Secured Mobile Devices

A forensic transplantation (Figure 3.39[18]) consists of taking the electronic components off a defective phone [Androulidakis, 2016] and re-soldering them on a functional phone board. This transplantation is similar to human organ transplantation.

A healthy organ (electronic component) in a failing body (broken main board) is transplanted in place of a failing organ in a healthy body (donor board).

[18]Ecole Normale Supérieure: https://inutile.club

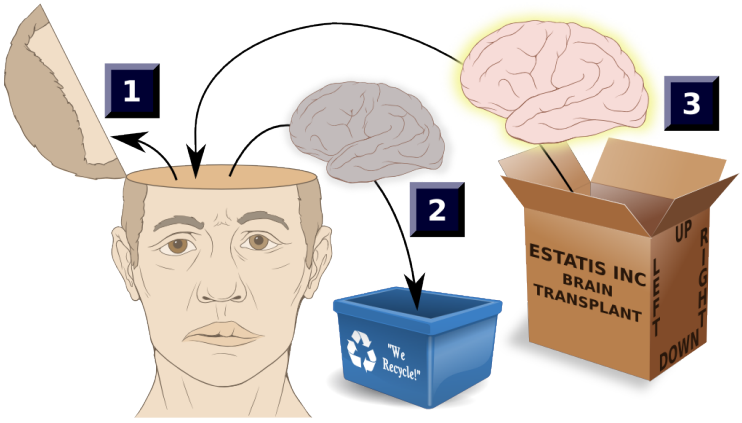


Figure 3.39: A Healthy Body (Board) Used To Salvage a Rescued Brain (Chip)

The functional phone board is called a donor board (Figure 3.40). The preparation of the donor board is an essential step in removing the components from it without damaging the board.

Usually, in this operation, the components of the board have been destroyed by lapping or chip-off.

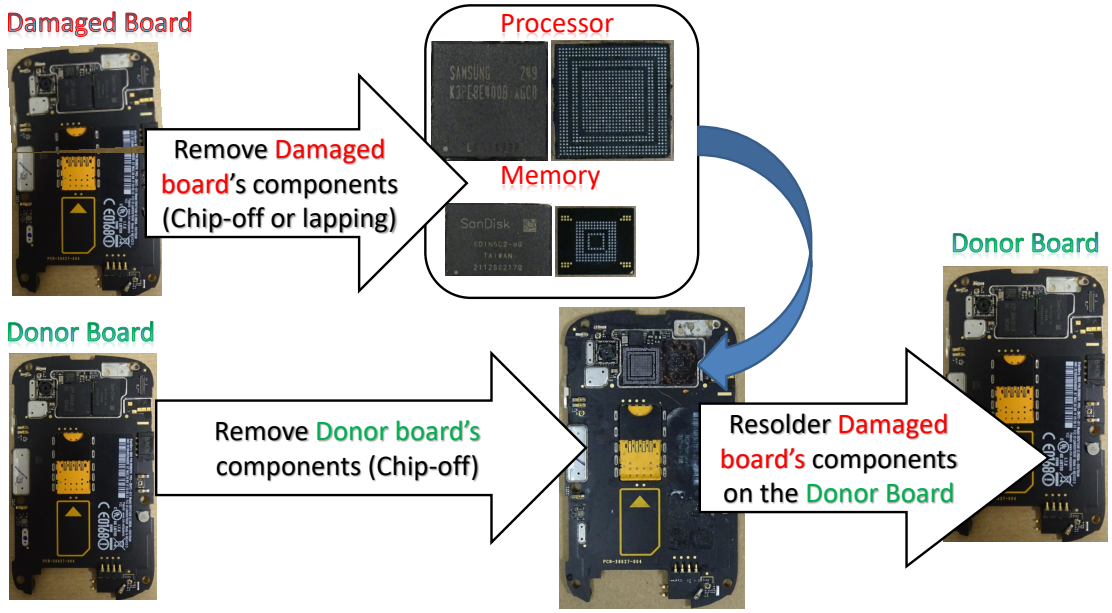


Figure 3.40: Transplantation Mechanism

On the other hand, the components of the damaged board (main board) are removed and then re-soldered onto the donor board. Once these two steps are completed, all that remains is to turn on the phone and make it operational again. This allows the investigator to retrieve the data inside (or bypass some security mechanism) that will

serve as evidence in court.

Similarly, other transplantation methods [Vidas, 2010] allow the reading of the volatile memory (DRAMs retain their contents for several seconds after power is lost), such as a "cold start attack" involving cooling memory chips with liquid nitrogen and transplanting them into a donor device capable of reading memory chips [Halderman et al., 2009].

Thus, we can use forensic transplantation to extract and analyse the data held on the damaged device, whether the memory acquisition is volatile or not.

Nowadays, new technologies are emerging that make the traditional transplant technique outdated. One could mention the implementation of Package on Package (PoP) components, the epoxy underfill between the neighbouring components and between components/PCB, and the implementation of coupled cryptographic chips that complicate the traditional process of transplantation. In chapter 7 we will develop the limitations of transplantation and bring innovative solutions.

4

Chip-off Improvement: the Role of 42Sn/58Bi Solder

As we write these lines, the desoldering and reading technique of the flash memory chip [Breeuwsma et al., 2007] has become the de facto standard. Furthermore, the re-soldering of electronic components to the exhibit device or a test board is sometimes necessary during forensic investigations. In some cases, data manipulation, performed during chip-off and chip-on, may also help to recover forensically significant data (Chapter 3).

Chip-off/chip-on techniques involve two risky steps during which forensic data can be irreversibly lost. Low-temperature soldering has a decreasing risk factor [Jongh, 2014] but can be considered an expensive process when considering the cost of commercial balls and difficult to implement without specialised expensive equipment. In our study, we proposed and developed a new low-cost and low-temperature reballing technique.

4.1 Context: Do Not Stress Components During Investigations

The re-soldering of micro-BGA (Ball Grid Array) components is often necessary during forensic investigations. Such re-soldering usually occurs in two scenarios. In the first *in vivo* scenario, a component is extracted from the exhibit board and analysed (or unlocked) externally before being re-implanted in the exhibit board. In the second *in vitro* scenario, the extracted component is implanted in an external test board or an unlocked device of the same brand and model. As we saw in chapter 3, we call such manipulations chip-off/chip-on procedures.

Although at first glance, re-soldering a chip on a Printed Circuit Board (PCB) may be seen as a simple operation, doing so may be far from harmless for both transplant chip and the receiving PCB. Chip-off/chip-on soldering requires the reflowing of parts using high-temperature hot air or infrared heating. This can damage the transplant chip or

the board due to overheating. In addition, this procedure may also cause close electronic components to slide and move away from their initial position.

To solve these issues, the first idea we had for this thesis was to think that it is desirable to use low-temperature chip-off/chip-on processes. Reflow techniques in which the temperature does not exceed 138 °C prevent 67Sn–33Pb or traditional Pb-free alloys (205 °C for 86.4Sn–5.1Ag–8.5Au and 183 °C for 63Sn–37Pb [Vianco et al., 2004]) from melting and do not stress the chip-off/chip-on components too much. The high price of such alloys has been for a long time an obstacle to their adoption.

Despite constant technical progress, low-temperature reballing 300-microns spheres of 42Sn/58Bi still cost several thousands of euros per jar and their use requires a lot of time to be implemented. In a limited economic context, investigators must also control the expenses of their investigations and must be vigilant about the legal costs incurred during their investigations.

This chapter focuses on using 42Sn/58Bi alloy (€30/500g[1]) and reballing stencils (€13/12 Chinese stencils[2] or €10/1 homemade stencil[3]) to achieve affordable and efficient low-temperature reballing.

We start by reviewing the material used in our study, the 42Sn/58Bi alloy properties and analyse the sourced paste's composition. Next, we describe our new reballing method and its application to the PCB rework.

4.2 Material

4.2.1 X-ray Tomography

As we saw in the previous chapter, X-ray micro-tomography can be used by the investigator to diagnose a damaged phone. This diagnosis asserts or excludes any damage to an electronic component. If the damage results from a silicon fracture, the investigator should perfom a micro-reading if the memory area is not damaged. Otherwise, the component will be classified as destroyed. If other parts of the chip are damaged, such as bonding wires, or insulation components, the investigator must then repair the component.

As part of this thesis, we have implemented a new technique for repairing bonding

[1] http://souvignet.net/r/zk4j538
[2] http://souvignet.net/r/gwe7gox
[3] http://www.laser-techno.com/

wires, which we develop in more detail in chapter 5. In addition to the diagnosis, tomography can be used to check the quality of soldering of the component on the PCB. This control makes it possible to see if there is no short circuit during the component's soldering process. Since the investigator has to re-solder the memory component after reading it (or during the realisation of a transplant), this tomography is essential.

We now introduce the fundamental principle of tomography. The X-ray micro-tomography imaging technique involves acquiring several 2D radiographic projections of an object viewed from different angles and reconstructing a 3D image using reconstruction algorithms. As part of our thesis (Figure 4.3), the X-ray tube and the detector are fixed and the object to be observed is placed on a turntable. The computed tomography consists of three phases: an acquisition phase (setting defined by the investigator), a reconstruction phase (mathematical algorithms) and an analysis phase (post-processing).

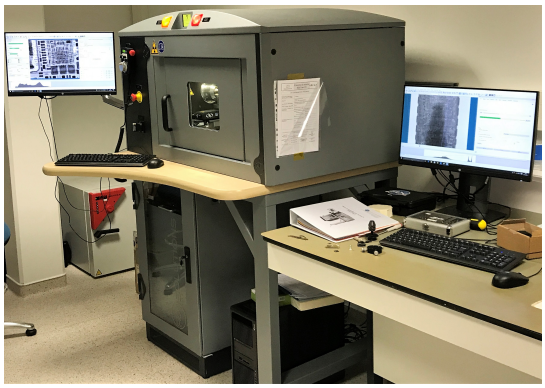


Figure 4.1: General View

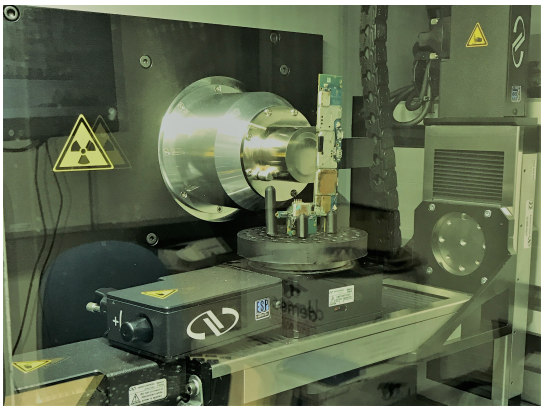


Figure 4.2: X-ray Tube

Figure 4.3: IRCGN's X-ray Tomography Equipment

4.2.1.1 Acquisition

X-ray Generation

The X-rays we use in this thesis come from an X-ray generator. We make use of a beam of primary electrons that are extracted at the cathode by applying a potential difference. These electrons are then accelerated by a high potential difference and bombard a heavy metal anode. This anode is usually tungsten but this can vary depending on the model.

It is the variation of the intensity and the voltage of the potential difference which defines the energy and the flux of X-rays. Thus, a high voltage will lead to X-rays of high energy, and a high intensity will lead to a steady flux of X-rays. When the electrons

arrive at the target, several phenomena occur:

- The slowing of electrons as they pass close to the nucleus of an atom of the target. The electrons are slowed and deflected by the electromagnetic field of the nucleus. The kinetic energy lost by the electrons is emitted in the form of polychromatic X-photons.
- The second phenomenon is the collision of electrons with the electrons of an atom of the tungsten target. During this phenomenon, polychromatic X-ray fluorescence photons are created.

Finally, the photons are filtered, using very dense materials, to allow the passage of the photons that have enough energy to cross the target material. The adjustment of the X-ray intensity and flux is a parameter set directly by the investigator. The denser the material (metal shield protection, mobile phone screen) the more the investigator will have to increase the power to obtain X-rays able to cross the material entirely. However, the more X-rays, the less the contrast is important. The investigator must strike a balance based on his needs for observation.

X-Ray–Matter Interactions

When an incident beam of X-rays passes through a given medium, we observe a progressive disappearance of the number of directly transmitted incident particles, due to the different radiation–matter interactions: absorption by photoelectric effect, Compton and Rayleigh diffusions (Figure 4.4).

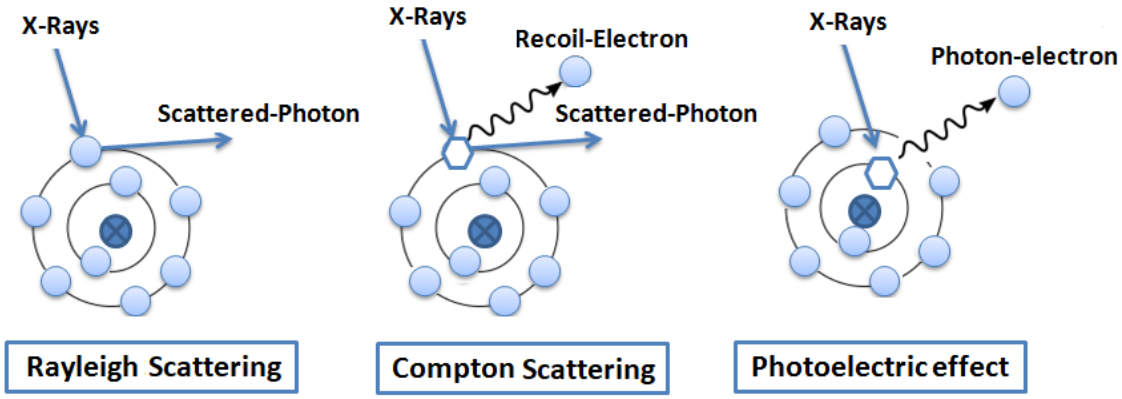


Figure 4.4: X-Ray–Matter Interactions

Depending on the material of atomic number Z crossed and the X-ray energy, we define the linear attenuation coefficient $\mu(E, Z)$. For silicon, the NIST gives the total linear attenuation coefficient, taking into account all the X-ray–matter interactions (Figure 4.5).

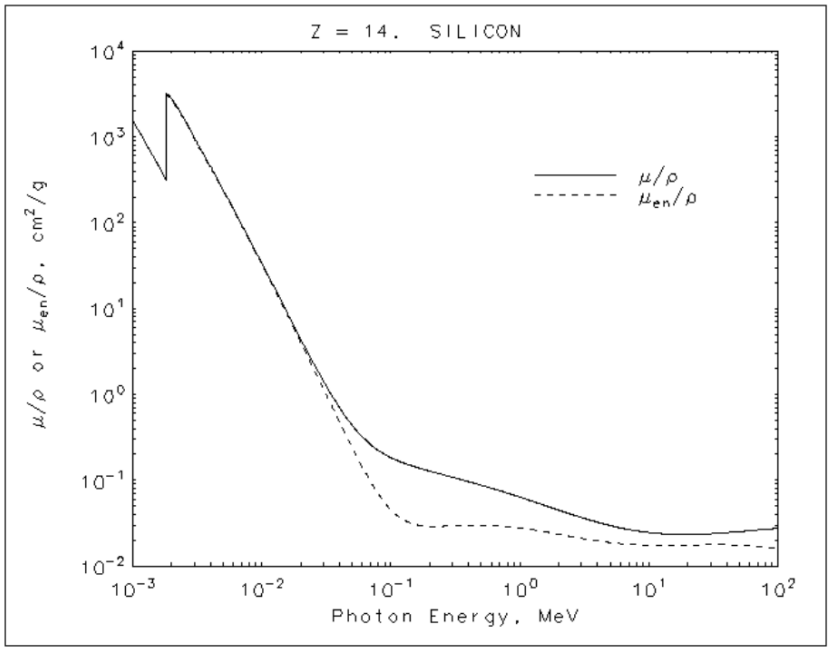


Figure 4.5: X-ray–Silicon Total Linear Attenuation Coefficient (NIST)

The Beer–Lambert law gives the fraction of the incident photons having passed through a thickness l of matter without interacting:

$$\phi_{rx} = \int_S \phi_0(E).e^{-\int_L \mu(E,Z(l))l.dl} dE \tag{4.1}$$

with a polychromatic beam S and for a distance L between the source and the detector.

It is thus possible, by measuring the outlet flow, to determine the density and the thickness of the material traversed by the X-rays.

4.2.1.2 Reconstruction

Tomographic reconstruction is an inverse problem (starts with the results and then calculates the causes) because we need to find the 3D volume of the object from a series of 2D measurements acquired during its complete rotation (Figure 4.6, [Flannery et al., 1987]).

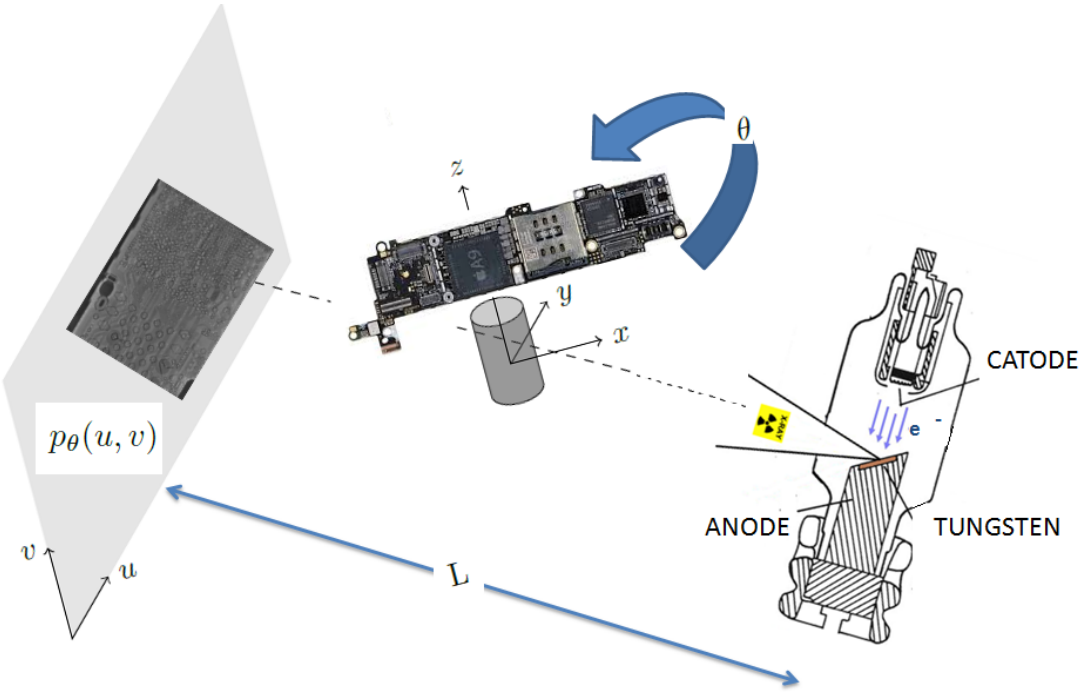


Figure 4.6: X-ray Tomography Principle

The filtered backprojection is based on the Feldkamp algorithm and the inversion of the Radon transform:

$$p_\theta(u,v) = \ln(\frac{\phi_0}{\phi_{rx}}) = \int_L f(x,y,z)dl \tag{4.2}$$

$p_\theta(u,v)$ is what is measured by the detector at the point (u,v) for an angle θ. What the investigator wants is to determine the function $f(x,y,z)$ which is the volume field we are seeking to reconstruct.

The Radon transformation allows us to define the projection operator $p_\theta(f,u)$ which sums the projection set for an angle θ on the length L:

$$p_\theta(f,u) = \int_{L,\theta} f(u.\cos(\theta) + v.\sin(\theta), -u._s in(\theta) + v.\cos(\theta))dl \tag{4.3}$$

By summing all the contributions on the complete rotation, we can express the projection operator as a simple mathematical function dependent on angle theta:

$$R_p(x,y) = \int_0^{2\pi} p_\theta(x.\cos(\theta) + y.\sin(\theta)d\theta \tag{4.4}$$

The reconstruction goes on afterwards in the frequency space (no longer temporal):

$$TF[p_\theta(w)] = TF[f(w.cos(\theta), w.sin(\theta)] \tag{4.5}$$

The technique of filtered back-projection resides in the introduction of a filter in the frequency domain by a ramp filter $[w]$ and a low-pass filter $H(w)$. The filtered back-projection comes into the temporal space again by an inverse Fourier transform:

$$f(x, y, z) = TF^{-1}[TF(p_\theta(w).[w].H(w))] \tag{4.6}$$

The function $f(x, y, z)$ is therefore perfectly determined at all points. A processing of the signal in grey scale makes it possible to translate this function into a 3D representation (Figure 4.7).

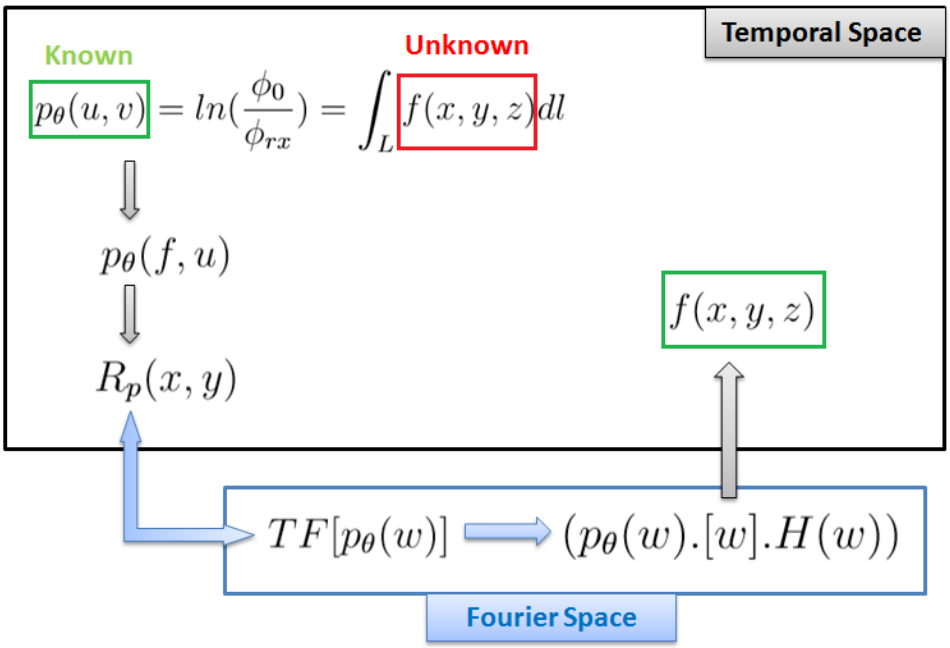


Figure 4.7: Filtered Back-Projection Algorithm

There are several tomographic reconstruction algorithms, such as Algebraic Reconstruction Technique ([Gordon et al., 1970]) or Monte Carlo methods ([Kaipio et al., 2000]). We will not go into details here, but the investigator has the possibility of selecting the algorithm which seems to be the most appropriate. This selection is part of the expert's work, who must select the right reconstruction algorithms according to his observation needs. The reconstruction's artefacts, specific to each algorithm, must be known in order to select the correct reconstruction method.

4.2.1.3 Post Treatment

In the reconstruction phase, a mathematical reconstruction algorithm then allows access to the volumetric data of the object from the various 2D projections. After reconstruction, it is possible to carry out a post-processing of the images obtained (Figure 4.8):

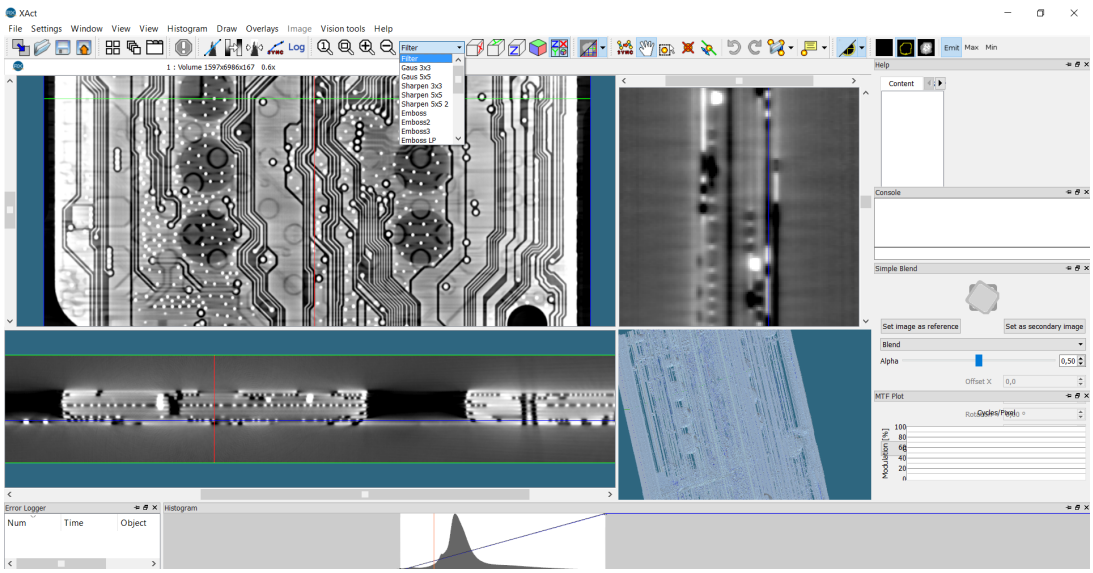


Figure 4.8: Tomography Post Treatment Using X-Act Software

This step corresponds to the post-processing filter application (Gauss filter 3x3, Gauss filter 5x5, Sharpen, Emboss, etc.), contrast change (allows the localisation of density element of interest), measurement distance, tracking and more generally all reverse-engineering operations (chapter 7).

4.2.2 Scanning Electron Microscopy (SEM)

The Scanning Electron Microscopy (SEM), as we saw in the previous chapter, can be used when reading transistors on the silicon components that are severely damaged (physical extraction by micro-read). We will show in this chapter that the SEM (Figure 4.9) is a precious tool in the determination and optimisation of the temperature curves used during soldering or desoldering (chip-off/chip-on methods). In this chapter we have bypassed the principle of first use of SEM used generally by investigators in electronics (micro-reading technique) and we have made use of the SEM to solve more complex hardware problems inside modern mobile phones: low temperature soldering.

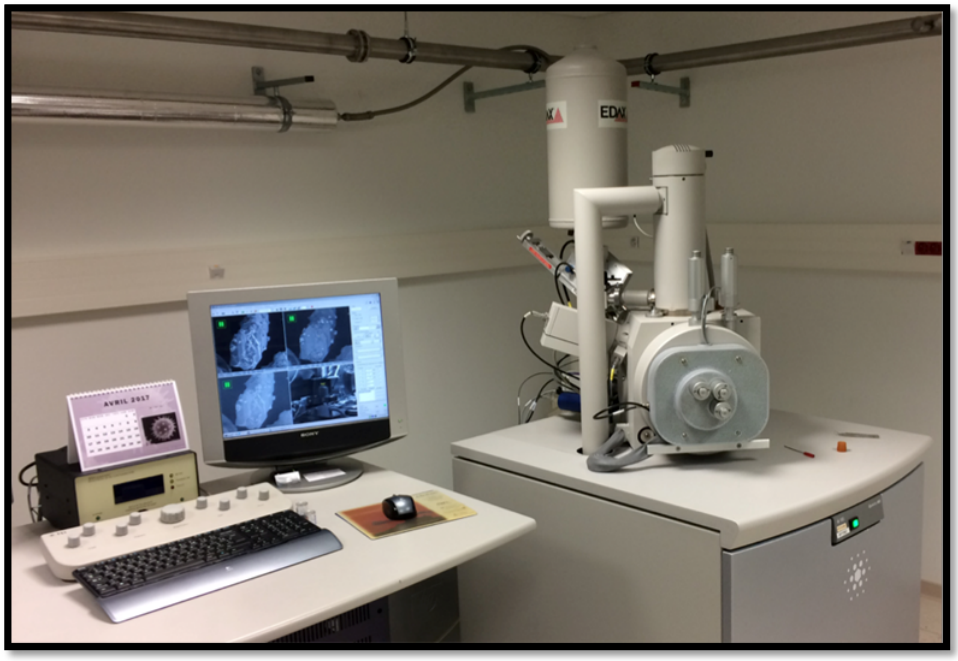


Figure 4.9: IRCGN's SEM

Electron microscopy is based on the same principle as optical microscopy, except that the incident ray consists of an electron beam instead of visible rays of light. X-ray radiography uses the radiation–matter interactions to characterise the state of the internal layers of a component.

SEM uses particle–material interactions to characterise the surface condition of the object to be analysed. This surface analysis is a topography analysis (surface condition and visualisation), as well as a quantitative analysis (chemical composition).

Unlike X-ray micro-tomography in which an electron beam accelerates on a tungsten anode and whose impact creates X-rays which are then sent onto the sample, in the case of SEM, it is the accelerated electrons that are sent directly to the sample (Figure 4.10). The electrons are accelerated by an electrical potential difference and are condensed by magnetic lenses.

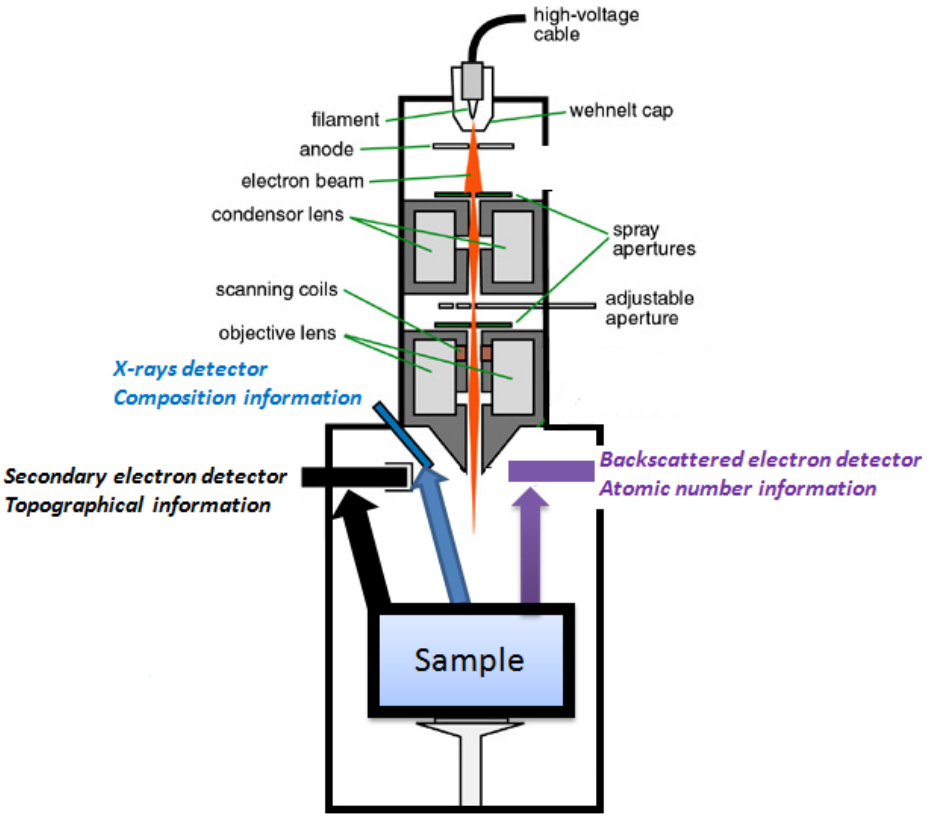


Figure 4.10: SEM Fundamental Principals

Fundamental physics describes the electron beam behaviour as it enters the sample and comes into contact with the nuclei and electrons of the electron cloud of the target material. The primer electron is dispersed by a succession of elastic and inelastic interactions with the target (Figure 4.11):

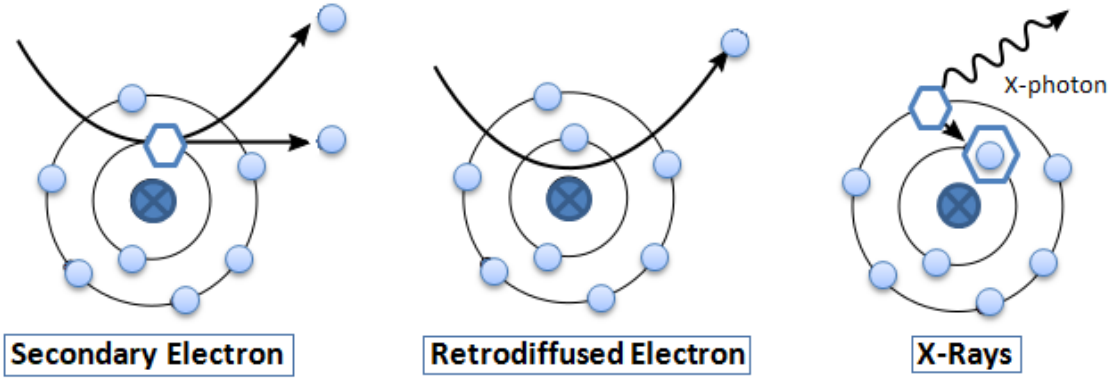


Figure 4.11: Particle–Material Interactions

- The first reaction is an inelastic reaction. In this case, the primary electrons interact with the electrons of the atomic electron cloud, and they emerge with a loss of energy. A secondary electron is emitted, and the target atom is ionised. The

detection of these electrons provides information on the topography of the sample.

- The second reaction is an elastic reaction. The primary electrons emerge without loss of energy, keeping their kinetic energy and momentum. In this reaction, the electrons do not exchange energy with the atoms of the sample. They undergo an interaction due to the Coulomb force within the nuclei of the atoms. The higher the atomic number of the atom, the stronger the signal and the clearer the area of the image. This is the phase contrast. The image obtained is, therefore, a function of the chemical composition of the sample.
- Finally, the third interaction is the impact of a high-energy primary electrons that ionise the atoms within their inner layers of electrons. Thus an electron of an inner layer is ejected, and it is replaced by an electron of an upper layer. A photon of energy equal to the difference between the two levels of electron energy is emitted. The hole in the upper layer is filled by another electron of an even higher layer with the emission of polychromatic X-photons. A cascade is thus created. The study of X-ray photon cascades allows a quantitative analysis of the chemical composition of the sample (contrast as a function of the atomic number of the surface elements).

SEM images are formed from a mixture of variable proportions of the secondary electrons, backscattered electrons and X-ray recording signals. The different signals are recorded by three different sensors: a secondary electron sensor, a broadcast electron sensor and an X-ray sensor. In a scanning electron microscope, the image is obtained sequentially point by point by displacing the primary electron beam on the surface of the sample. The image is then reconstructed using the signal generated by the various detectors to modulate the brightness of a cathode ray tube.

4.2.3 42Sn/58Bi Fundamental Properties

Our 42Sn/58Bi solder alloy used in this thesis is a eutectic composition [Song et al., 2004, Puttlitz and Stalter, 2004]. The alloy melts and solidifies at the same temperature, in contrast to the usual mixtures. From a merger point of view, the alloy behaves like a pure body (it melts and solidifies at the same temperature) [Miao and Duh, 2001, Okamoto, 2000]. This eutectic mixture reaches a solidus-liquidus phase transition at 138 °C [Braga et al., 2009, Rechchach, 2011] (Figure 4.12). The advantage of this solder is that it has a low melting point [Predel, 1991, Torres et al., 2012].

In forensics, the main advantage of low temperature soldering is the protection of evidence, because it has little impact on neighbouring components (temperature/matter interaction), which grandly reduces deterioration risks (data modification, destruction).

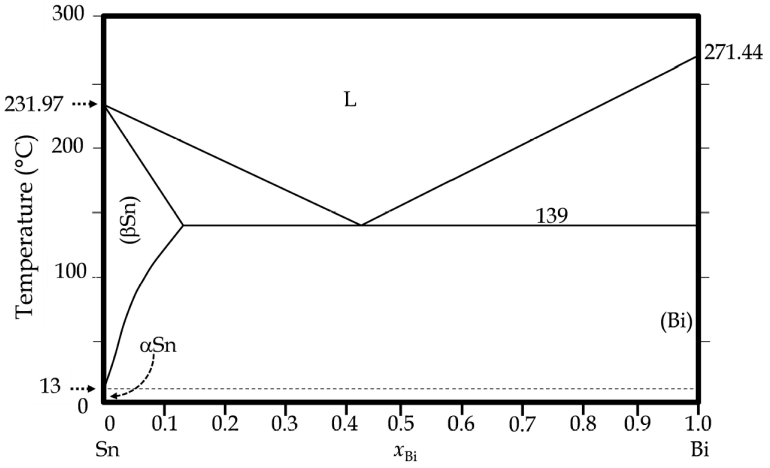


Figure 4.12: 42Sn/58Bi System Phase Diagram [Aller et al., 1996].

We will hereafter call *"the assembly"* the physical circuit consisting of the target BGA component, the PCB and the solder paste. For the study of our solder ball, we will use our scanning electron microscope coupled to an Energy-Dispersive X-ray Spectroscopy (EDS analysis). This reveals the characteristic spectrum of the mixture of substances contained in the analysed part. By observing our past using our SEM, we see inhomogeneous beads ranging from 25 microns to 45 microns embedded in a solvent (Figure 4.13).

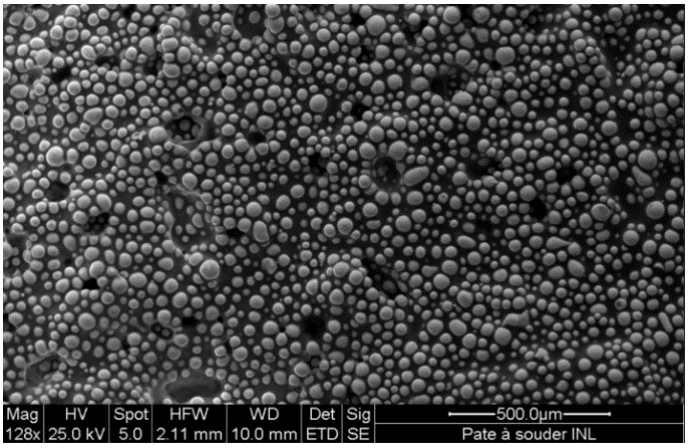


Figure 4.13: 42Sn/58Bi Past Observed at 500 µm Scale (SEM)

The study shows that the balls appear in two colours with uniform and homogeneous spherical distribution solidification (Figure 4.14).

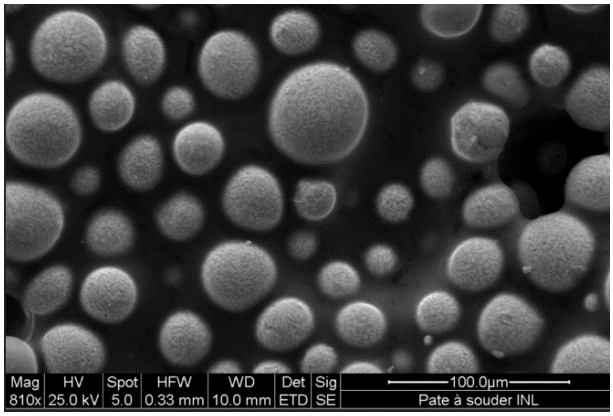


Figure 4.14: 42Sn/58Bi Alloy Observed at 100 µm Scale (SEM)

We carried out the analysis of the micro-beads without the solvent by elemental chemical microanalysis of the balls' surface by Energy Dispersive X-ray Spectroscopy (Figure 4.15).

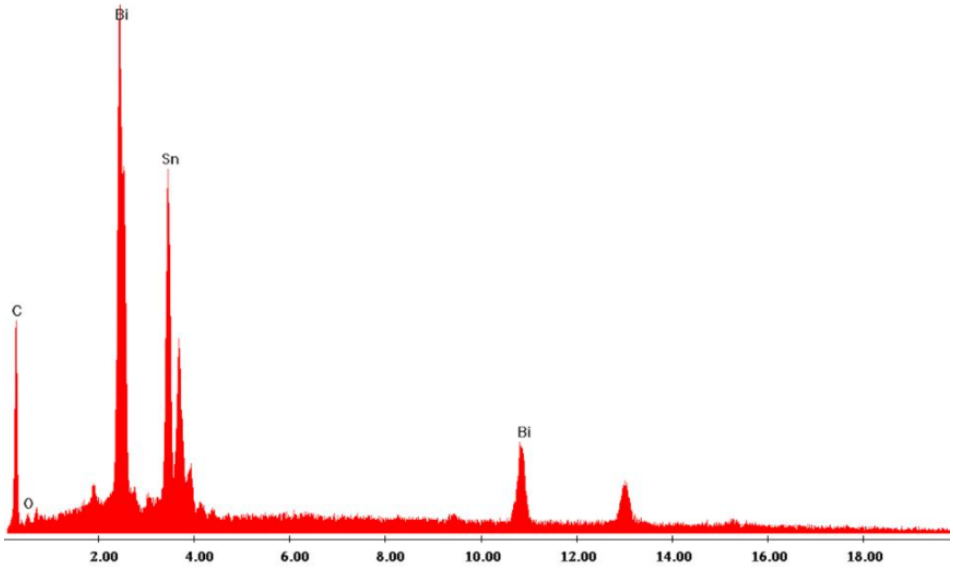


Figure 4.15: 42Sn/58Bi Alloy Energy Dispersive X-ray Analysis

Our past is a tin–bismuth (Sn–Bi) binary composition. The bismuth precipitation takes place in a tin matrix, and the cooling rate has a great influence on the microstructure properties of the alloy [Zeng and Tu, 2002, Crocker et al., 1973, Puttlitz and Stalter, 2004]. Due to a high bismuth concentration, cracking issues can be expected and will be characterised in the experimental results section in order to help the investigators select the right process.

4.3 Method Developed

We now describe our new re-balling step process. This process is necessary before re-soldering the BGA components into devices for which no alternative ways to access the device's contents are known. Our study focuses on the re-balling step of a flash Hynix H9DP4GG4JJAC chip (Figure 4.16).

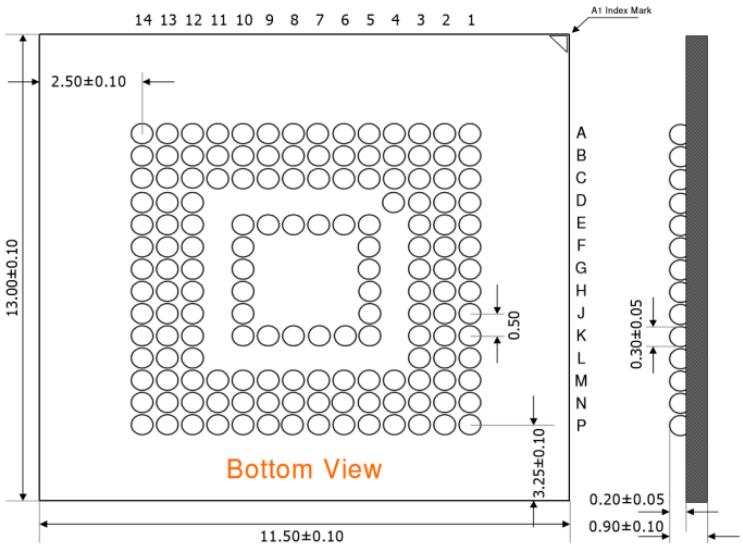


Figure 4.16: Hynix H9DP4GG4JJAC Ball Grid

The re-balling process we used was realised in steps, as follows:

Step 1: Clean up the chip using a soldering iron and remove flux (Figure 4.17).

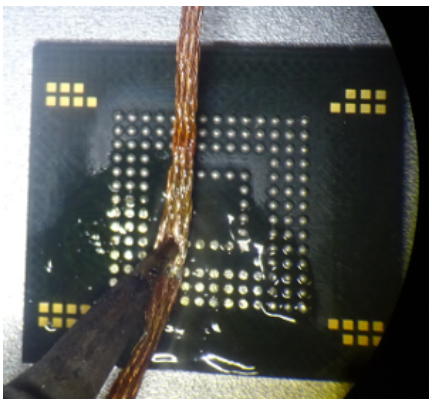


Figure 4.17: Step 1

Step 2: Stencil selection adapted to the chip to reball (Figure 4.18). Better results were obtained using stencils of 135-micron thickness, 500-micron pitch and 300-micron diameter holes.

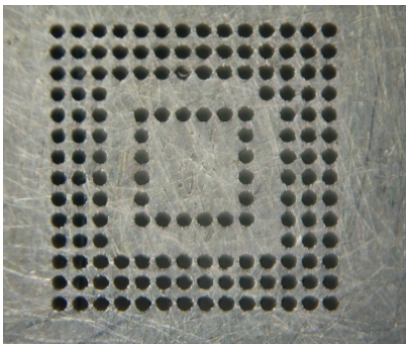


Figure 4.18: Step 2

Step 3: We position the stencil by superimposing it on the target chip to align the stencil (Figure 4.19).

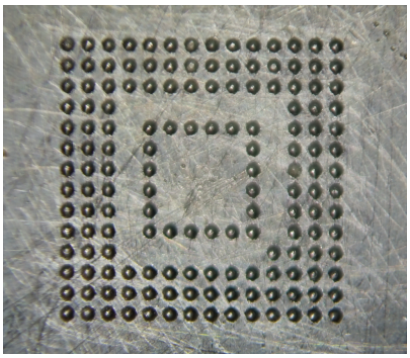


Figure 4.19: Step 3

Step 4: We apply the soldering lug composed of beads 25–45 microns in diameter presented in its solvent (Figure 4.20).

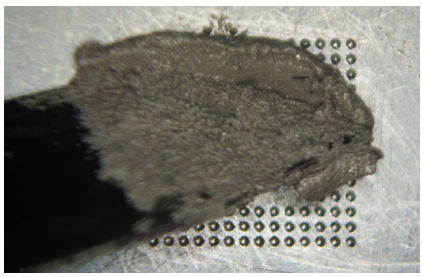


Figure 4.20: Step 4

Step 5: We remove the excess dough with a flat knife (Figure 4.21) to have the same amount of alloy and thus avoid obtaining balls of different sizes.

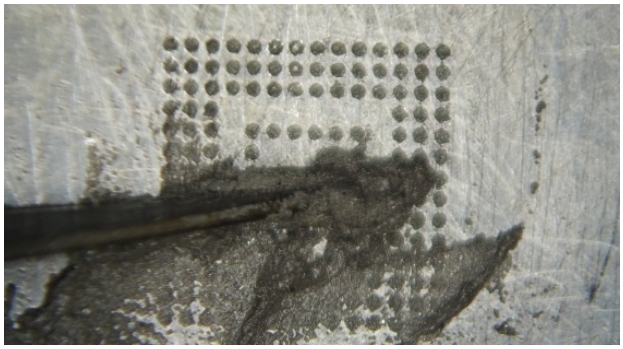


Figure 4.21: Step 5

Step 6: We evenly apply a gentle air flow (8 l/min) at 160 °C (measured using a thermocouple) for 4 seconds on the stencil (Figure 4.22) to begin the agglomeration process between the beads. It is then ready to apply an air stream (14 l/min) at a more consistent temperature of 200 °C.

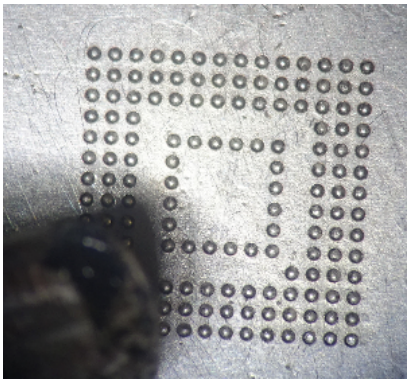
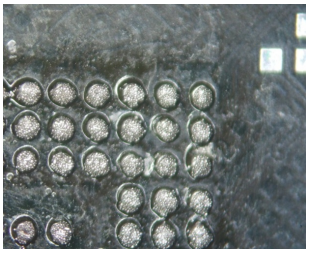


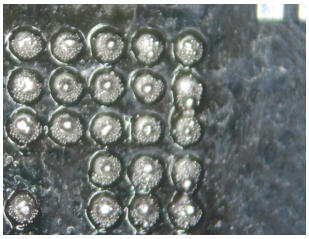
Figure 4.22: Step 6

The balls will naturally fall off the stencil and position themselves on the PCB by capillarity. Due to this capillarity, the stencil does not have to be cleaned afterwards. Thus, the stencil can be used several times by the investigators without risk of destruction. The temperature factor is the only possible source of alteration.

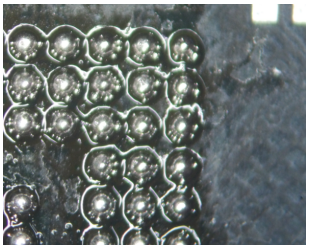
At this stage, the beads are not completely agglomerated, and many microballs remain mixed with the solvent (Figure 4.23a). To get nice 300-µm balls, we apply an air stream (14 l/min) of 200 °C for 12 seconds (Figures 4.23b to 4.23d).



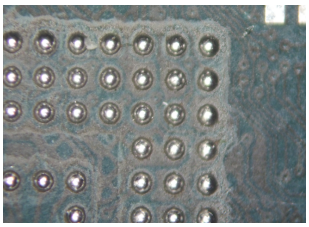
(a) Stencil Removed After 4 Seconds of Heating at 160 °C (8 l/min)



(b) Air Stream at 200 °C (14 l/min) for 5 Seconds



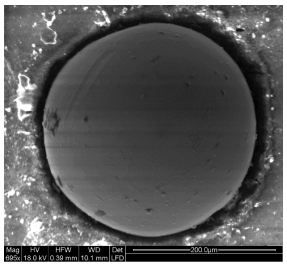
(c) Air Stream at 200 °C (14 l/min) for 8 Seconds



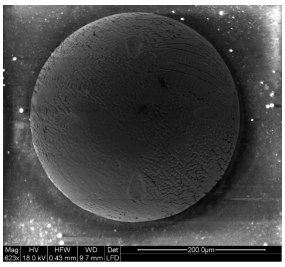
(d) 12 Seconds of Heating at 200 °C (14 l/min) and removed flux

Figure 4.23: Balls Observed at Binocular Microscope During the Agglomeration Process

Using the SEM, we decided to look at a ball that we created with the stencil (Figure 4.24a) and compare it with a commercial leaded solder ball (Figure 4.24b). The underlying goal is to check if our mixture ball is close to one of the industrial balls.

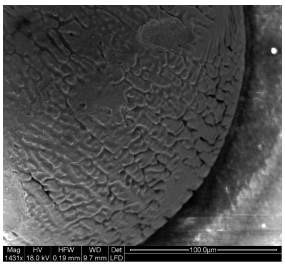


(a) Sn–Bi Alloy



(b) 67Sn–37Pb Alloy

(c) Sn–Bi Alloy (50 µm scale)



(d) Sn–Pb Alloy (100 µm scale)

Figure 4.24: Sn–Bi vs Sn–Pb Alloy Comparison

4.4 Experimental Results

4.4.1 Main Results

The microstructure of brased joints is significantly influenced by the following reflow conditions: maximum peak temperature, cooling rate and volume of solder [Rechchach, 2011]. We are mostly interested in investigating the influence of the cooling rate on the solidification dendrites. The maximum temperature was set to 160 °C and the solder volume corresponded to a sphere of 300 µm diameter (Figure 4.25).

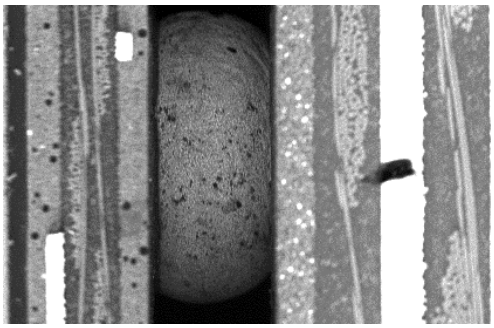


Figure 4.25: (SEM) Sn–Bi Sphere between eMMC (left) and PCB (right), Slow Cooling

The following experiments investigate the quality of the solder joint as a function of different slopes of temperature drop. The first experiment refers to a rapid reduction in temperature. In the following reflow curve, the maximum temperature (160 °C) is reached in 135 seconds (Figure 4.26). The maximum temperature is maintained for 60 seconds before starting the cooling down phase at a cooling rate of 1.9 °C/s for 58 seconds.

Figure 4.26: Fast Cooling (1.9 °C/s) Reflow Profile, X-axis (seconds), Y-axis (Celsius)

In the case of rapid cooling, the solidification pattern has a uniform structure with parallel solidification dendrites. Soldering is performed over the entire reception area

[Torres et al., 2012]. It has the same width over the entire wafer. In our experiments, the solder samples were mechanically polished using standard metallography techniques before being examined under a scanning electron microscope (Figure 4.27).

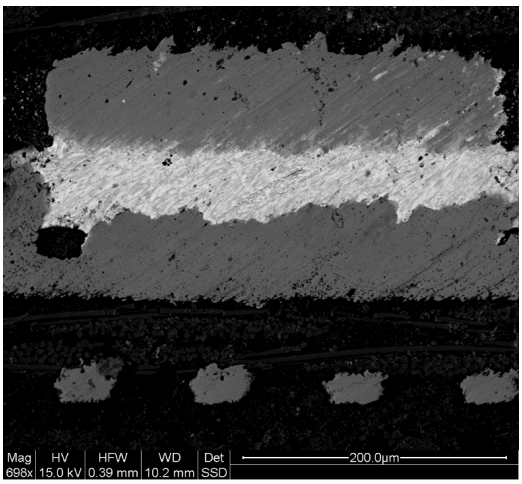


Figure 4.27: Polished Section of Soldering Using Standard Metallography Techniques

The second experiment focuses on a prolonged temperature descent. The slope is 0.13 °C per second. Figure 4.28 shows the temperature profile used to make the junction.

Figure 4.28: Slow Cooling (0.13 °C/s) Reflow Profile, X-axis (seconds), Y-axis (Celsius)

With the same temperature increase rate and the same maximum temperature, Figure 4.31 shows an example of the microstructure of 42Sn/58Bi solidified balls at two different cooling rates (Figure 4.29: 1.9 °C/s and Figure 4.30: 0.13 °C/s) examined under scanning electron microscope and each time using the same procedure.

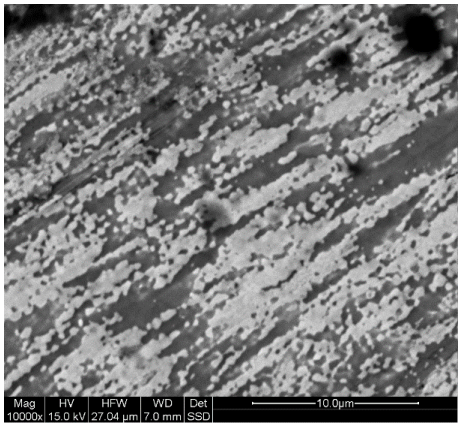


Figure 4.29: At Fast Cooling Rate

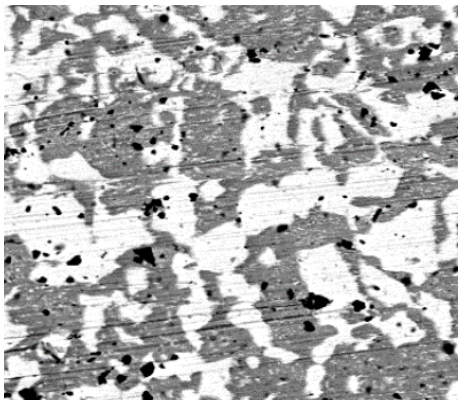


Figure 4.30: At Slow Cooling Rate

Figure 4.31: Microstructure of 42Sn/58Bi Solidified Balls

Slow cooling rates lead to crack formation and large grain size, which leads to questioning the reliability of the solder joint [Zeng and Tu, 2002, Crocker et al., 1973]. The results confirm that by using slow cooling, the beads will reform and detach; the soldering is not successful; bismuth and tin will agglomerate into coarse structures (Figure 4.32).

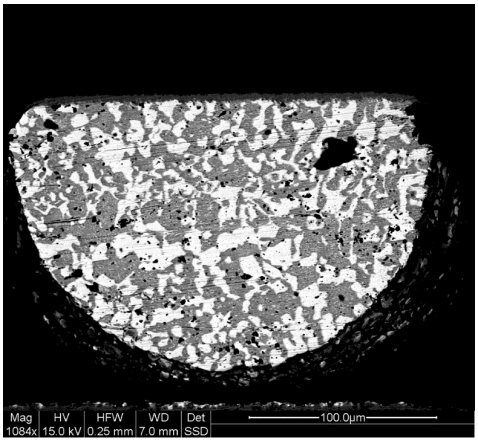


Figure 4.32: Bismuth and Tin Agglomerated Into Coarse Structures

We also observe an irregularity of the solidification structure on the cut edge when it is cooled slowly. This irregularity is all the more important when the temperature descent curve is slow. If the descent is too long the structure is weakened and small cracks appear. The addition of several small cracks generates even larger cracks, which are directly visible. The results also show the appearance of a crack within the ball that was caused by mechanical stresses between the chip and the PCB (Figure 4.33).

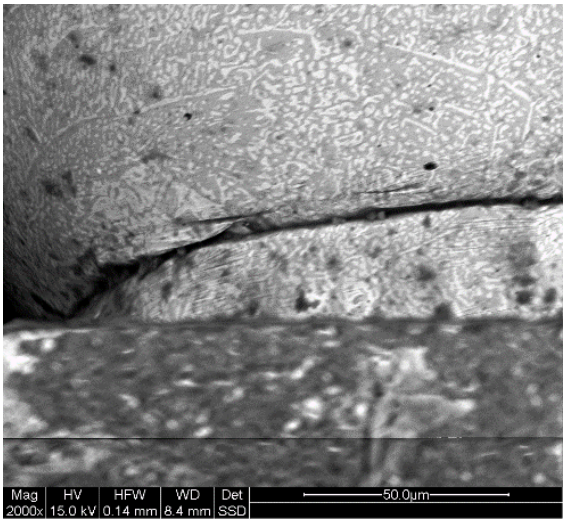


Figure 4.33: Crack Within the Ball Caused by Slow Solidification

We observe that the cooling rate affects the distance between the PCB contacts and the chip. During rapid cooling, the liquidus phase tends to spread and thus reduce the PCB-to-chip distance (which averages 50 microns). Rapid solidification tends to keep this configuration (Figure 4.34).

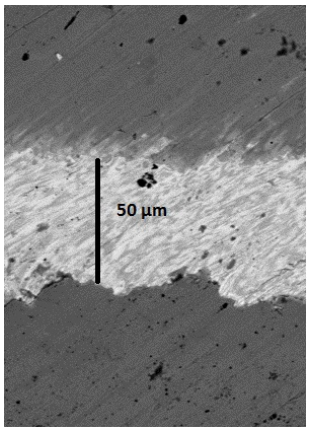


Figure 4.34: Fast Solidification

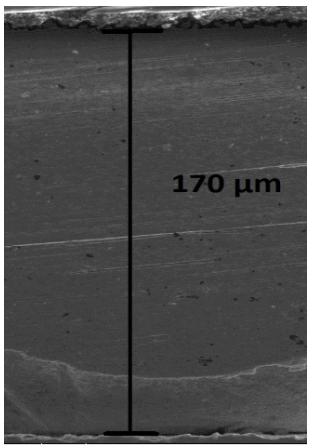


Figure 4.35: Slow Solidification

Figure 4.36: Distance Between PCB and Chip

Conversely, during slow solidification (Figure 4.35), the PCB-to-chip distance is greater than the PCB-to-chip distance obtained by rapid cooling (170 microns average). Soldering quality is directly impacted and the investigator will have to take this into account during the low temperature soldering process.

4.4.2 Stencil Selection

Although the Chinese stencils tested are of interest because of their low price, they have several disadvantages in terms of their integrity and delivery. Paste/stencil usage becomes the main issue compared to sphere reballing when considering a chip that is not supported by the current stencils. While low-temperature balls can be manually placed, the paste cannot be easily spread over the chip pads.

It is then necessary to quickly get a suitable stencil. We also sourced local stencil designers[4] and managed to obtain affordable stencils on demand (less than €10/stencil including our design and production costs) delivered within 3 days (Figure 4.37).

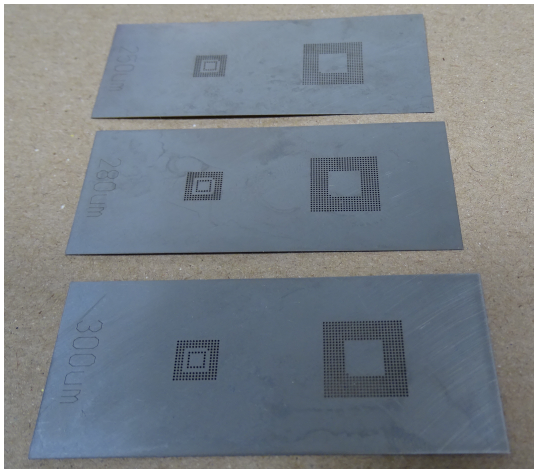


Figure 4.37: On-Demand Stencils of Our Own Design

These stencils are made using laser cutters which are normally used to etch aluminium frames. Based on empirical tests using 117-, 127-, and 152-micron thickness stencils, we selected stencils of 127 microns in thickness with holes of 300 microns (diameter) to make quality balls of 300 microns in diameter. In fact, thin stencils will produce balls which are not strong enough to withstand mechanical stress. Conversely, stencils that are too thick will produce oversized balls which would tend to marge. We have therefore determined and quantified that the investigator has the possibility of adjusting these parameters himself to arrive at the best results according to the component that he has to treat. Thus stencils of small thickness ($\leq$ 117 microns) will be recommended for the re-balling of very fragile components and therefore the distance between each ball centre is low. Stencils of high thickness ($\geq$ 152 microns) will be recommended for the re-balling of less fragile components and therefore the distance between each ball centre is large.

[4] http://souvignet.net/r/jj3tel4

4.4.3 Reflow Error Analysis

In order not to create differences in the amount of material per ball, it is important to apply again a high-temperature air flow (especially with a slow air flow of 160 °C at 8 l/min) during step 6 of the process described in section 4.3. Otherwise, some micro-balls could migrate with the solvent and position themselves on adjacent balls (Figure 4.38).

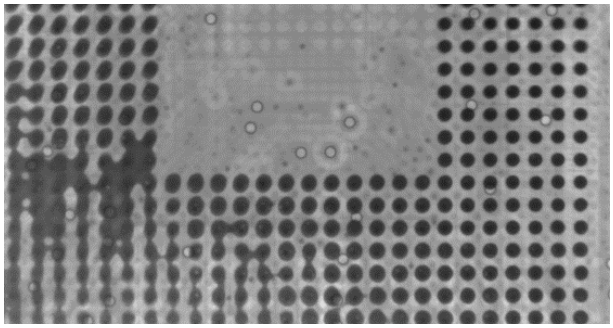


Figure 4.38: Ball Migration Observed at Radiography

Well-performed soldering must show soldering points during X-ray radiography with 90° incidence (Figure 4.39).

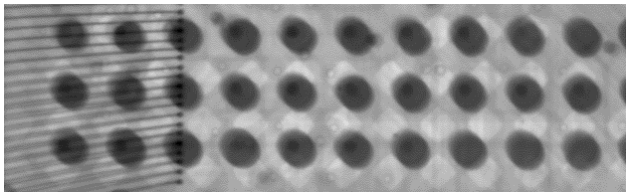


Figure 4.39: Soldering Performed Correctly: X-ray View Using 90° Incidence

To confirm good soldering, a second 45° incidence X-ray is necessary to establish the soldered joint (Figure 4.40). Radiography with such an incidence (called tilt radiography) should confirm uniform spheres.

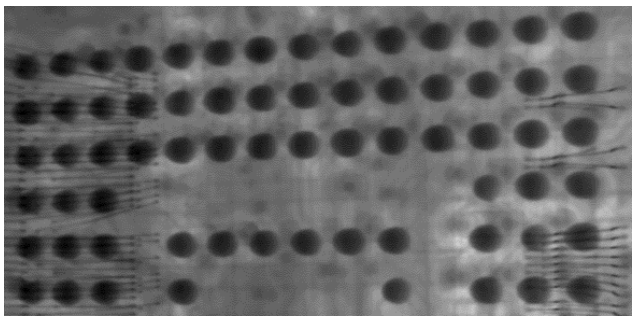


Figure 4.40: Well-Prepared Soldering Confirmed by Tilt Radiography (45° Incidence)

In the next tilt radiograph (Figure 4.43), we observe non-spherical consistency (the soldering is of poor quality) so we have a Chip/PCB misalignment.

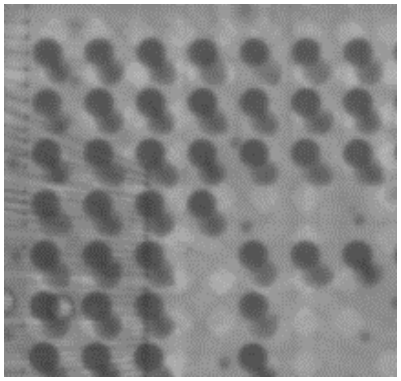


Figure 4.41: Incidence 90°

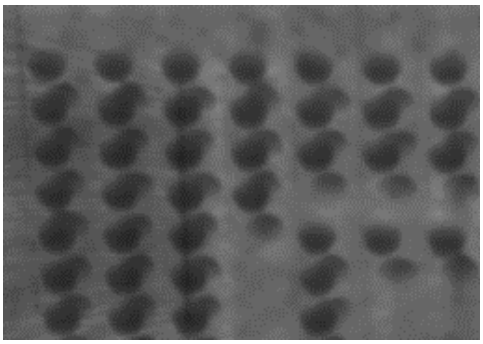


Figure 4.42: Incidence 45°

Figure 4.43: X-ray Views of Incorrectly Performed Soldering: Chip/PCB Misaligned

Note that non-uniform neighbouring beads can create migration beads. The non-flatness of the ball grid array is critical to achieve good soldering without internal migration that can lead to poor contact (Figure 4.44) and contact faults.

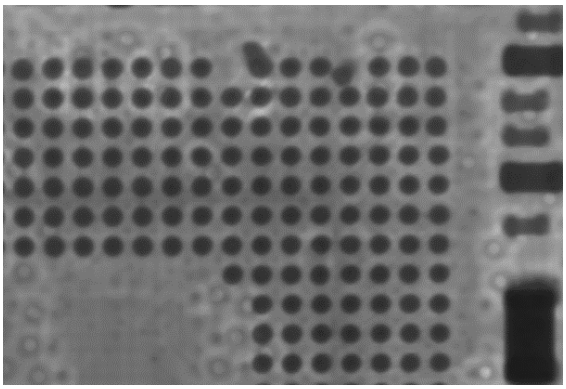


Figure 4.44: Migration of a Micro-Ball to the eighbouring beads: X-ray Visualisation

4.4.4 Conclusion of the Experiments

To verify and optimise the quality of our soldering technique, we conducted experiments on the chemical composition and physical appearance of the beads created while taking into account the slow and fast cooling rates. Rapid cooling improves the soldering, exploiting in particular the properties of resistance to mechanical stresses, and thus reduces the number of cracks. In our experiments, we did not see any trace of cracking when the balls are cooled rapidly, unlike some cracking when the balls are cooled slowly. The number of porosities (voids) is smaller with rapid cooling because it will quickly fix the Sn–Bi joint. The soldering will start on all of the host pads, thus reducing the PCB-to-chip distance by a levelling phenomenon.

Slow cooling promotes the presence of pores and tends to re-form the beads by detachment. Rapid cooling levelled the component to the PCB without the balls being able to regain shape. The solidification of the structures is different too.

In rapid cooling, we have a homogeneous structure with tin plaques and bismuth perfectly parallel to each other and homogeneous across the wafer. In slow cooling, solidification reveals an inhomogeneity on the wafer, with a coarser structure.

We have just seen a possible improvement in chip-off techniques. This technique will also be used again during the process of transplanting damaged and encrypted phones, that we will develop in chapter 7. We now present techniques that we realised during this thesis to directly repair damaged components.

5

Adhesives in Advanced Forensics Data Extraction

We have seen in chapter 3 that investigators have to perform an X-ray to diagnose the level of repair needed and thus select the correct extraction method. If the component or PCB is damaged, they must implement methods for repairing it. As we present in this chapter, traditional methods are rapidly reaching their limits with new generations of electronic components. Thus, in this chapter and in chapter 6, we have implemented new repair methods to push back the current limitations.

5.1 Context

Recent publication [Cui et al., 2014] underlines the interest of using polymers in microelectronics. Polymers are the ideal interconnect alternative to solder materials containing lead. Electrically Conductive Adhesives (ECAs) [Li and Wong, 2006], Thermally Conductive Adhesives (TCAs) [Felba, 2011] and UV Adhesives (UVAs) [Asif et al., 2005] mainly consist of a polymeric resin (epoxy, silicone, polyurethane or polyimide) that provides physical and mechanical properties such as adhesion and mechanical strength, and conduction of electricity (with the addition of metal fillers like silver, gold, nickel or copper) [Luo et al., 2016]. Currently, it is possible to find really cheap polymeric resin. Using these resins for digital forensic purposes is the focus of this chapter, which we demonstrate in a hardware reverse engineering prototype case study.

5.2 Material

5.2.1 Thermally Conductive Adhesives

This polymer family is designed to dissipate heat ([Felba, 2011] and [Falat et al., 2007]). Generally, this adhesive covers memory components and CPUs to reduce chip heating and improve sturdiness, which makes chip-off analyses seriously more difficult (Figure 5.1).

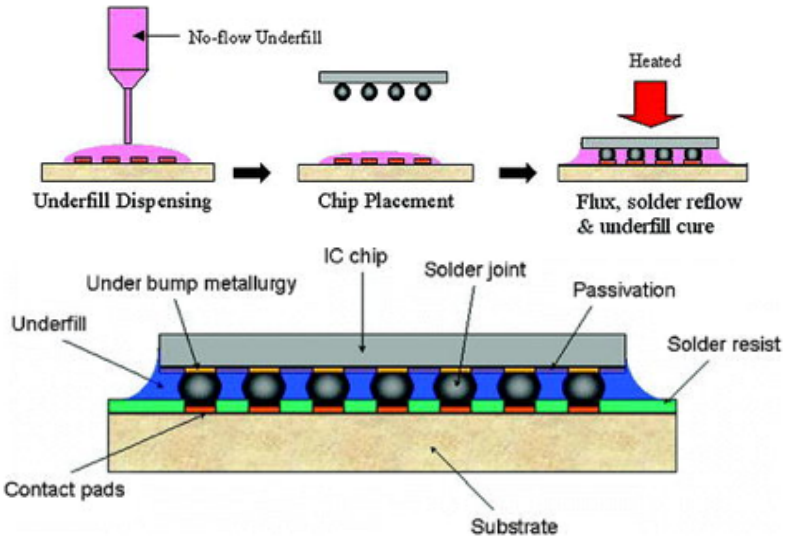


Figure 5.1: TCA Used for BGA Soldering With Board [Zhang et al., 2013]

The consequence of using a classical chip-off process is the need to increase the de-soldering temperature, which may seriously damage the chip and the board (Figure 5.3).

Figure 5.2: Good De-Soldering

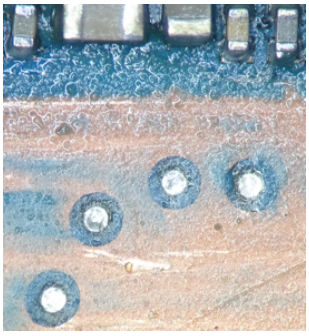


Figure 5.3: Bad De-Soldering Due to Underfill

Figure 5.4: Printed Circuit Boards After De-Soldering

The thermally conductive adhesive studied in this thesis (Polytec[1]) consists of two components: resin and hardener. The mixing ratio by weight is 100 resin units for 6 hardener units. Viscosity at 23 °C is 9000 mPa.s and minimum bond line cure time is 16 hours at 23 °C and 15 minutes at 100 °C (Figure 5.5).

[1] http://www.polytec-pt.com/int/products/

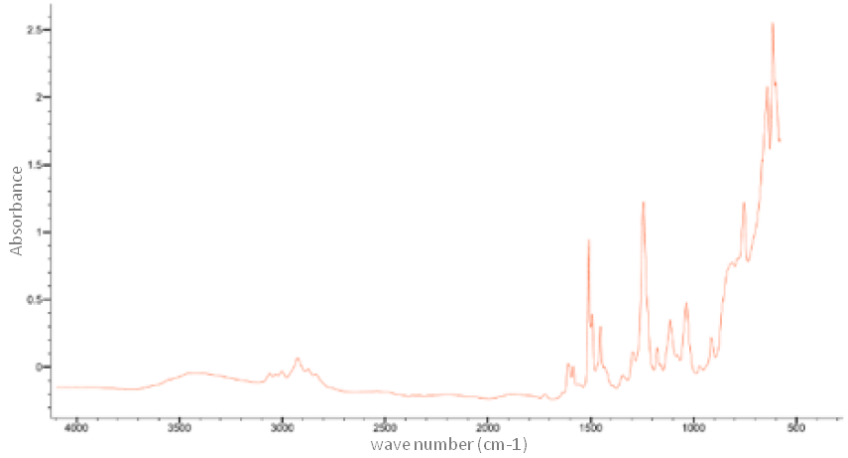


Figure 5.5: Thermal Adhesive Absorbance (Polytec)

Thus, this adhesive is very well-suited for the fixing, covering and coating of electronic parts. The solidification is slow, which allows for shaping of the fixation. It is possible to vary the solidification rate by modifying the temperature.

5.2.2 Underfill: A Special Type of Thermally Conductive Adhesive

5.2.2.1 What is an Underfill Epoxy?

In addition to traditional high/low temperature eutectic soldering [Heckmann et al., 2016], the underfill is systematically inserted in the manufacturing process between the electronic components and the PCB, to increase the shock resistance and the thermal resistance of the electronic components (Figure 5.6).

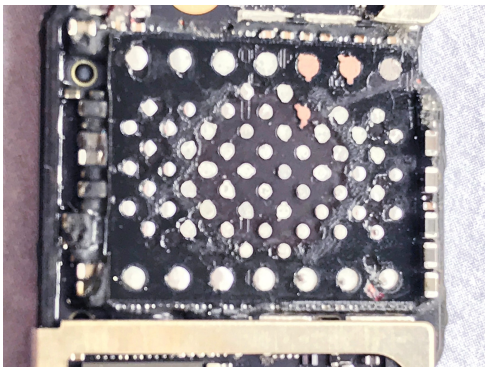


Figure 5.6: iPhone 7's Underfill Between Memory and PCB

Underfill has a coefficient of thermal expansion which is close to that of the eutectic solder alloy. Many papers have demonstrated an interest in the use of underfill [Mishiro et al., 2002] in increasing thermal and physical stress resistance (Figure 5.7,

[Okura et al., 2000]).

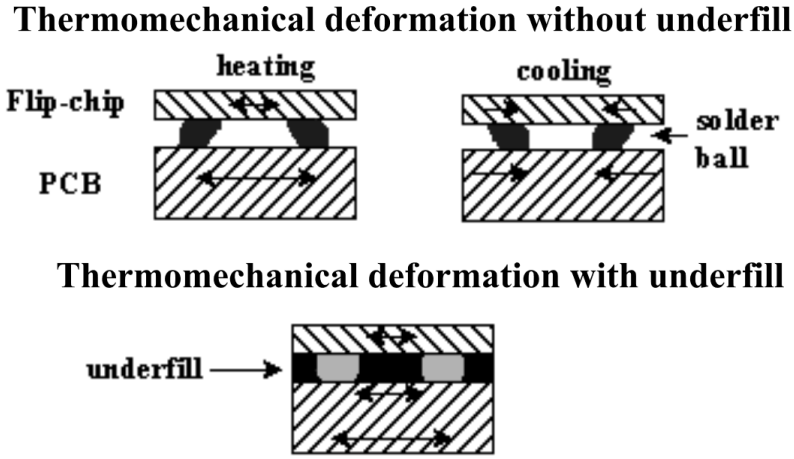


Figure 5.7: Thermomechanical Deformation Without and With Underfill

Some other papers [He et al., 2000] describe methods to characterise the fundamental properties using thermo-mechanical analysis, differential scanning calorimetry, and thermo-gravimetric analysis. There are several types of thermally conductive adhesives on the market, which vary according to several parameters, such as: the method of preparation (mono-component, two-component with resin and hardener), the coefficient of viscosity, the degradation temperature, the thermal conductivity, the Young's modulus values, the percentage of elongation at break, and the continuous resistance temperature.

Among all these different parameters, the temperature of degradation is the one that particularly interested the investigators. Indeed, in the context of forensic operations, the investigator must heat the adhesive and the electronic component at this temperature in order to remove the component and be able to perform its physical reading or forensic transplantation.

Thus, epoxy glues with a degradation temperature of less than 300 °C are characterised by forensic investigators as low-temperature glues. The ones above 300 °C are called high-temperature epoxy glues. To guarantee a greater robustness of assembly, manufacturers now fix the electronic components close to each other, thanks to this type of glue, which also increases the possibility of compactness of the latest generation phones (Figure 5.8).

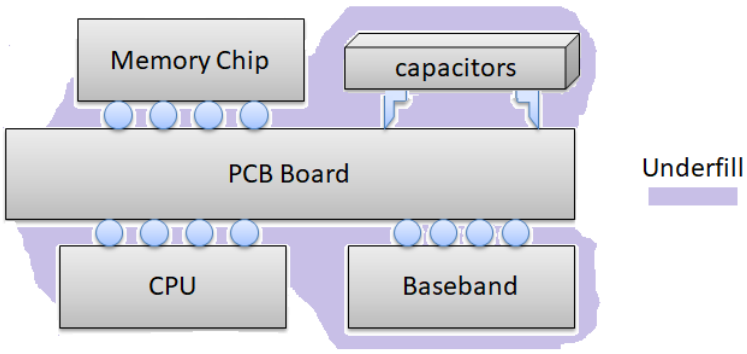


Figure 5.8: Underfill Modelling

In the case of the iPhone 7, the memory and the peripheral capacitors are fixed closely together by means of a low-temperature underfill (Figure 5.9).

Figure 5.9: iPhone 7 Memory and Neighbourhood Capacitors Glued on Board

It is the same on the opposite face of the electronic card: the processor, the baseband co-processor as well as the peripheral capacitors are also fixed together. This overall fastening makes the assembly robust to impact and greatly increases the resistance of the components to water.

The IP Code for International Protection Marking (IEC standard 60529) is an international standard of the International Electrotechnical Commission relating to waterproofing. The iPhone 7, 8 and X, the Samsung Xcover 3, Android Phone D6, BlackBerry Motion, Huawei Mate 10 Pro and Google Pixel 2 have an IP rating of 67. Thus, these mobile phones can be submerged in up to 1 metre of water for 30 minutes. The Samsung Galaxy Note 8, S7 and S7 Edge, LG X Venture, HTC U11 Plus, Motorola Moto X4, LG G6 and LG V30 have an IP rating of 68. Thus, these mobile phones can be submerged

in up to 1.5 metres of water for 30 minutes.

5.2.2.2 Underfill Problem for Forensic Investigations

Investigators must now confront the widespread use of underfill in the assembly of components on an electronic card. The realisation of the chip-off technique [Breeuwsma et al., 2007] on one of these components will cause an overall movement and thus the destruction of the electronic card (Figure 5.10).

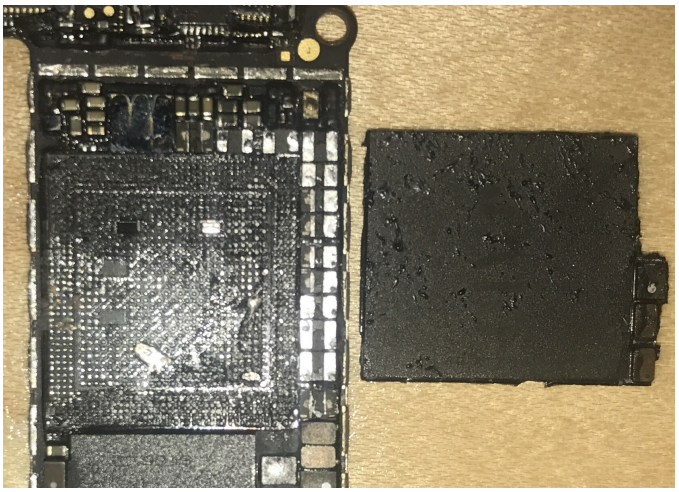


Figure 5.10: Destruction of the Neighbourhood Capacitors Due to Underfill on iPhone 7 CPU After Classical Chip-Off

This overall movement is due to the fact that to remove the component, the investigator does not get to the degradation temperature of the glue as well as the soldering materials (BGA ball). When the investigator exerts a perpendicular mechanical force to remove the component, as is the case with the conventional chip-off method, all the components that are poured into the same block of underfill will be set in motion. These movements will create a short circuit of the other components and thus destroy the electronic card (Figure 5.11). Thus, techniques that were quite well-mastered until now are no longer applicable to current generations of mobile phones, even in cases in which the investigator must make a simple change of one of the electronic components.

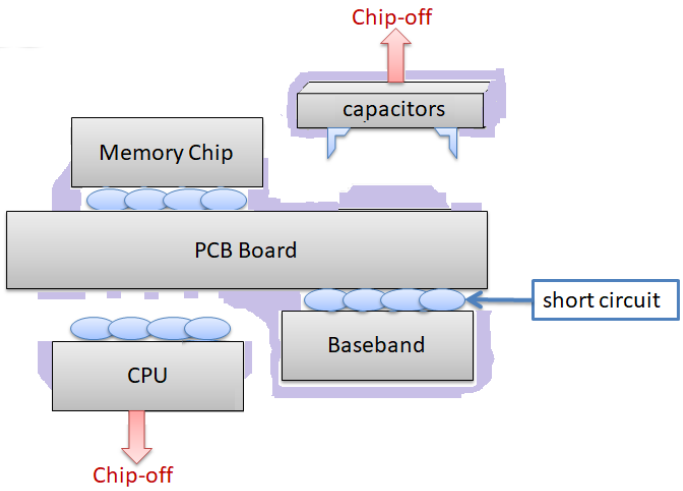


Figure 5.11: Destruction of the Neighbourhood Capacitors After Classical Chip-Off

To do this, the investigator usually proceeds to replace the defective components to make the phone operational again. But basic replacement, which has been a perfectly mastered technique, becomes inapplicable on underfilled phones.

5.2.3 UV-Curable Adhesives

The UVA we studied (Polytec UV[2]) is a single liquid acrylic adhesive component curable by UV light (Figure 5.12).

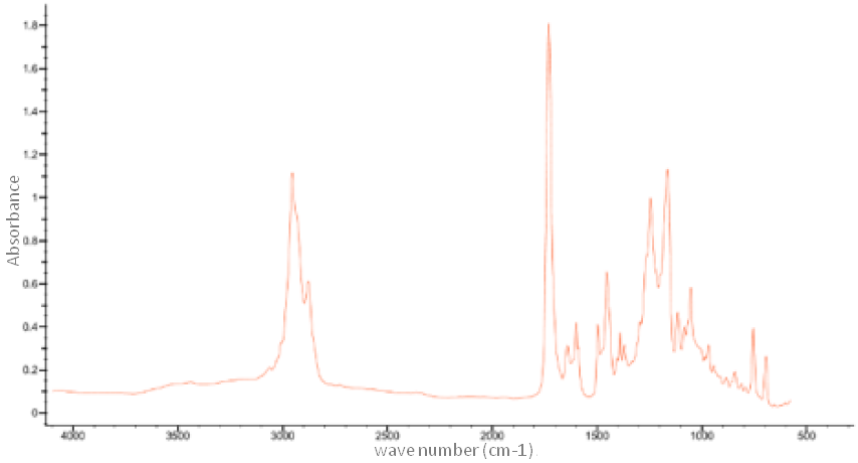


Figure 5.12: UV-Adhesive Absorbance (Polytec)

UV-curable adhesives ([Asif et al., 2005], [Kim et al., 2002], and [Zhang et al., 2014]) can be categorised by their mechanical properties, viscosity, hardness and the desired

[2] http://www.polytec-pt.com/int/products/uv-and-dual-curing-adhesives/

adhesive strength on selected substrates. Thus, reading the properties is essential in order to choose the right feature for the desired effect. Viscosity is 900 mPa.s and this adhesive will cure within 45 seconds upon exposure to UVA light at 320-400 nm (Blacklight). UV-curable adhesives are very well-suited for instantly fixing of electrical wires.

5.2.4 Electrically Conductive Adhesives

Research efforts [Cui et al., 2013, Li and Wong, 2006, Moon et al., 2004] have focused on two lead-free alternatives, lead-free metal [Heckmann et al., 2016] solder alloys and polymer based ECAs [Rechchach, 2011].

ECAs consist of an organic/polymeric binder matrix and metal filler (Figure 5.13). ECAs are the ideal interconnect alternative to lead-based solder materials. The conductive fillers provide the electrical properties and the polymeric matrix provides the physical and mechanical features [Kishi et al., 2016].

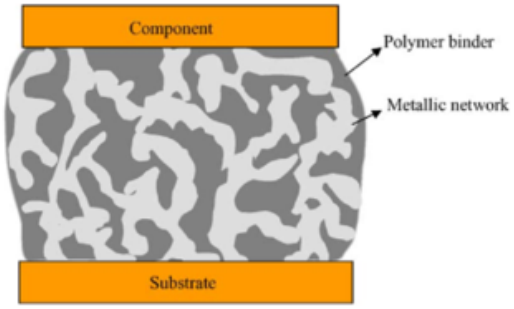


Figure 5.13: Ball Created by an Assembly of a Polymeric Binder Matrix and Metal Filler [Li and Wong, 2006]

The ECA that we studied (Polytec[3]) is an electrically conductive, mechanically stable and flexible polyurethane dispersion. It is suggested for electrically conductive bonding and coating applications.

Polyurethane combines a very high degree of flexibility with good mechanical stability. The melting point is 250 °C and the polyurethane's softening range is from 180 °C. Drying time at 23 °C is 8 minutes and viscosity is 5000 mPa.s. Thus, this adhesive is very well suited for bonding when a quick attachment is necessary.

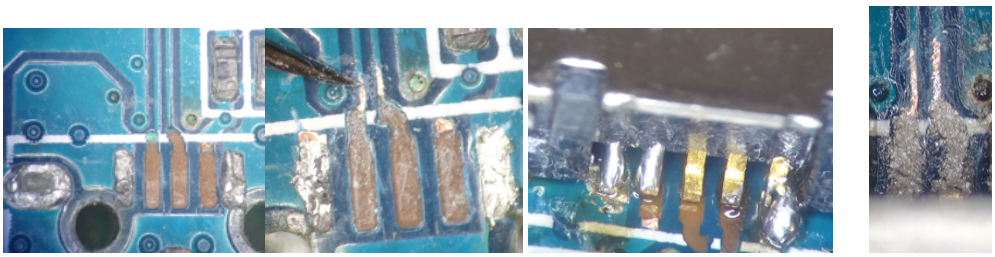
[3] http://www.polytec-pt.com/int/products/epoxy-adhesives/electrically-conductive-adhesives/

5.3 Digital Forensic Applications and Developed Methods

As indicated in the previous sections, the adhesives have various and multiple properties. The forensic application of polymer-based adhesives is as wide as the choice of adhesives. However, we will only focus on three main challenges in forensic data extraction: electronic circuit repair, low temperature reworking, and prototyping.

5.3.1 Restoring Conductivity

Damaged exhibits (received broken, or damaged during the extraction process) may require restoration of conductivity of data lines, control lines or power lines. Conductive adhesives are suitable to obtain such a result. A basic approach consists in identifying the broken part (Figure 5.14a), finding the remaining parts of the broken line(s) (Figure 5.14b), preparing the parts to be connected (Figure 5.14c) and joining them using a suitable conductive adhesive mixture (Figure 5.14d).



(a) Damaged PCB pads (b) Recover Lines (c) Prepare Pads and Connector (d) Restore Connectivity

Figure 5.14: Damaged Micro-USB Connector Repair

While most repairs require only a rough repair using a basic conductive adhesive, some may require quite fine techniques. Some specific conductive adhesives (such as Polytec[4], Aremco[5] or Loctite[6]) may also be used to repair bonding wires.

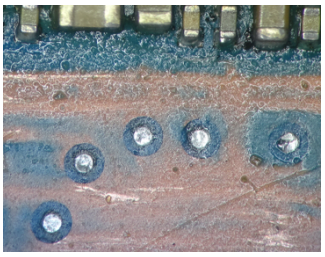
In this case, we are able to communicate with the silicon chip by creating a conductive bridge between the two parts of the broken bonding wires.

[4] http://www.polytec-pt.com/int/products/epoxy-adhesives/electrically-conductive-adhesives/
[5] http://www.aremco.com/
[6] http://na.henkel-adhesives.com/industrial/equipment-solutions-19440.htm

5.3.2 Restoring Insulation

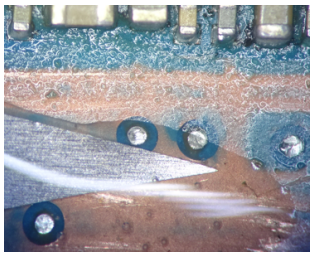
Fixing the PBC may also be necessary to replace the damaged insulation layers. This is particularly suitable when a chip is over-lapped or when the PCB underlayer is accidentally removed while de-soldering an underfilled chip.

To restore the insulation of a damaged board, it is first necessary to clean the damaged area using a flux remover (Figure 5.15a). The prepared insulating glue bi-phase mixture (Figure 5.15b) applied in its viscosity state (Figure 5.15c) will then create a thin insulation layer (Figure 5.15d). Once dry, the excess insulating adhesive can be removed using a micro-knife (Figure 5.15e) to reveal the PCB connection (Figure 5.15f). Finally, filling in the hole with some electrically conductive adhesive (Figure 5.15g) permits recreating the pad over the insulation layer (Figure 5.15h).

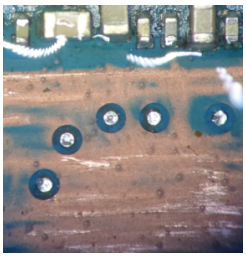


(a) Cleaned PCB

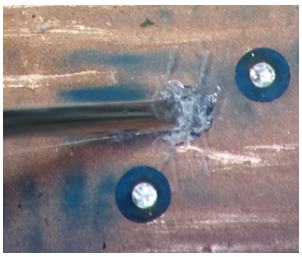
(b) Prepared Insulating Bi-Phase Glue Mixture



(c) Insulating Adhesive Being Applied

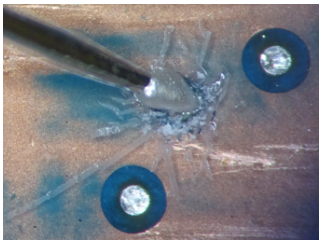


(d) Thin Insulation Layer

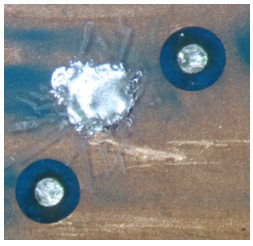


(e) Micro-Knife Removing Insulation Layer

(f) PCB Pad Revealed



(g) Electrical Conductive Adhesive Applied



(h) Restored Connectivity

Figure 5.15: Pads Restored

As demonstrated, ECA can be used to recreate a very thin insulation layer. Thermal resistance of the polymer mixture is a crucial parameter if the related surface has to be

reheated for reworking purposes.

5.3.3 Reworking and Prototyping

Reworking without damaging the evidence is the real forensic data extraction challenge. It is sometimes necessary when the only way to extract the data is to replace a de-soldered component (e.g. memory) back onto the donor board or its original one. To prevent overheating, recent articles [Jongh, 2014, Heckmann et al., 2016] propose different ways to use bismuth–tin alloy in order to reduce heating within the re-soldering process at only 150 °C. Although such low temperature profiles can save most of the chip from overheating damage, it may not prevent some highly sensitive or heat-protected memory chip from being damaged. Therefore, ECA can be used to re-solder a chip at room temperature (Figure 5.16). The challenge here is to reball the chip prior to chip-on. A dedicated pump, such as Nordson Performus II, could also be used for this purpose as it allows to carefully choose time/pressure, key parameters to create small, round balls ready to be fixed long-term.

Figure 5.16: Nordson Performus II Pump

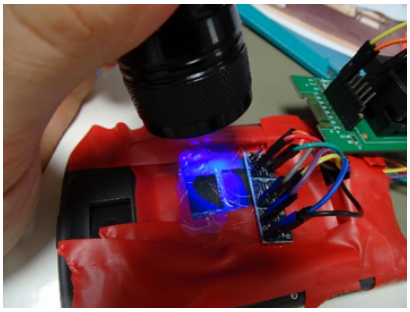


Figure 5.17: Thesis' Memory Man-In-the-Middle Prototype Being Prepared

Figure 5.18: Adhesives in Reworking and Prototyping

Prototyping is often required within digital forensic processes, either to create extraction environments (donor board, attack board), or to understand exhibit modus operandi (reverse engineering, Figure 5.17). Such prototypes require expensive on-demand design boards or adapters. Polymer-based adhesives' properties can also be used as a cornerstone for homemade prototypes, as demonstrated in use in the next section. In fact, we have shown in this thesis that the adhesives are suitable to fix, connect or isolate parts of any system in a very flexible and affordable way.

5.4 Memory Man-In-The-Middle Attack

This case study is related to the need to understand the security mechanism of a smartphone, which was accessed by prototyping a memory man-in-the-middle (MIM) attack. The aim of the prototype is to capture/analyse exchanges between the smartphone's processor and its non-volatile memory (eMMC) in order to be able to easily manipulate (read/write) the memory data.

5.4.1 eMMC: Embedded Multimedia Card

We first present the memory component that we used to make our MIM attack: the eMMC. In addition, the investigator most often finds the eMMC during the extractions on mobile phones. Naturally, there are other components, but we chose this one to carry out our studies because the investigator will find it in use 90% of the time. The whole study carried out in this chapter is easy to generalise to other electronic memory component sets.

The eMMC is therefore often the element that the investigator must be able to resolder at low temperature. The data found in this memory is very fragile and sensitive to temperatures above 160 °C. An eMMC is an advanced, managed NAND flash memory (Section 2.2.2) for mobile applications (smartphone, tablets, GPS).

The eMMC is not only a NAND flash memory but also a controller/interface circuit (Figure 5.19, [JEDEC, 2010]). Electronic components need to be electrically connected for power, ground and signal transmissions.

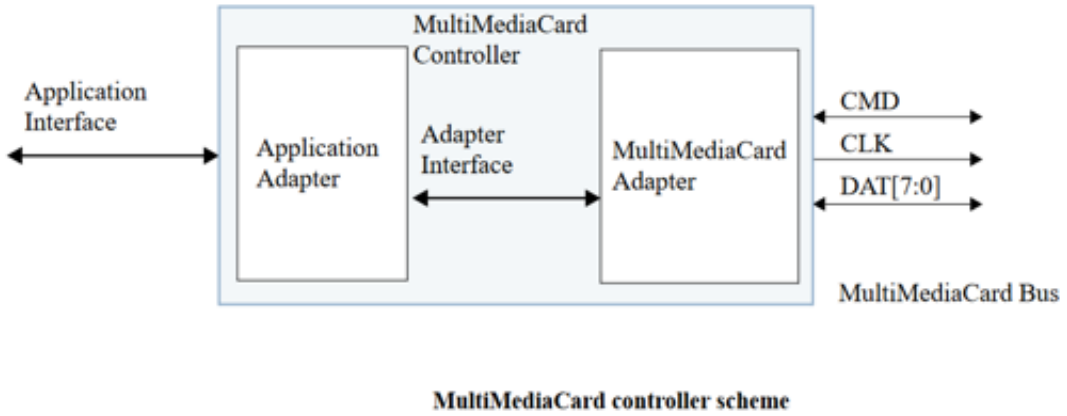
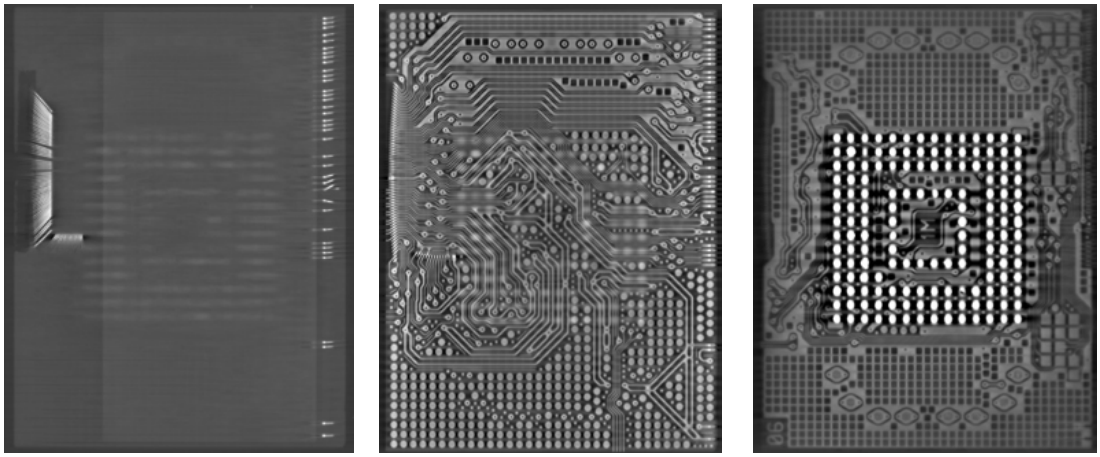


Figure 5.19: eMMC Controller Scheme [JEDEC, 2010]

These interconnection technologies can be: Pin Through Hole (PTH), Surface Mount

Technology (SMT), Ball Grid Array (BGA) (Figure 5.20), Chip Scale Package (CSP) and Flip Chip Technology (FCT) (Section 2.3). Signals are transmitted through BGA balls located between the component and the board.



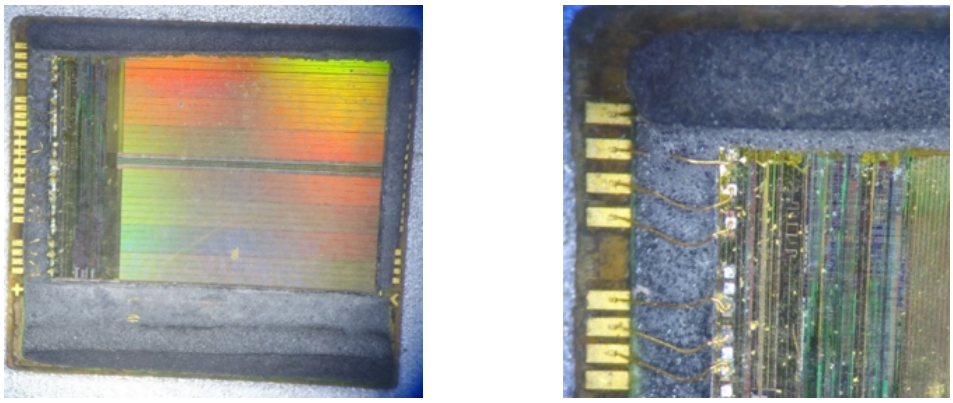
(a) Connections Between the Silicon Dies at BGA Level

(b) Copper Plate Level

(c) Ball Grid Array level

Figure 5.20: IRCGN's X-ray 3D Micro-Tomography of the H9DP4GG4JJMCGR Hynix eMMC

Then the bonding wires are used to make the junction between the BGA balls and silicon dies (Figure 5.21). Wire bonding can be attached to silicon in two different ways. The first is the Ball Bonding method and the second is the Wedge Bonding method.



(a) Memory Dies Exposed

(b) Bonding Wires Exposed

Figure 5.21: eMMC Decapsulated

Ball bonding is associated with thermocompression and thermosonic joining and the wedge bonding process uses ultrasound and pressure to create a bond between the wire and the bond pad.

The eMMC, contrary to UFS[7], has a parallel interface (half-duplex) which means it can only send data in one direction at a time: it can either be read or write. UFS has a Low-Voltage Differential Signaling (LVDS) serial interface which has separately dedicated read/write paths. This allows full two-way interaction: UFS can be read and written simultaneously (Figure 5.22).

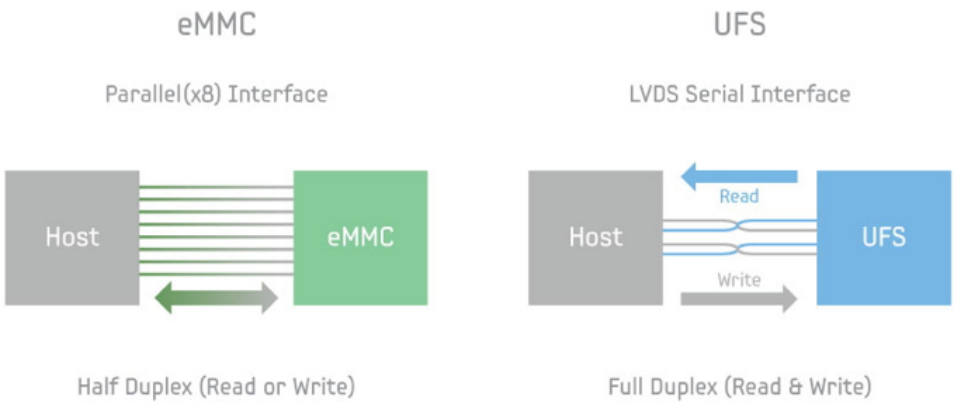


Figure 5.22: eMMC UFS Interface Comparison [Samsung, 2015]

For an eMMC, as described in the JEDEC Bus protocol, after a power-on reset, the host must initialise the card by a special message-based MultiMediaCard bus protocol. For each data line, the data can be transferred at the rate of one bit SDR (single data rate) or two bits DDR (dual data rate) per clock cycle.

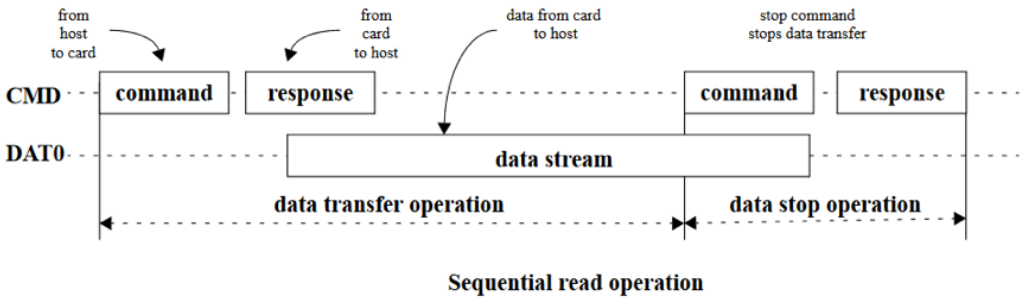


Figure 5.23: Sequential Read Operation: 1 Bit Data Bus

[7] http://www.jedec.org/sites/default/files/docs/JESD220C.pdf

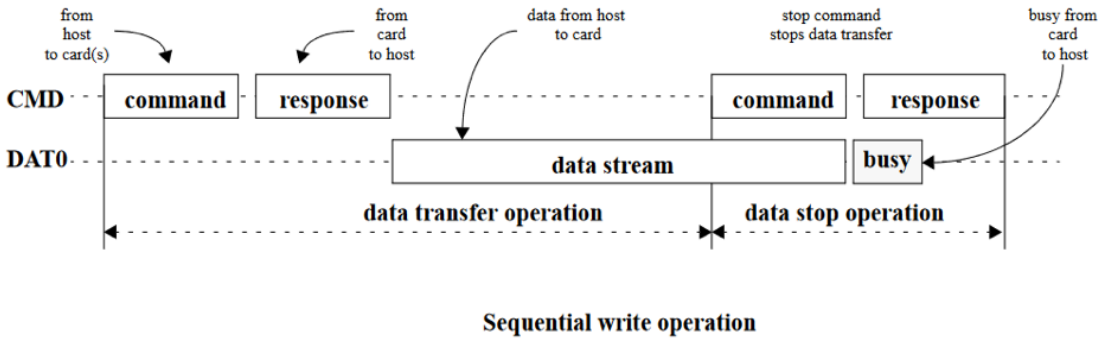


Figure 5.24: Sequential Write Operation: 1 Bit Data Bus

The data bus is a flexible data bus of 1, 4 or 8 bits. After power-up (or hardware reset), only D0 is used for data transfers, but the host can configure the device to use a wider data bus, D0, D0-D3 or D0-D7 for data transfer.

Our chip is packaged on a 153-ball BGA with 0.5 mm pitch, 0.3 mm diameter ball, lead- and halogen-free (Figure 5.25).

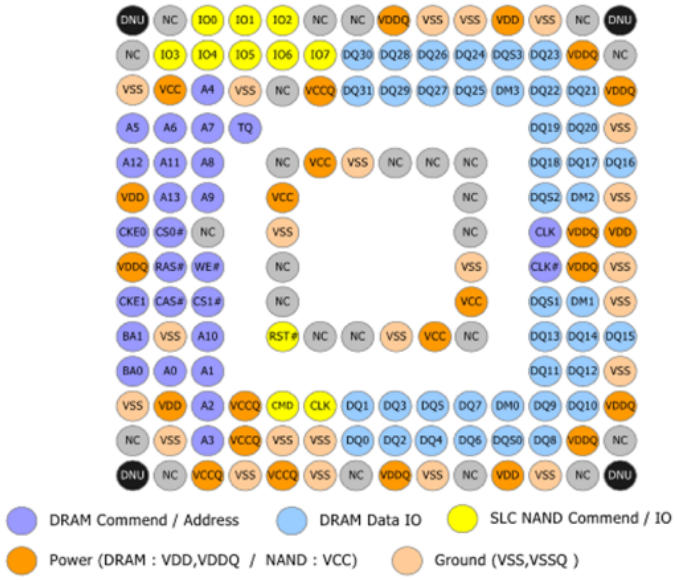


Figure 5.25: Mapping Schematic Balls of the H9DP4GG4JJMCGR Hynix eMMC

We noticed, by performing reverse engineering under tomography control, that even if the data sheet indicates common bead values (VSS, VDD) they may not be interconnected on chip (different GND domains). Thus on some models of eMMC, the internal VSS are not connected to the external VSS. In order to write on the eMMC, it is essential to link the internal and external VSS to the ground to avoid creating an electrical charge that can destroy the chip.

We now present the manufacturing steps of our prototype which allows us to perform our MIM attack.

5.4.2 Steps of the Prototype

Our study focuses on the steps of a new MIM attack on the eMMC, but these steps can be used for other BGA-components (CPU or crypto-chips). We illustrate here new capacities of adhesives. This platform aims to provide (reading/writing) access to data exchanged between the phone controller and its non-volatile memory (Figure 5.26).

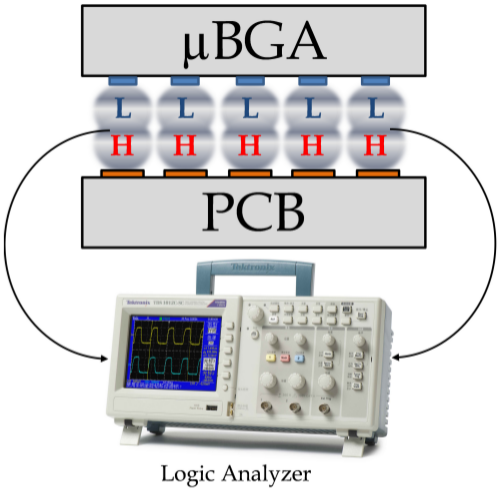


Figure 5.26: Our MIM Attack Using Two Levels of BGA: (L) Low-temperature, (H) High-temperature

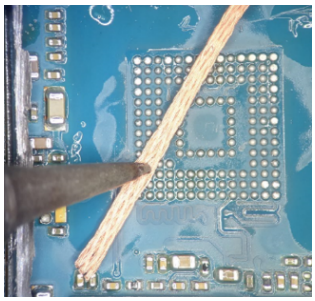
The prototype we have developed enables us to interface several types of devices (logic analyser, FPGA, oscilloscope, injection computer, etc.). The choice of the hardware interface depends on the investigators' needs during their reverse engineering operations. In this case, we first present an interface with a computer for analysis, making it possible to make the injection and modification of the data in the memory. In a second step, we perform the interface with a logic analyser in order to characterise the real-time exchanges between the CPU and the memory.

The re-working process we used was in several steps, as follows:

Step 1: Cleaning the receiving PCB, using a soldering iron tip to clean it (Figure 5.27a).

Step 2: High-temperature reballing (250 °C) of the PCB except the balls we want to observe (CMD, CLK, VSS, VCC, VCCQ (DRAM), VDD (DRAM) and D0) (Figure

5.27b).



(a) Step 1: Cleaning the Receiver PCB

(b) Step 2: High Temperature Reballing

Step 3: Depositing electrically conductive adhesive using a micro-instrument on the balls we want to observe (Figure 5.28).

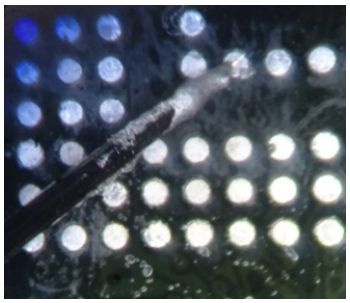
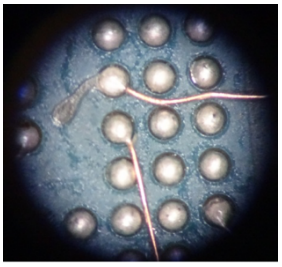
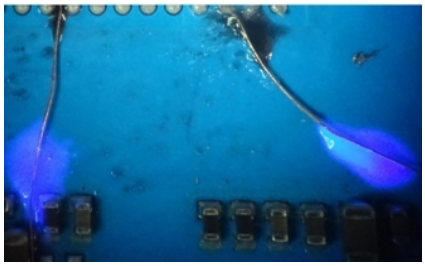


Figure 5.28: Step 3: Depositing ECA

Step 4: Attaching wires to the ECA (Figure 5.29a). The wires we use have a cross section of 30 µm and are self-insulated copper (AMETEK ECP). We use ECA instead of high-temperature solder balls because of the viscosity that allows a more easy insertion of the wires inside the ball (without creating deformation). It may be possible to directly use a PCB flex, but the precision is limited for CPU attacks whose balls are very close. Our low-cost technique can be executed quickly on all components bypassing this difficulty and without design phase.



(a) Step 4: Attaching Isolated Wires

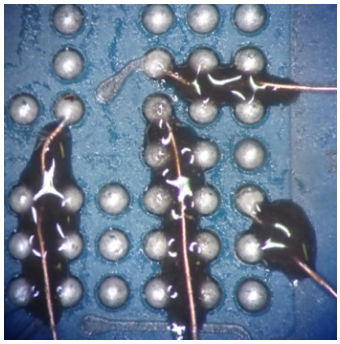


(b) Step 5: Instantly Fixing Wires

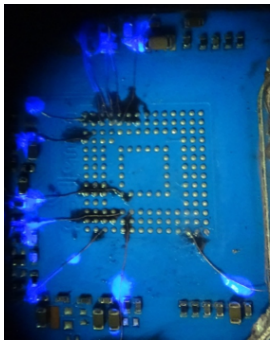
Step 5: Fixing wires with UV adhesive (Figure 5.29b). The goal here is to prevent the wires from interfering when positioning the eMMC. We note that UV adhesives cannot

withstand the re-soldering temperature, so the only goal is to make instant and mouldable fixing wires. This adhesive will allow insertion of the second adhesive (step 6), which will withstand the reflow temperature.

Step 6: Fixing wires between the balls with TCA (Figure 5.30). Here the setting rate is longer than UVA's (step 5) but thermal conductive adhesive will withstand the temperature.



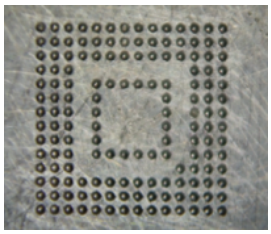
(a) Fixing Wires With TCA



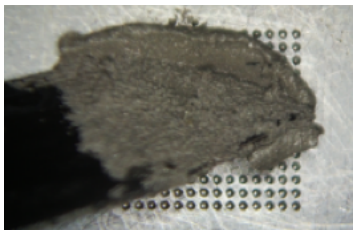
(b) Global View

Figure 5.30: Step 6: Fixing Wires With TCA and UVA

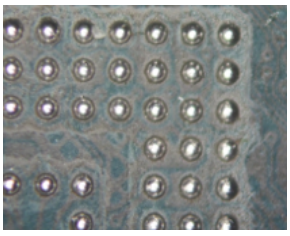
Step 7: eMMC Low temperature reballing using the technique in chapter 4 (Figure 5.31).



(a) Stencil Positioning



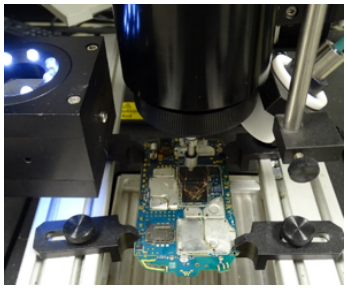
(b) 42Sn/58Bi Paste Being Applied



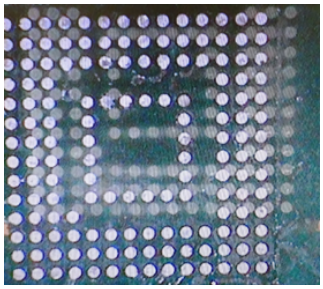
(c) Balls Agglomerated

Figure 5.31: Step 7: Low Temperature Re-Balling Technique

Step 8: Positioning the eMMC on the PCB through the soldering BGA station. At this stage high temperature balls now interface with low temperature ones. This follows the application of the low temperature curve re-soldering (Figure 5.32).



(a) BGA Station



(b) Positioning the eMMC Beads

Figure 5.32: Step 8: Using the Soldering BGA Station

Step 9: Checking the final position by 2D X-ray (Figure 7.18).

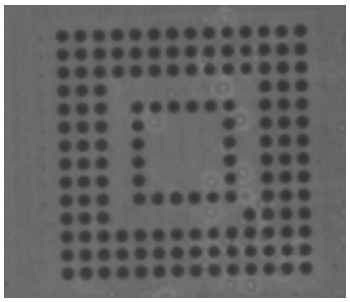
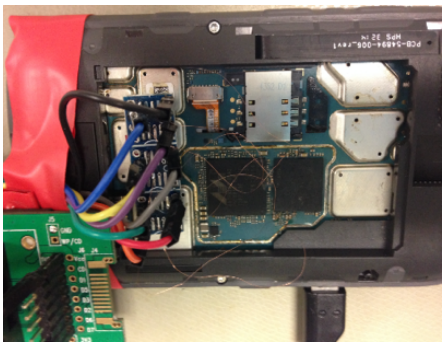


Figure 5.33: Step 9: X-ray to Check the Position

Step 10: Restarting the reassembled phone (Figure 5.34) and manipulating the data.



(a) SD Bus Connection

(b) Pin-out Dies

Figure 5.34: Step 10: General View of the Prototype

The objective is the reading of the eMMC, re-injection of the modified data in the eMMC, restarting the phone, etc. It should be noted that real-time editing of data (e.g. data modifications via script) would not be by connecting low and high temperature balls, but by connecting each of the components to an intrusive piece of electronic equipment (FPGA).

5.4.3 Reading Phase in Forensic Conditions

To read the eMMC we connect our cables (Io, VDD, etc.) to an SD adapter. This adapter is connected to a write blocker (forensic mode) as shown in Figure 5.35.

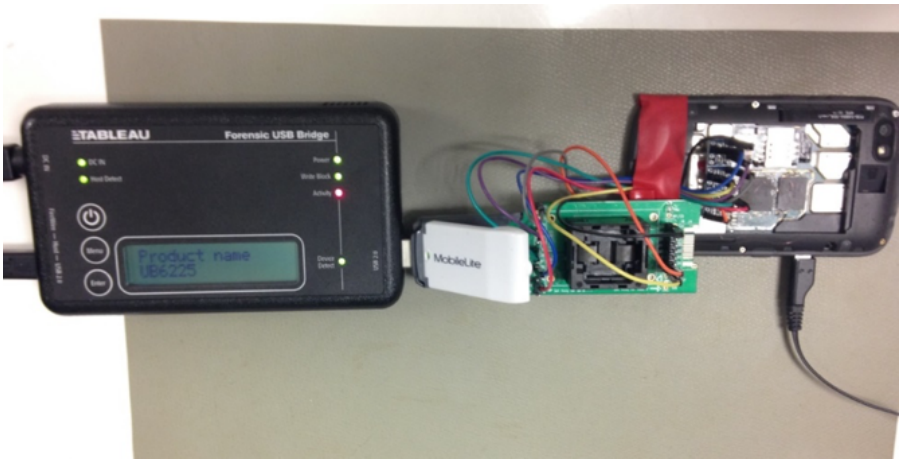


Figure 5.35: Reading Phase Process: BlackBerry 9790 PGP

We connect the blocker to the computer via USB and use analysis software in order to extract the memory. Note that we need to power the phone via its USB port as the power provided to the eMMC via SD adapter is not enough (loss in the PCB). At the end of this step, it is possible to turn on the phone and to make changes on all the internal elements related to embedded cryptography. For example, the reading phase takes about 2 minutes for an 8GB eMMC BlackBerry 9790 memory.

Subsequently, it is possible to perform a complete readback of the memory almost simultaneously. This new reverse-engineering method is a considerable time-saver for the investigator who, before this system, had to re-ball and solder the components between each reading. In addition, we have shown in the previous chapter that the temperature applied during the soldering operations had an impact on the memory's operation. Thus, our MIM attack aims to also preserve the integrity of memory cells without thermal stress.

5.4.4 Injection Phase

In the previous step, the investigator had the possibility to modify internal phone data (cryptographic application, password, etc.). The investigator (performing the reverse engineering) usually needs to directly modify the binary data of the memory and watch the influence of its modifications when the phone re-starts. Therefore, to inject the new image in the eMMC we do the same mounting as above but without the write blocker (Figure 5.36). The electrical power to the phone should always be via USB. The

injection has now been performed correctly. For example, the injection phase lasts less than 2 minutes for an 8 GB eMMC memory and only a few seconds if the change does not affect the entire memory (a few bytes).

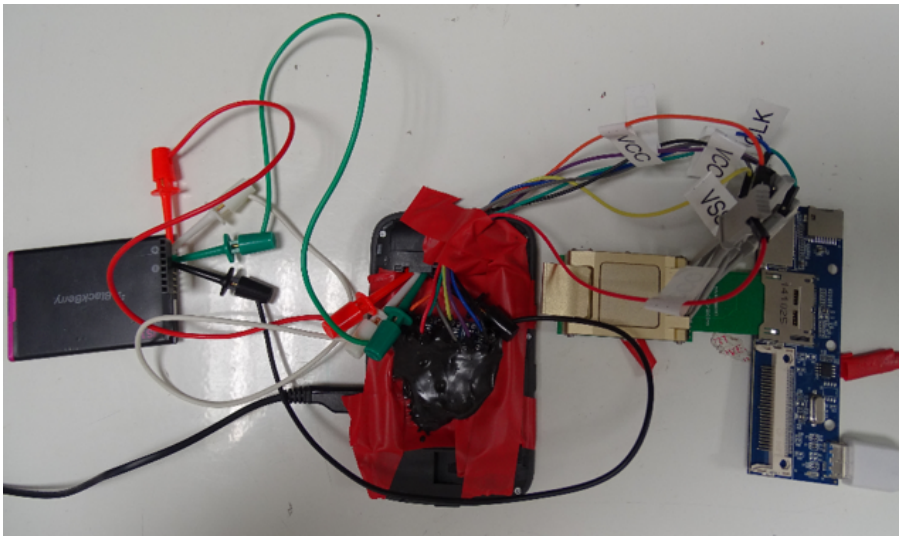


Figure 5.36: Reinjection Phase in the eMMC: BlackBerry 9790 PGP

The complete reverse-engineering of the system can therefore be achieved by combining the reading and injection phases while being able to see the influence of the changes when the phone is switched on. Since the component no longer needs to be unsoldered, the operations can be successfully carried out several hundred times without compromising the system integrity.

5.4.5 Tracking Signals Using a Logic Analyser or an FPGA

The two previous steps use a reading and writing device with the MIM interface. But we can also interface an FPGA or a logic analyser in order to have real-time crypto-information. As shown in the photo in Figure 5.37, techniques using adhesives can enable us to spy on the exchange of information between the controller and the eMMC in real time. In our example, we listened only to IO_0 because only this cable is fixed on by adhesive.

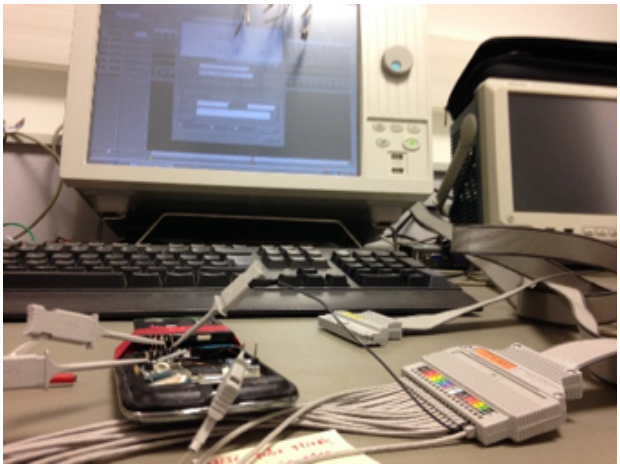


Figure 5.37: Tracking Signals Using a Logic Analyser or an FPGA

We can achieve this by listening on all IO and also performing this operation on the CPU or RAM to characterise all real-time exchanges. This global method gives us complementary information to that of the simple memory data modifications. We have therefore set up in this thesis several means of characterisation of the security system by changing the device according to the needs of the investigator.

5.4.6 Conclusion of the Experiments

Laser attacks, electric probing (chip-on), soldering, and re-soldering attacks (chip off) are very powerful level 4 methods. However, they quickly become time-consuming when the forensic investigator needs to make numerous changes of encrypted data or to make several change injections to characterise the secure system. These two methods are therefore preferred in the routine.

This study is twofold. It first introduces the multiple properties of adhesives (electrically conductive, thermally conductive, and electrically insulating curable by UV) and their concrete possibility to solve significant and long-standing hardware forensic challenges. Then the study successfully applies the adhesive capabilities to propose a memory man-in-the-middle platform. This platform is intended to provide advanced access to the inspected devices: easy access to the memory content (reading/writing), but also data exchange live analysis and manipulation.

This method is currently used by the IRCGN experts for the reverse-engineering of secure systems. This method has the advantage of being able to be applied on any type of BGA components and its implementation becomes trivial if the technique is mastered. The cost of setting up the attack is actually reduced to the purchase of solder paste, conductive and insulating glues, micrometric connection wire and BGA mask. It should

be noted that it is the frequency and the number of required channels observed that will drive up the cost. The final cost will be dependent on the interface equipment: FPGA, oscilloscope, spectrum analyser, logic analyser, real-time processing unit, etc.

In chapter 6, we will discuss methods dealing more with repairing components damaged as a result of disasters (bombing, air crash).

6

Laser Attacks on Pigmented ECA

Polymeric adhesives are of interest in the digital forensics domain. They can be used to perform more or less complex repairs or even to realise advanced man-in-the-middle attacks in order to carry out reverse engineering of secure systems (chapter 5). The main aim of this chapter is to develop a technique that makes polymeric adhesives sensitive to laser decapsulation attacks while decreasing the laser power deposited during ablation. Indeed, when the investigator has a mobile phone that has suffered a shock, bonding wires are generally the first impacted. These must be made operational again in order to proceed with the phone's analysis. However, current techniques are very difficult to implement for the repair of bonding wires without causing irreversible damage to the component.

The aim of this work is to develop a protocol that can be used for the micro-repair of broken bonding wires in areas where traditional techniques using the wire-bonder are not applicable or are likely to create additional damage to neighbouring bonding wires. This chapter shows how to make conductive bonding, using Electrically Conductive Adhesive (ECA), with an accuracy of 15 micrometres. This chapter led to the deposit of two industrial patents.

6.1 Context

As a rule, Electrically Conductive Adhesives (ECAs) are not very sensitive to laser decapping attacks. For it to have an effect on the ECA, it is necessary to strongly increase the power of the laser in order to increase the amount of absorbed energy. However, an excessive increase of this energy deposit can have devastating effects on the memory component (destruction of the bonding wires, radiation damage and interactions on the silicon chip). In some forensics cases encountered, the investigator may have to resort to laser or chemical decapsulation [Staller, 2010] in order to proceed with the repair of the broken bonding wires, reverse engineering, or chip-on method to read the memory. This laser/chemical decapsulation can create the risk of destroying bonding wires (Figure 6.1).

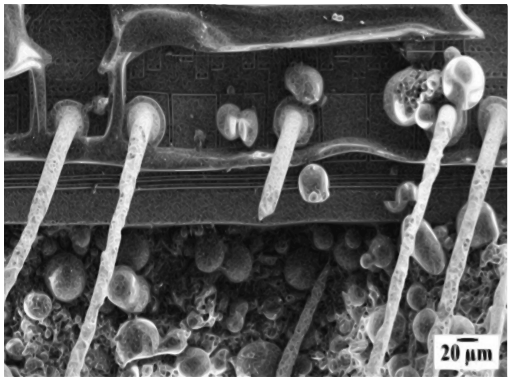


Figure 6.1: SEM Images of Broken Copper Wires After Complete Decapsulation (Laser and Chemical) [Kor et al., 2014]

This damage can occur in three ways. The first is when the laser or chemical is used on the package to expose the bonding wires (Figure 6.2).

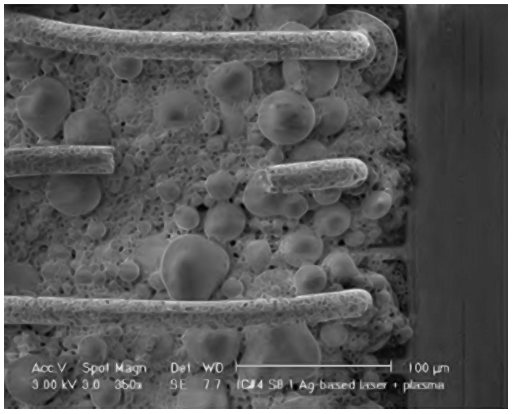


Figure 6.2: SEM Image of Attacked and Broken Wire [Kerisit et al., 2014a]

The second is when the investigator manipulates the probes to contact the bonding wires for the realisation of a forensic dump (Figure 6.3).

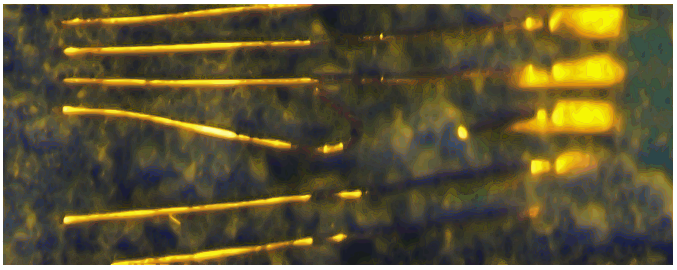


Figure 6.3: Image of Broken Wire After Probe Manipulation

The third is destruction resulting from a terrorist attack, an air crash, a violent car/bus accident, defenestration, a deliberate destruction act by the perpetrator of a crime during arrest, brawling, fire, etc.

The repair of the bonding wires is therefore a very complex and indispensable step which entails a risk of further destruction. As we saw in chapter 3, component repair is part of the investigator's work. The repair of the damaged memory component allows the investigator to return to the stage of a simple undamaged mobile phone analysis without having to do the micro-reading technique (Figure 3.35). For forensic day to day cases, the micro-reading (using FIB and SEM) is a long, expensive and one-time technique, which is why the investigators reserve it for operations of reverse engineering.

6.2 Purpose of the Developed Technique

The main aim of this section is to develop a technique allowing us to vary the absorbance of our target matter (polymeric adhesives) which will have an immediate and fundamental influence on the equation of heat. To vary the absorbance, we can experiment on two parameters, as shown in chapter 3 and in equations 6.1, 6.2 and 6.3.

$$\rho(T)C_p(T)\frac{\partial T(x,y,z,t)}{\partial t} - \nabla(\kappa(T)\nabla(T(x,y,z,t)) = Q(x,y,z,t) \tag{6.1}$$

$$Q(x,y,z,t) = (1-R)I(x,y,t)\frac{A}{L}e^{-\frac{A}{L}z} \tag{6.2}$$

$$A = \sum_{i=1}^{n} A_i(\varepsilon_\lambda, i, l = 1cm, c_i) \tag{6.3}$$

First parameter: we can experiment with the concentrations of the mixture components by adding pigments or dyes. With this first factor of influence, we will vary the absorbance of the ECA by adding either dyes or pigments of different concentrations. The goal is to make the ECA sensitive to laser decapping attacks and decrease the laser power deposited.

Second parameter: we can also experiment on the selection of pigments or dyes with a high molar extinction coefficient at the wavelength $\lambda = 1064$ nm (corresponding to the wavelength of our ablation laser). For this second factor of influence, we will select the pigments or dyes with the highest molar extinction coefficients at the wavelength $\lambda = 1064$ nm.

6.2.1 Materials

6.2.1.1 Electrically Conductive Adhesives: EC151L

ECA consists of an organic/polymeric binder matrix and metal filler. The ECA that we studied, Polytec EC151L[1], is a 100% solid, two components, electrically conductive epoxy. This ECA is suggested for electrically conductive bonding and coating applications. The pot life at 23 °C is 2 days and the viscosity is 4800 mPa.s at 23 °C. The minimum bond line cure schedule is 60 minutes at 90 °C, 15 minutes at 150 °C and 40 s at 180 °C. The thermal properties give a 400 °C degradation temperature, and between −55 °C and 300 °C for the intermittent temperature. The special chemistry of this epoxy also allows rapid cure cycles at higher temperatures and will withstand reflow and wire bond processes up to 300 °C.

6.2.1.2 Laser IC Decapsulation

The Laser IC decapsulation we used is the one described in section 3.2.4.2. We use in this chapter laser attacks on the wavelength $\lambda = 1064$ nm.

6.2.2 Theoretical Method

6.2.2.1 Theoretical Study of Main Pigments and Dyes

There are two types of colouring matter: dyes and pigments (Table 6.1).

The dyes (soluble in solvents and substrates) can be separated into 25 classes according to the chemical groups present in the molecule of the dye [Waring and Hallas, 2013]. The best known families are the Azo, Anthraquinone, Indigo and Xanthene dyes.

The pigments (not soluble in solvents and substrates) can be separated into two main families: organic pigments [Herbts and Hunger, 2004] and inorganic pigments [Buxbaum, 2008]. Organic pigments can also be separated into two subfamilies: organic Azo pigments (Monoazo, Diazo, Beta-Naphthol, Naphthol AS, Benzimidazolones, etc.) and Polycyclic Non-Azo organic pigments (Phthalocyanine, Anthraquinone and Perylene pigments). In the family of inorganic pigments we find iron oxide (Fe_2O_3), chromium oxide (Cr_2O_3), carbon black and mixed metal oxides.

[1] http://www.polytec-pt.com/int/products/epoxy-adhesives/electrically-conductive-adhesives/

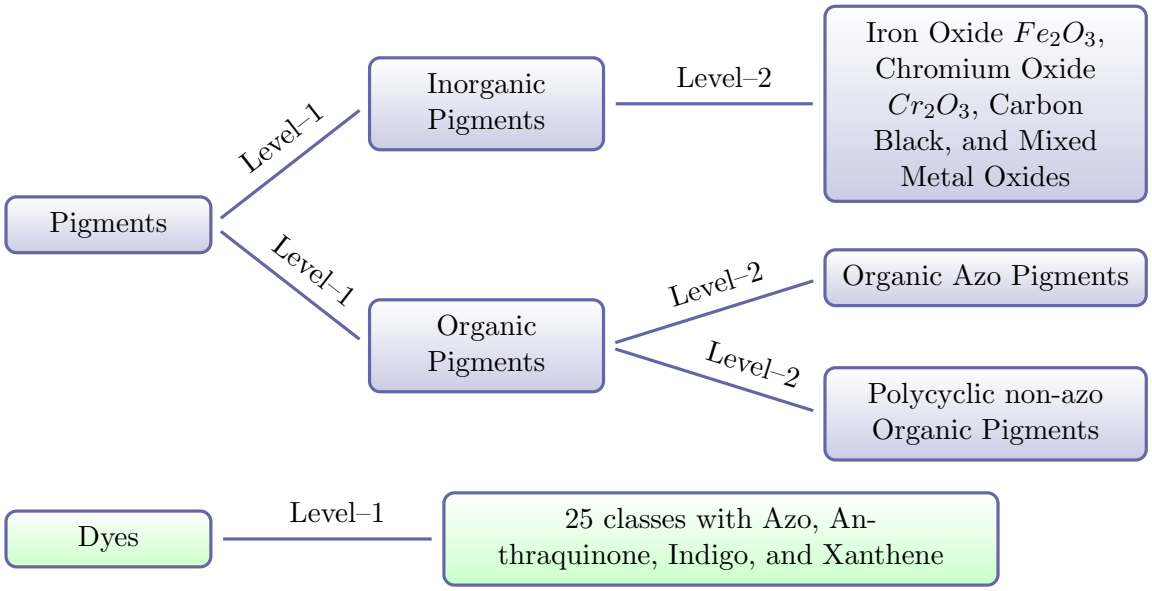


Table 6.1: Synthesis of the Two Types of Colouring Matter

6.2.2.2 Choice of Our Dye and Pigments

As shown by the calculation of the absorbance (Equation 3.7), the molar extinction coefficient at the wavelength λ plays a major role in the calculation of the total absorbance of the mixture. We have therefore looked more closely at the reflectance spectrum (the complement of the absorbance at zero transmittance) of several dyes at the wavelength $\lambda = 1064$ nm, using an infrared spectrophotometer.

The reflectance spectra were acquired from 320 to 2500 nm, step 1 nm. We have selected Indigotin ($C_{16}H_{10}N_2O_2$) [Leona et al., 2004], Eriochrome Black T and Sudan Black because their NIR (Near Infrared) reflectance spectra show high absorbance at the 1064 nm wavelength. As an example, Indigotin ($C_{16}H_{10}N_2O_2$) has the reflectance spectrum on Figure 6.4.

By contrast, we did not select Maya blue pigment [Leona et al., 2004] because its NIR reflectance spectrum (Figure 6.4) shows a small absorbance at the wavelength 1064 nm.

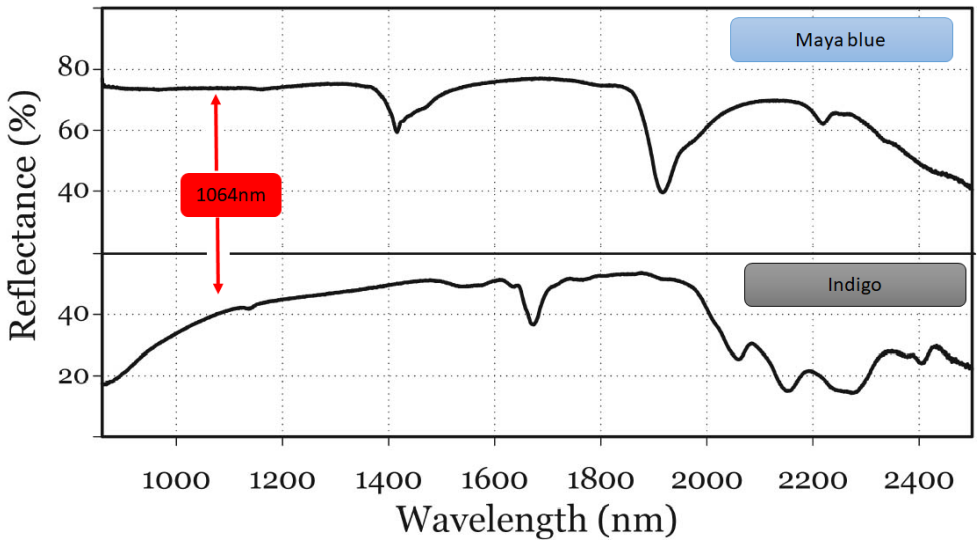


Figure 6.4: Indigotin ($C_{16}H_{10}N_2O_2$) and Maya Blue Reflectance Spectra, 1064nm

We note here that for 1064 nm the absorbance of Indigotin is better than the absorbance of Maya blue. Thus, when using our laser, a mixture of our glue with Indigotin will be theoretically preferable for a better laser–material interaction (which our experimental results confirm). Similarly, if we look at the NIR reflectance spectrum of Maya blue pigment, we observe a better absorbance at 1920 nm wavelength than the Indigo (Figure 6.5).

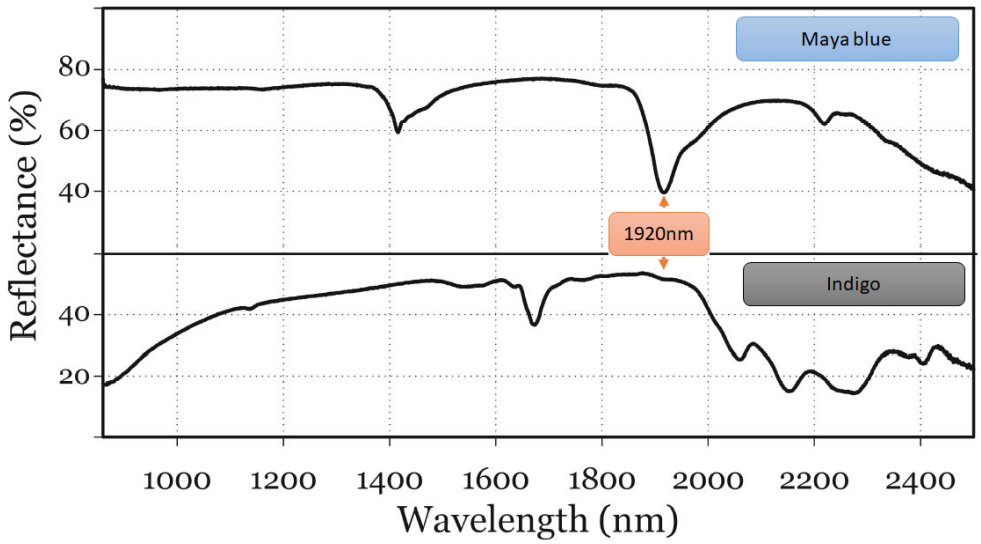


Figure 6.5: Indigotin ($C_{16}H_{10}N_2O_2$) and Maya Blue Reflectance Spectra, 1920 nm

Thus, if we now use an ablation laser at 1920 nm wavelength, a mixture of our glue with the Maya blue will be theoretically more interesting for a better laser (1920 nm)–material

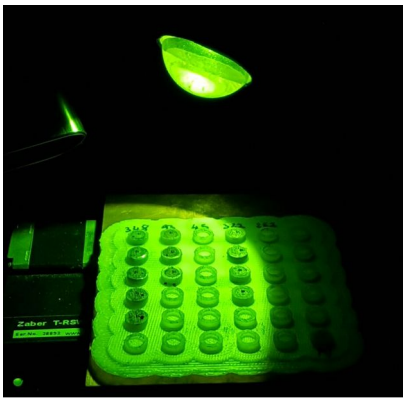
interaction. A form of attack selectivity can even be implemented with laser attacks. Here we find an analogy with attacks with acids. Indeed, certain mixtures of acids make it possible to create a layer of passivation (protective layer) and therefore to be selective on the target to be destroyed. As discussed in chapter 2, a mixture of fuming nitric acid with a 100% concentration (3 volumes) and sulphuric acid with a 100% concentration (1 volume) is an example of selective mixing. The first theoretical results show the possibility of using pigments and dyes with conductive glues. Let us now look at their practical implementation.

6.3 Experimental Results

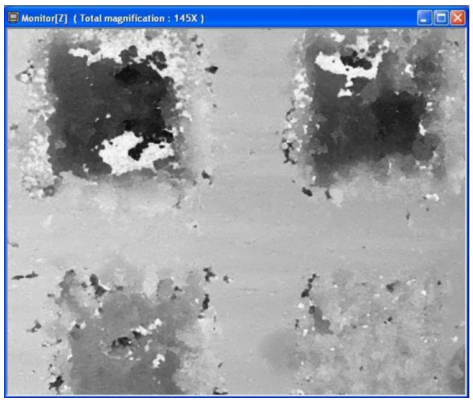
6.3.1 Experimental Method

All our experiments were carried out with a specially defined protocol in order to quantify the effectiveness of laser interference on our target material that we prepared. In this part we performed all our tests in the following way:

- The first step is the **Production** of mixtures with different concentrations of pigments and dyes.
- Laser attacks with fixed power settings: 40%, 50%, 60%, and 70% of the maximum laser power (30 W) (Figure 6.6a).
- Finally the last step is the **Measurement** of the average ablated depth using a focal plane microscope (Figure 6.6b).



(a) Glue mixtures with different concentrations of pigments and dyes under the laser

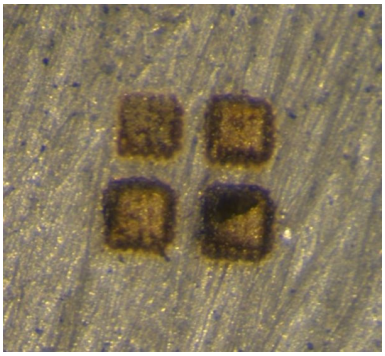


(b) Measurement of the average ablated depth using a focal plane microscope

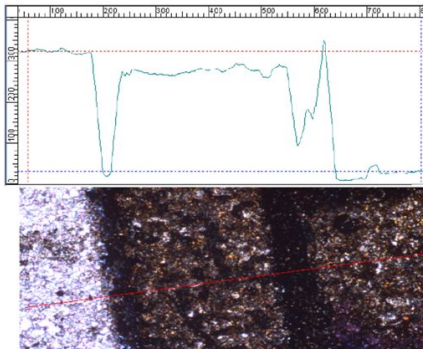
Figure 6.6: Experimental Protocol

6.3.2 Interaction Results (1064 nm Laser–EC151L) Without Dye and Pigment

In the first experiment we applied the protocol directly to the ECA adhesive with 0% dye or pigment. We made 4 patterns, 500ms, 15kHz, (Figure 6.7a) with: **Pattern 1,** 4 ablation passes with the laser at 40% of its maximum power; **Pattern 2,** 4 ablation passes with the laser at 50% of its maximum power; **Pattern 3,** 4 ablation passes with the laser at 60% of its maximum power; and **Pattern 4,** 4 ablation passes with the laser at 70% of its maximum power.



(a) Laser power: 40% top left, 50% top right, 60% bottom left, and 70% bottom right



(b) Measurement of the average ablated depth using a focal plane microscope

Figure 6.7: Laser Interaction on ECA Without Dye or Pigment

Our experimental results (Figure 6.8) show that without dye and with laser power below 50%, the effect of the laser is nearly zero on the glue (less than 20 µm).

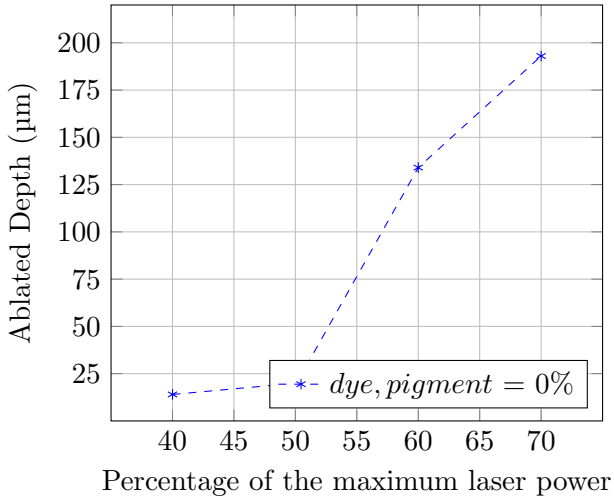


Figure 6.8: Without Dye and Pigment: Measurement of the Average Ablated Depth (Averaged Over 5 Measurements) Using the Focal Plane Microscope

An increase in laser power, above 50%, can have adverse effects on silicon chips. Thus, the aim of this work was to make the glue sensitive to the laser for laser powers below 50% of the maximum laser power and thus protect the silicon chip.

6.3.3 Interaction Results (1064 nm Laser–EC151L) With Indigotin Dye

In this part we made mixtures of ECA with several concentrations of Indigotin ($C_{16}H_{10}N_2O_2$) dye (Figure 6.9). Our experimental results (Figure 6.10) show that the dye has an immediate effect on the ablated depth, and this even at 40% of the maximum laser power.

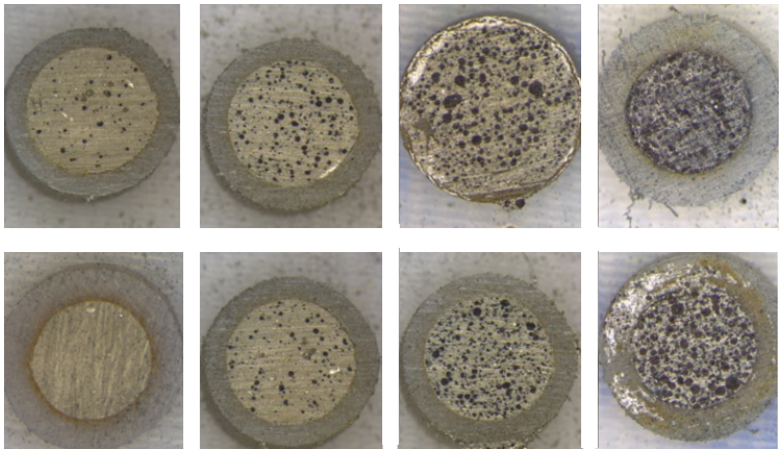


Figure 6.9: (Left to Right), Bottom: 0%, 1.5%, 3%, 7%, Top: 4.5%, 8.5%, 17% and 25% of Indigotin Dye

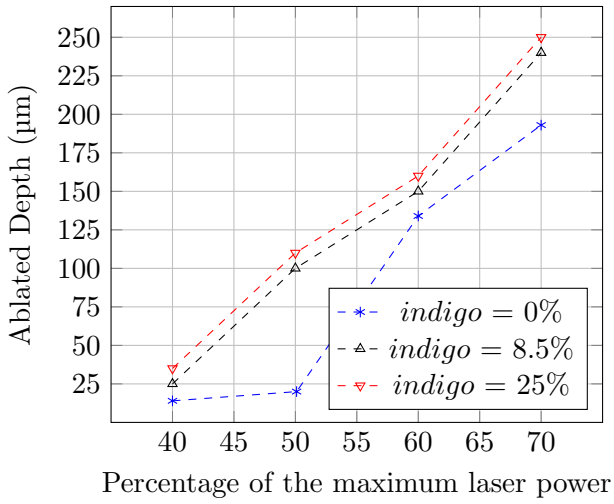


Figure 6.10: Indigo: Measurement of the Average Ablated Depth (Averaged Over 2 Measurements) Using the Focal Plane Microscope

With an 8.5% mass percentage of dye in our glue we multiplied the ablated depth by a factor of 1.8 (40% laser power) compared with our glue without dye. With a 25% mass

percentage of dye in our glue we multiplied the ablated depth by a factor of 2.5 (40% laser power) compared with our glue without dye. We noted that mixtures below 5% dye have no effect on the ablated depth. Mixtures above 30% have the effect of rendering the adhesive very pasty and therefore very difficult to apply, and caused the glue to set too quickly without giving us enough time to apply it.

6.3.4 Interaction Results (1064 nm Laser–EC151L) With Eriochrome Black T Pigment

In this part we made mixtures from our ECA glue (151L) with several different concentrations of Eriochrome Black T pigment (Figure 6.11).

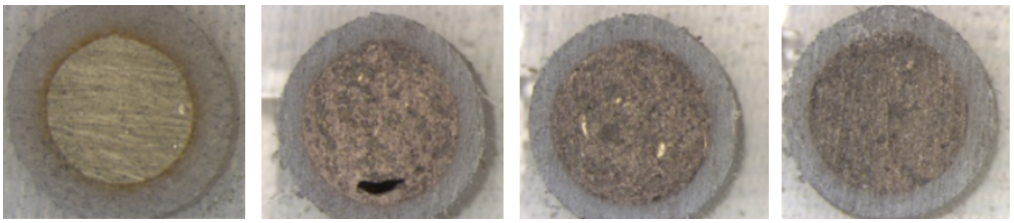


Figure 6.11: (Left to Right) 0%, 12%, 36%, and 60% of Eriochrome Black T

We noted that the Eriochrome Black T had an immediate effect on the ablated depth (Figure 6.12) even at 40% of the maximum laser power.

With a 12% weight percentage of Eriochrome Black T in our glue, we multiplied the ablated depth by a factor of 1.8 (to 40% laser power) compared with our glue without pigment. With a 36% weight percentage of Eriochrome Black T in our glue we multiplied the ablated depth by a factor of 2.5 (to 40% laser power) compared with our glue without pigment. With a 60% weight percentage of Eriochrome Black T in our glue we multiplied the ablated depth by a factor of 2.8 (to 40% laser power) compared with our glue without pigment.

We noted that mixtures below 5% of Eriochrome Black T have virtually no effect on the ablated depth. Mixtures above 60% have the effect of making the glue very pasty, granular and very difficult to apply, as well as causing the glue to set very quickly without having time to apply it.

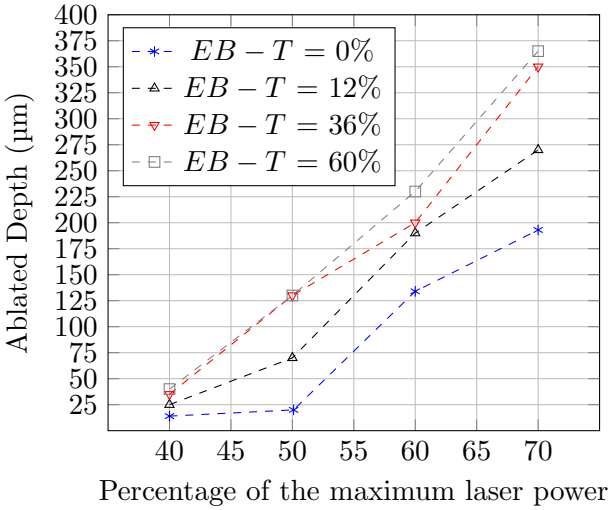


Figure 6.12: Eriochrome Black T: Measurement of the Average Ablated Depth (Averaged Over 2 Measurements) Using the Focal Plane Microscope

6.3.5 Interaction Results (1064 nm Laser–EC151L) With Sudan Black Pigment

In this part we made mixtures of our ECA glue (151L) with several different concentrations of Sudan Black pigment (Figure 6.13).

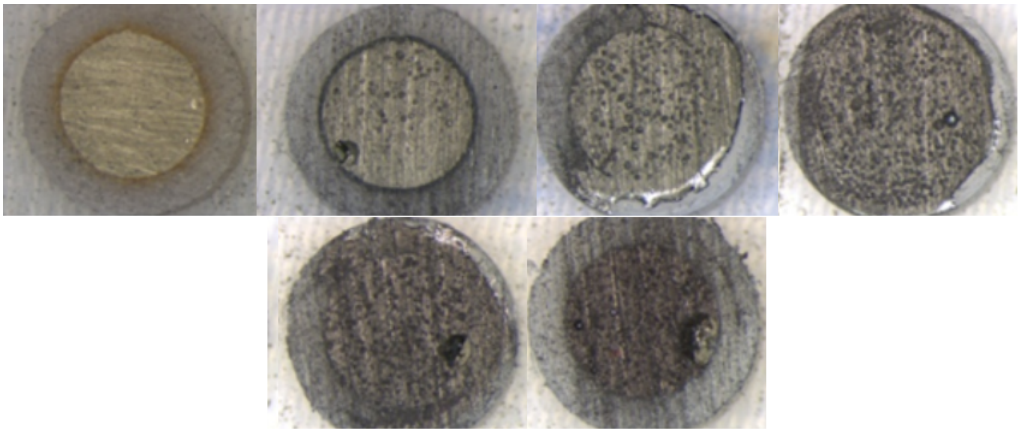


Figure 6.13: (Left to Right, Top: 0%, 1.5%, 7.5%,15%, Bottom: 25%, and 35% of Sudan Black Pigment

We noted that the Sudan Black has an immediate effect on the ablated depth (Figure 6.14) even at 40% of the maximum laser power.

With a 7.5% mass percentage of Sudan Black in our glue we had no influence on the ablated depth (at 40% laser power) compared to our pigment-free glue. The effect began

at 50% laser power. Indeed, with a 7.5% mass percentage of Sudan Black in our glue we multiplied the ablated depth by 3.5 (to 50% laser power) compared to our glue without pigment.

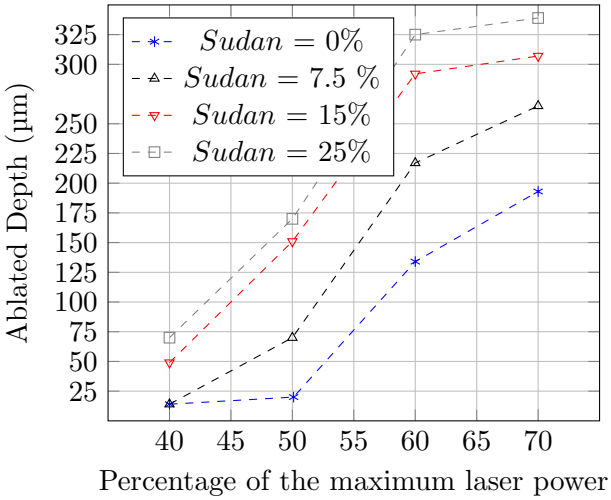


Figure 6.14: (Left to Right, Top to Bottom) Sudan Black: Measurement of the Average Ablated Depth (Averaged Over 2 Measurements) Using the Focal Plane Microscope

With a 15% mass percentage of Sudan Black in our glue we multiplied the ablated depth by a factor of 3.5 (at 40% laser power) compared to our glue without pigment. With a 25% mass of Sudan Black in our glue we multiplied the ablated depth by a factor of 5 (at 40% laser power) compared to our glue without pigment. We noted that a major effect of Sudan Black was to make the normally electrically conductive glue electrically insulating with 5% or more mass percentage added.

6.4 Direct Forensic Application: Repair of a Broken Bonding Wire

We now present a concrete example of application on an eMMC type memory. A laser decapsulation was performed (Figure 6.15) in order to access the bonding wires (the goal was to do nanoprobing).

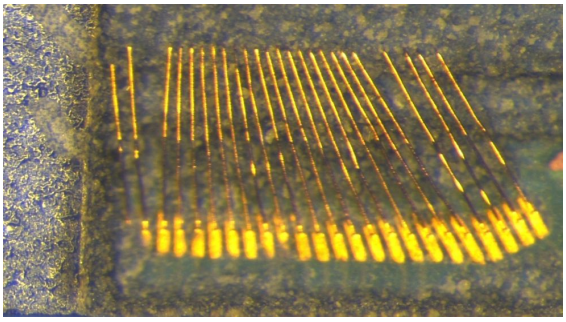


Figure 6.15: Bonding Wire Access After 1064 nm Laser Decapsulation

When performing the nanoprobing, the micro-probe broke a bonding wire (Figure 6.16) in an area where repair would be very complicated due to the bonding wires being close together.

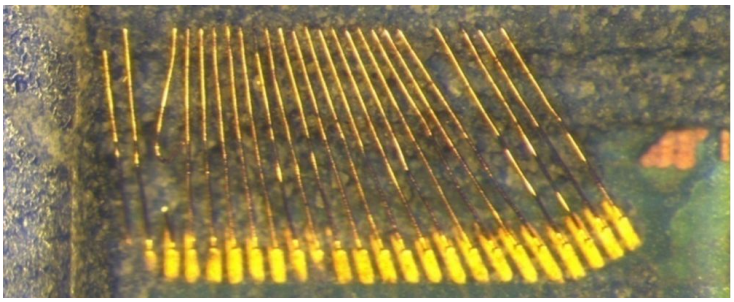


Figure 6.16: Micro-Probe Bonding Destruction

Given the high density of bonding wires in the immediate vicinity, it is very difficult to recreate the bonding wire using the wire-bonder station without the risk of destroying the neighbouring bondings. Using a micro-instrument, a thin layer of conductive glue containing 36% Eriochrome Black T (Figure 6.11) was applied to the area Figure 6.17):

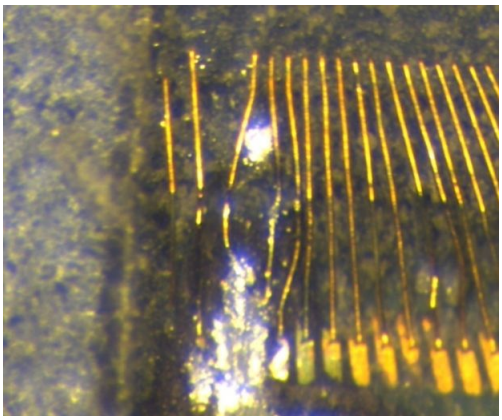


Figure 6.17: Application of the Pigmented ECA

To remove the excess ECA, which creates false contacts between the bonding wires, we superpositioned a mask that defined the path of the laser beam (Figure 6.18).

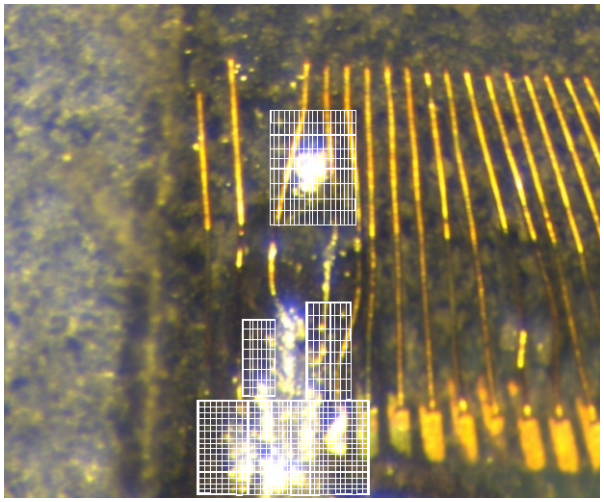


Figure 6.18: Mask Superpositioned to Define the Laser Beam Path

We applied between the bonding wires, 4 passes of the laser (Figure 6.19) with 40% of the maximum power (30W).

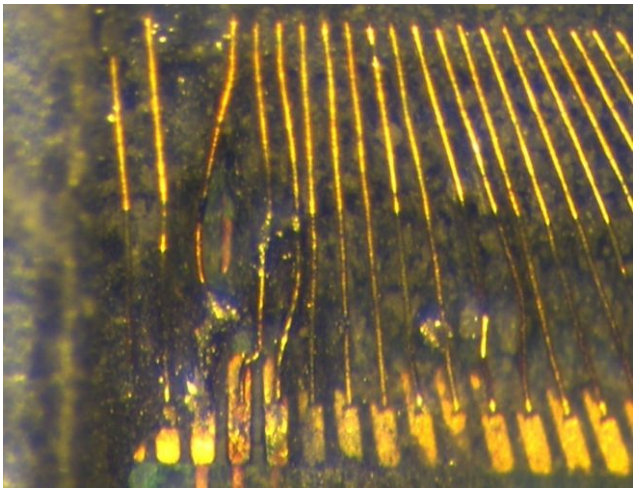


Figure 6.19: Repair Result After 4 Passes of the Laser at 40% of the Maximum Power

Our protocol allowed us to separate different conductive zones and isolate them. The repair was effective and the memory re-run was successful. The technique that we have developed allows us to repair the bonding wires with an accuracy of the order of 15 µm. We cannot go below 15 µm because we reached the limits of the precision of our laser. A superior laser would allow repairs below 15 µm.

6.5 Discussion of the Experimental Results

6.5.1 Limit of the Amount of Dye and Pigment on the Structure of the Adhesive

We found experimentally that we could not go on increasing the quantity of dye or pigment in the glue mixture. Our theoretical study showed us that a greater quantity of dye or pigment produced more effective laser ablation, and that we could reduce the power of the laser. But we found in practice that there was a threshold beyond which the behaviour of the glue mixture changed. The glue became difficult to model for micrometric applications. For large concentrations of dye or pigment, the adhesive also lost its conductive properties. Thus, since the maximum quantity of dye and pigment is limited experimentally, we cannot indefinitely reduce the energy deposited by the laser.

Sudan Black is not a good pigment for use in conductive conditions. Indeed, even if the effect of the laser is significant at 40% of the maximum laser power, we noted a total loss of the conductivity of the glue mixture. However, this loss of conductivity can be used to carry out isolation operations and then laser modelling (inverse operation).

6.5.2 How to Choose Between Pigment and Dye?

The indigo dye tended to create dye clusters while the Eriochrome Black T blended uniformly (Figure 6.20).

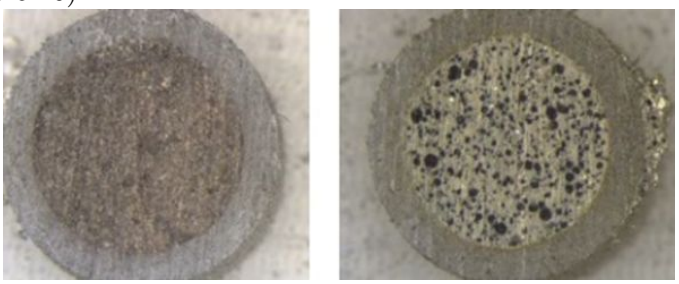


Figure 6.20: Eriochrome (Left), Indigo (Right)

When the laser attacks an indigo dye bubble (area where the molar extinction coefficient at the lambda wavelength is important), it created a cavity that was much larger than the area we wanted to ablate (Figure 6.21).

Thus the use of Eriochrome Black T is recommended for precise ablations because the pigment creates a uniform zone in which the molar extinction coefficient at the wavelength is increased in comparison to our glue without pigment.

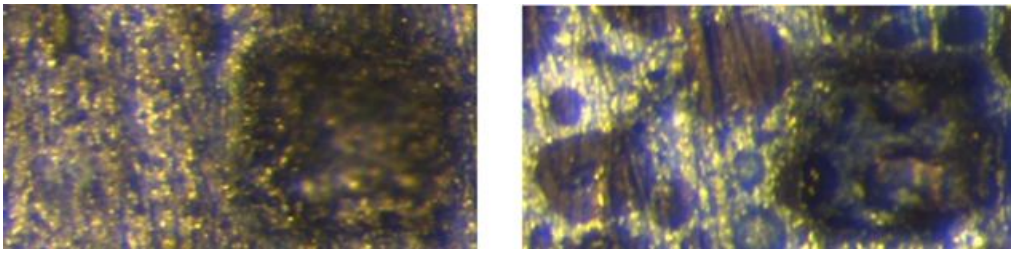


Figure 6.21: Eriochrome (Left), Indigo (Right): Cavity After Laser

6.6 Conclusion of the Experiments

Now that we have characterised and optimised the energy deposited during laser attacks by the dyeing and pigmenting of the ECA, we want to apply these methods on two other types of adhesives: Thermally Conductive Adhesives (TCA) and UV-Adhesives (UVA). Unfortunately, UVA is quite challenging to work with because the pigments/dyes do not allow the ultraviolet radiation needed for curing to penetrate and consequently they do not allow the glue to set.

The repair of bonding wires is actually a very complex step which entails a risk of further bonding destruction. The goal of this chapter is to propose a method that couples pigments and dyes mixed with electrically conductive adhesive (ECA) with the precision of the ablation laser for the repair of damaged bonding wires.

This study is dual. It first introduces the interaction properties between ECA and 1064 nm laser decapping attacks. A new technique was developed using an ECA–pigment mixture and the study was successfully applied to the forensic micro-repair of wire bonding. The technique that we developed allows us to repair the bonding with an accuracy of the order of 15 µm.

7

Advanced Transplantation Technique

As we saw in chapter 3, total transplantation is the only solution available to the investigator when recovering data from an encrypted phone damaged after an air crash, accident or an attack. However, with the arrival of new cryptographic components and modern Package On Package (PoP) processors, the conventional transplant technique is reaching its limits. We propose in this chapter a new method of transplantation that we call: PoP chip-off/TCA method. This method is now applicable to new generations of encrypted phones (BlackBerry, iPhone, Samsung, etc.).

7.1 Background and Limits

7.1.1 What is a PoP Component?

Package on Package is a new semiconductor packaging process consisting of the stacking of two or more dies (memory die, CPU die, RAM, etc.) on top of one another (Figure 7.1)[1].

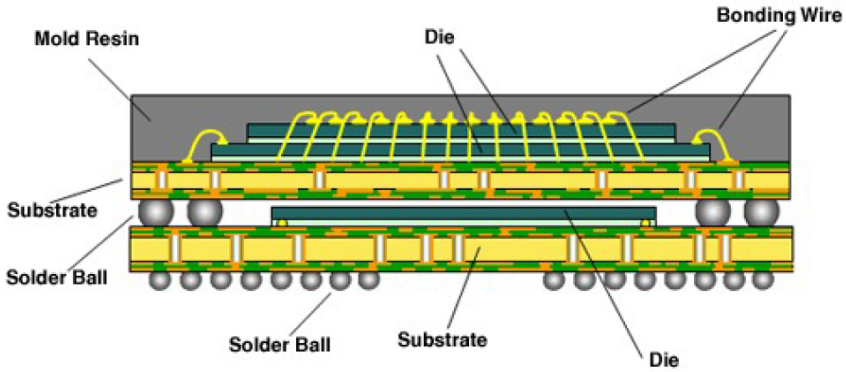


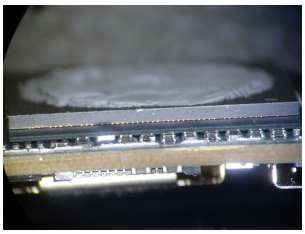
Figure 7.1: Package On Package Component Principle

The technology combines a vertical component, where two or more Ball Grid Arrays (BGAs) are stacked. As evidence, manufacturers are now increasingly adhering to this

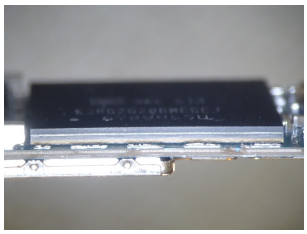
[1] https://electronics.stackexchange.com

technology (Apple iPhone 6S uses A9 APL0898 processor and Micron DRAM):

- BlackBerry Z10 (Figure 7.2a) uses Qualcomm MSM8960 Snapdragon S4 Dual-core 1.5 GHz Krait CPU stacked to 2 GB RAM.
- Samsung S7 edge (Figure 7.2b) uses Qualcomm Snapdragon 820 processor and SK Hynix H9KNNNCTUMU-BRNMH 4 Go LPDDR4 SDRAM.



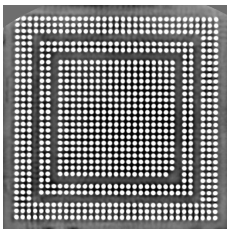
(a) BlackBerry Z10 PoP CPU: CPU (bottom) and RAM (top)



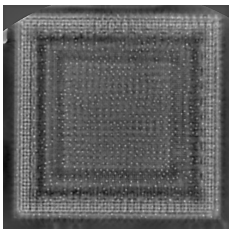
(b) Samsung S7 edge: Qualcomm PoP processor

Figure 7.2: BlackBerry and Samsung Stacked CPUs

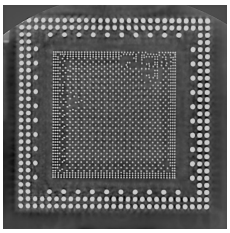
- BlackBerry 9900 (Figure 7.3) uses a single core, 1200 MHz, QC 8655 CPU stacked to 0.75 GB RAM.



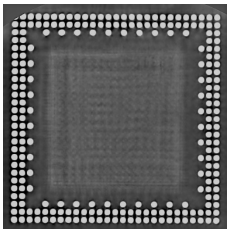
(a) Bottom CPU BGA on PCB



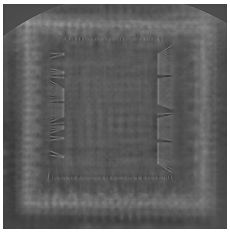
(b) CPU via connection



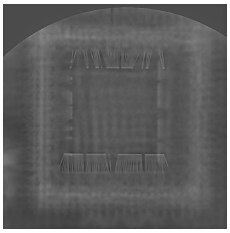
(c) CPU and RAM junction



(d) RAM BGA on the CPU



(e) Middle RAM bonding wires



(f) Top RAM bonding wires

Figure 7.3: BlackBerry 9900 PoP CPU X-ray Images

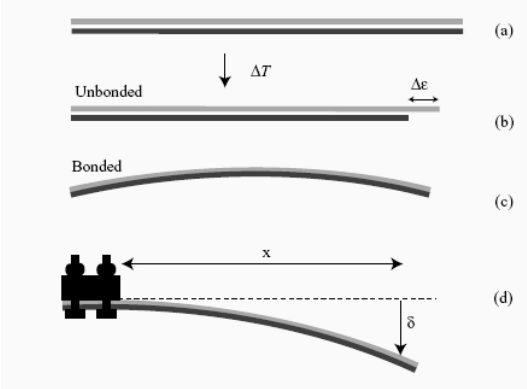
Manufacturers are increasingly using PoP components to save space on the electronic board. The race for optimisation is an important factor that reduces production costs and decreases the weight of the phone. But the main advantage is the increase in the level of secret information that no longer passes through the printed circuit, but directly inside the chip of the package.

7.1.2 Traditional Techniques and Their Limits

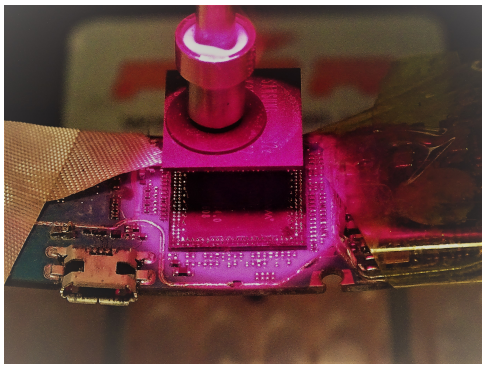
The traditional BGA unsoldering technique (chapter 3) quickly reaches its limits and leads to the destruction of the PoP component (Figures 7.4b and 7.4c). The mechanical forces exerted are too important (coefficient of thermal expansion α and Hooke's law, Equation 7.1).

$$\text{radius of curvature} = \frac{2\sin\tan^{-1}(\frac{\delta}{x})}{\sqrt{x^2+\delta^2}} \quad \text{with} \quad \Delta\epsilon = \Delta\alpha\Delta T \tag{7.1}$$

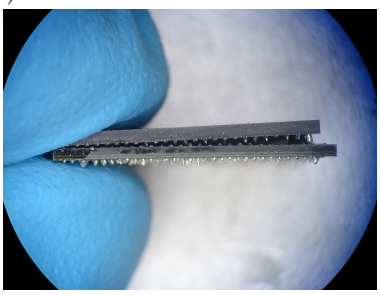
Those forces create false internal contacts making the component unusable (Figure 7.4d).



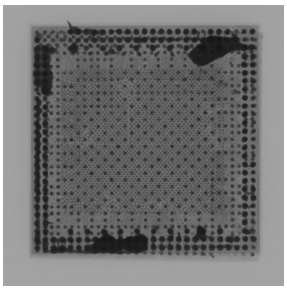
(a) Mechanical forces: Hooke's law



(b) BGA Station: CPU Destruction



(c) BlackBerry Z10 PoP CPU: CPU on the (bottom), RAM (top)



(d) BlackBerry Z10 PoP CPU After Chip-Off (X-ray Image)

Figure 7.4: Chip-Off PoP CPU Destruction and Hooke's law

Similarly, the lapping technique is not applicable because of the electronic components that face each other on the PCB (Chapter 3). The lapping would either destroy the memory or destroy the CPU. Knowing that encryption keys are in these two components, the method is no longer applicable.

7.2 PoP Chip-off/TCA Adhesive Method

7.2.1 High Temp Thixotropic Thermal Conductive Adhesive (HTTTCA)

This polymer family is designed to dissipate heat [Felba, 2011, Falat et al., 2007]. Generally, this adhesive covers memory components and CPUs to reduce chip heating and improve sturdiness, which makes chip-off analyses considerably more difficult. The thermal conductive adhesive studied in this thesis, Polytec TC430[2], consists of two components: a resin and hardener. The mixing ratio by weight is 100 resin units for 4 hardener units (Figure 7.5).

Figure 7.5: Resin (Left), Hardener (Right)

Viscosity at 23 °C is 13000 mPa.s and minimum bond line cure schedule is 60 minutes at 100 °C and 15 minutes at 150 °C. The pot life at 23 °C is 2 days, which allows us to make alterations. The degradation temperature is 400 °C (later in this chapter we will show the fundamental importance of this information). Another suitable interesting property is that TC430 is a thixotropic consistency paste. A fluid or material is said to be thixotropic if, under constant stress (or gradient of velocity), its apparent viscosity decreases with time. In addition, TC430 will not expand with the increase in temperature, which is a major advantage in the transplantation of PoP components.

7.3 Method

7.3.1 Applied Method

The new PoP chip-off/TCA method can now be described in steps. This process is necessary for transplantation, when there are no alternative ways to de-solder PoP

[2] http://www.polytec-pt.com/fileadmin/user_uploads_Polytec-PT/home/documents/Polytec_Klebstoffe_ENG/Polytec_TC_430_engl.pdf

components. The process used was realised in steps, as follows:

Step 1: Identification of the components that need to be transplanted. It is not necessary to transplant all the components, as many of them are unnecessary. Before any transplantation, it is necessary to understand the security mechanisms inside the phone by hardware/software reverse engineering. When a phone has been damaged, many small electronic components can be the cause of the malfunction that led to the forensic transplant. To minimise the risk of transplanting defective electronic components, it is necessary to transplant only the essential components to recover the forensics data (memory components, CPU, and cryptographic chips).

Step 2: X-ray tomography is performed to verify that the components to transplant are not destroyed, that all bonding connections are undamaged (Figure 7.6), and that there is no silicon fracture.

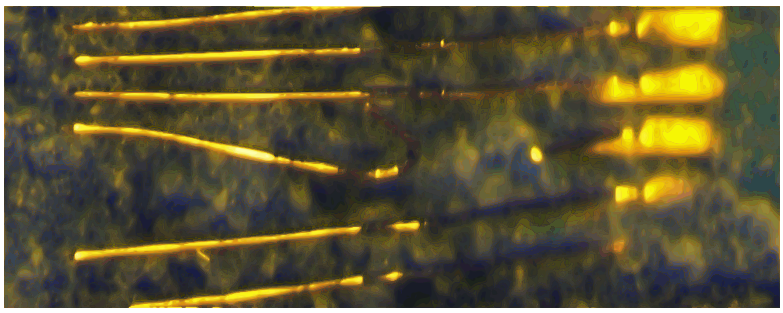
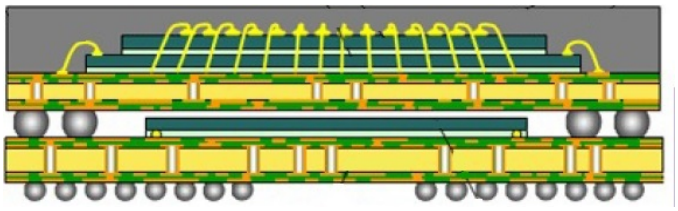


Figure 7.6: eMMC Memory Broken Bonding Wire

In the forensic cases encountered (crash, accident, attacks) this step is often essential because verification of the component's physical aspect makes it possible to avoid irreversible destruction (and thus loss of data) typically due to a broken bonding wire (false contact) or to silicon weakened during the impact.

Step 3: With a micro-tool, the High Temp Thixotropic Thermal Conductive Adhesive (HTTTCA) can be applied into the stack (between the RAM and the CPU) (Figure 7.7).



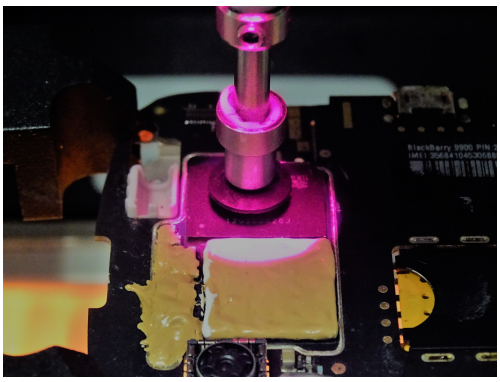
(a) PoP Without HTTTCA

(b) PoP With HTTTCA in Between Stack

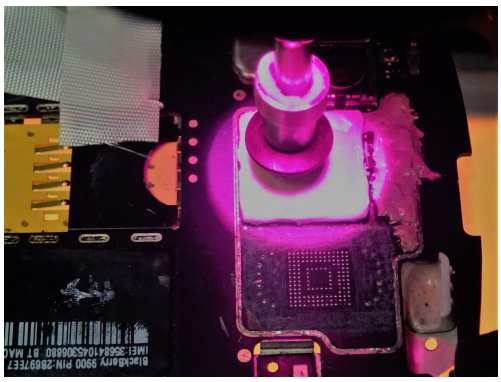
Figure 7.7: Application of the HTTTCA

Once the adhesive has been applied, the adhesive must dry according to the manufacturer's data. As the glue is thixotropic, it will not build up pressure (expand) during the drying process. There is hence no risk of breaking off the stacks and destroying the electronic component.

Step 4: A classic unsoldering process (chip-off) [Breeuwsma et al., 2007] is applied to all the necessary components to be transplanted (Figure 7.8). The components required for forensic transplantation are those determined at the end of Step 1.



(a) Memory Chip-Off



(b) PoP Processor Chip-Off

Figure 7.8: Chip-Off: Classical Process

Step 5: The PoP must be checked for stability by using X-ray tomography (Figure 7.9). The main advantage of the technique proposed in this chapter is the ability to realise the chip-off technique without creating false contacts between the balls and between the different levels of the PoP. This step is essential to verify that the technique has been performed correctly.

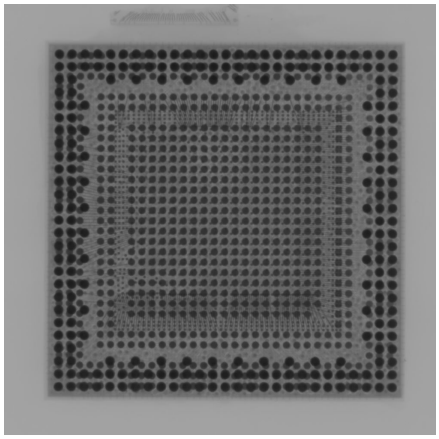


Figure 7.9: X-ray Control to Ascertain That the PoP Component Has Not Moved

Step 6: Clean the chip using a soldering iron and remove the flux. Once this is done, apply the low-temperature reballing step process. It is advisable to solder with a low temperature paste to minimise the thermal shock applied to the electronic component. Re-soldering with a high temperature paste is possible but riskier and/or potentially damaging to the component. It is important to remember that the component has already undergone a major shock during a traumatic incident and must be considered as already fragile.

Step 7: The donor board is prepared by removing the necessary components determined in Step 1. The main board's components will be soldered (Step 9) in its PCB's gap position. The components can be removed (Figure 7.10) either by de-soldering using the conventional chip-off technique, or by lapping.

Figure 7.10: Donor Board eMMC Gap Position

This step is important because no other components of the donor board should be damaged. The choice of one or the other method must be reflected upon judiciously to obtain a perfect donor board. Thus, in the case of heat-sensitive components, the lapping method will be preferable. However, when components are located in a high density area and these are sensitive to mechanical stress, the chip-off technique would be preferable.

Step 8: Re-attach, with a low temperature, the main board's components onto the donor board using a BGA station or manually (chapter 3).

Step 9: Using X-ray tomography, it must be verified that the components are well reworked on the donor board and checked that there are not false electrical contacts between balls (Figure 7.11).

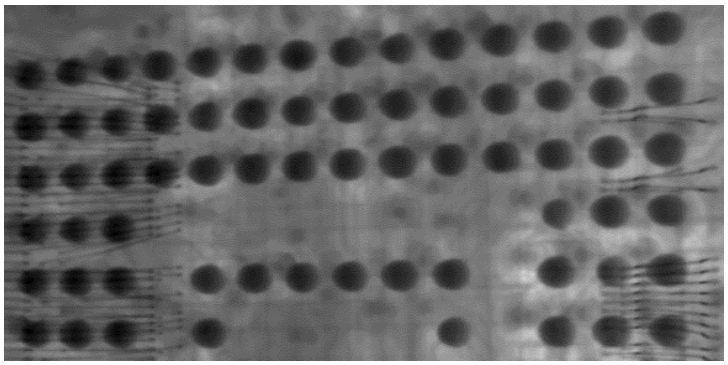


Figure 7.11: Soldering Confirmation on the Donor Board

This step allows us to verify that each ball is properly soldered. Note that an accidentally grounded output may destroy the component, as would a short circuit.

Step 10: Finally, the donor phone's components can be installed (screen, keyboard, etc.) and the phone turned on for forensic investigations. This stage confirms the successful completion of the forensic transplant. If this step is satisfactory, Step 1 will no longer need to be carried out for a phone of the same model (or GPS, etc.).

However, if Step 1 has been incorrectly performed, the phone will not boot, or anomaly indicators will appear (blue screen or others). These indicators will be the sign that all the necessary components have not been properly transplanted (e.g. cryptographic chips, CPUs, memories, etc.).

Finally, if the phone still does not boot even if Step 1 is performed correctly, Step 2 must be looked at more closely to find the malfunction. If the defective component is an essential component of Step 1, then it will be impossible to perform the forensic transplant.

7.4 BlackBerry 9900 PGP Transplantation

7.4.1 BlackBerry PGP Cryptography Process

The first step is to select the components to be transplanted. For this, it is essential to understand the encryption mechanisms used in the 9900 (Figure 7.1). The encryption mechanisms are described at the BlackBerry website [BlackBerry-Entreprise-Server, 2014].

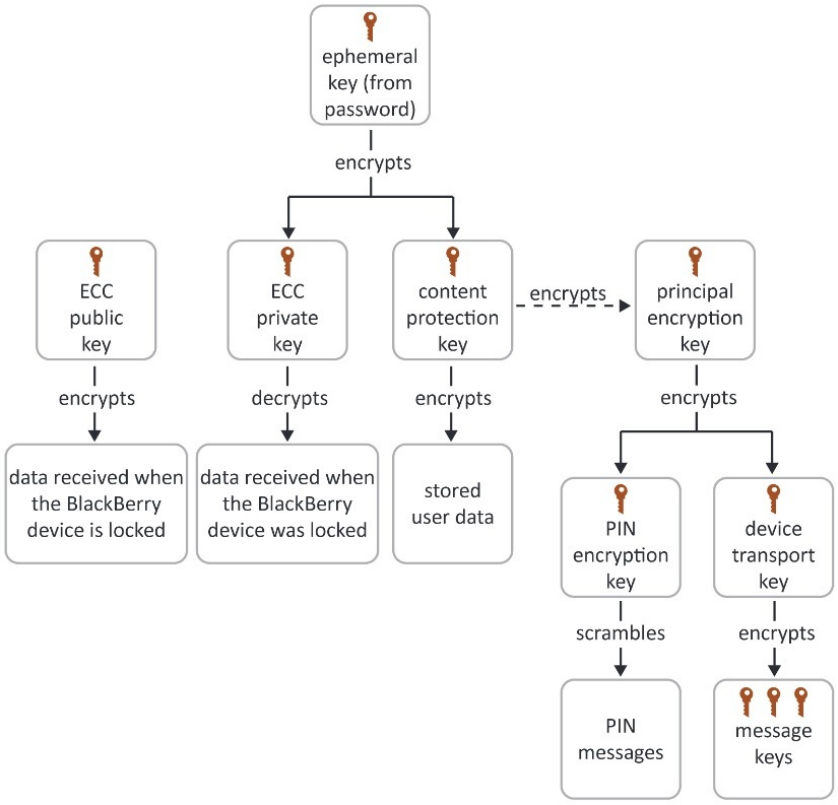


Table 7.1: BlackBerry PGP Key Hierarchy [BlackBerry-Entreprise-Server, 2014]

The encryption mechanism shows us that the hardware encryption key is etched in the silicon of the CPU, but we cannot physically access it in reading. This key is a kind of super key that will encrypt a container containing the Ephemeral Key.

The ephemeral key encrypts the ECC private key and content protection key on the device. This key is itself derived from the unwinding password entered by the user. This key will also allow investigator to decrypt the content protection key (this key encrypts user data on the device when the device is locked).

7.4.2 BlackBerry 9900 PGP Physical Transplantation

Step 1: The study confirms that the processor (containing the hardware key) and the memory component (containing the encrypted user data and enabling decryption) are the only two components to be transplanted (Figure 7.12).

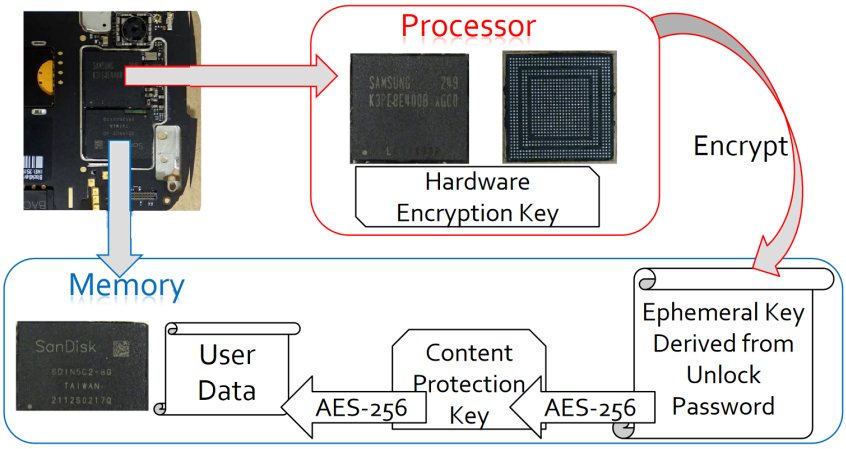


Figure 7.12: Simplified BlackBerry Encryption Process

Step 2: The two components of the main board are then radiographed to check that the components are not damaged (Figure 7.13). In this case, the components seem operational on X-ray diagnosis, so the investigator can move on to the next step.

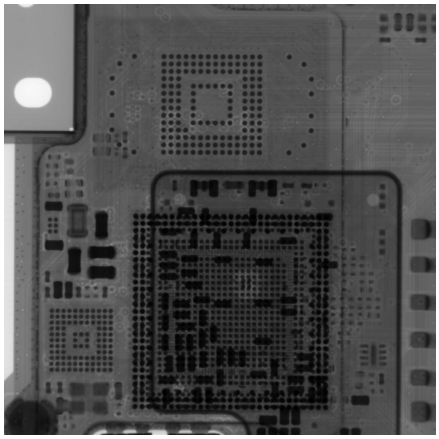
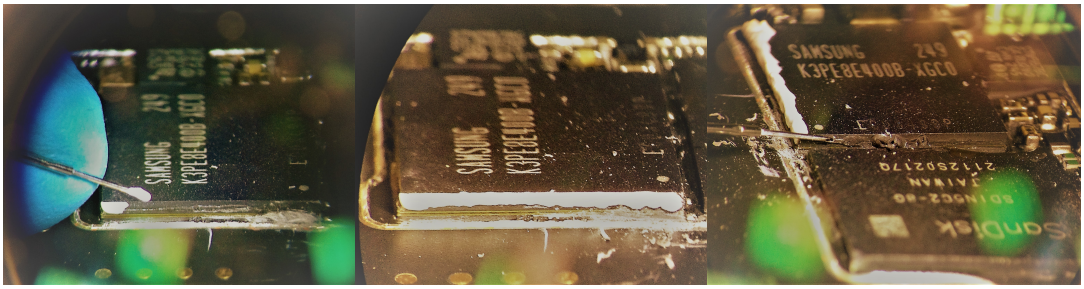


Figure 7.13: BlackBerry 9900 PoP CPU and Memory X-ray Soldering Control

Step 3: A High Temperature Thixotropic Thermal Conductive Adhesive (Polytec TC430-T) (Figure 7.14) is then applied with a micro-tool. If any underfill is present, it can then be removed to better apply our adhesive (Figure 7.14c). As soon as the four sides are glued (Figure 7.14e), the adhesive is left to dry (60 minutes at 100 °C or 15 minutes at 150 °C).

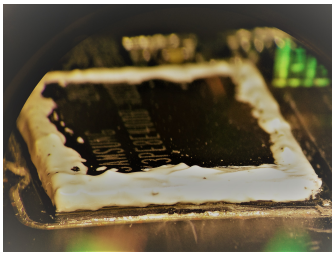


(a) Adhesive Applied With a Micro-Tool

(b) One Side Glued

(c) Underfill Removed With Micro-Knife

(d) Two Sides Glued



(e) Four Sides Glued

Figure 7.14: Step 3: Adhesive Deposited Between the Stack of the PoP CPU

Step 4: The PoP CPU and the memory are then de-soldered using the chip-off method (Figure 7.15). Blackberry 9900 phones have underfill under the processor and memory component.

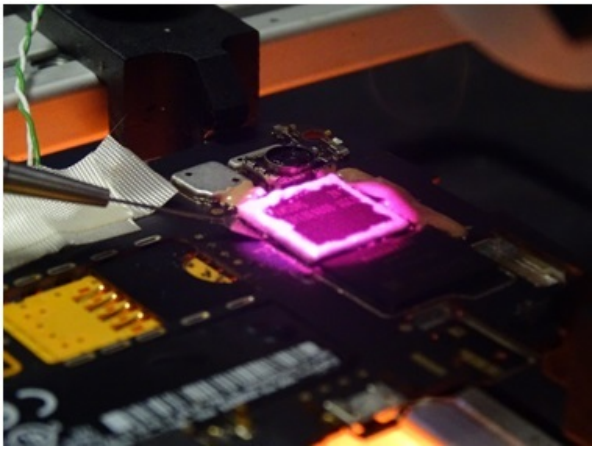


Figure 7.15: BlackBerry 9900 PoP CPU Unsoldering Using Chip-Off/TCA Technique

It is advisable to use the micro-cutting instrument (Ted Pella Micro-Tools) to facilitate the destruction of underfill glue. Since the temperature is important during the movement of the instrument, it is advisable not to carry out a pivot movement which could destroy the component.

The micro-instrument should only perform horizontal micro-movements, which will remove the underfill excess and thus facilitate the unsoldering of the electronic component.

As the degradation temperature of the TC430-T is much higher than the manufacturer's underfill, the stack does not move and the PoP remains in perfect condition.

Step 5: X-ray tomography is then performed to confirm that the technique has not damaged the stacked component. Each of the balls should be checked to confirm that there is no migration/junction of balls between them.

Step 6: The low-temperature reballing technique is then applied [Heckmann et al., 2016] (Figure 7.16). The masks are custom made and they perfectly fit the electronic component to be reballed. As has been studied in [Heckmann et al., 2016], the thickness of the mask is an important datum that will determine the quantity of material and therefore on the thickness of the ball. Thus, for the reballing of a CPU, with a large ball density, the thickness of the mask (127 μm) will be less than for a memory mask (152 μm).

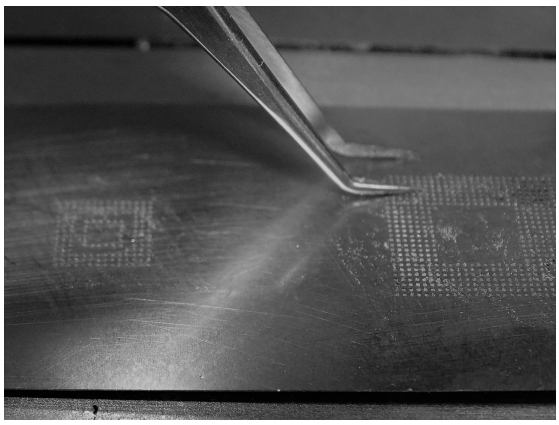


Figure 7.16: Reballing Process Using Stencil

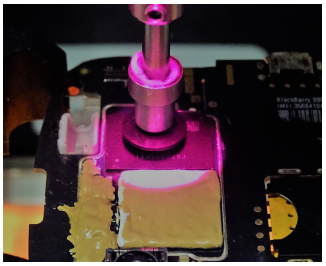
Step 7: The preparation of the donor board is then carried out (Figure 7.17). In order not to risk damaging the micro-electronic components located in the periphery of the memory and the CPU (capacity, etc.), the investigator must first deposit high-temperature thixotropic adhesive (Figure 7.17a). As much as possible, the donor board must be protected from the risks of thermal or mechanical shock applied while unsoldering and soldering.

Then unsoldering can be performed at 290 °C using a micro-instrument in order to remove the underfill and facilitate unsoldering (Figure 7.17b). A micro-cutting instrument and a heated micro-pane are used to gently remove underfill residues on the board (Figures 7.17d and 7.17e).

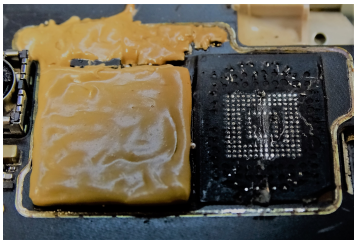
The same process for the processor is used (Figures 7.17f and 7.17g). As the processor soldering balls are smaller than the memory ones, it is necessary to be even more careful when cleaning the board. Indeed, the risk here is to tear off tracks present on the PCB

and thus destroy the donor board. After a final wash with flux remover, the donor board is ready to accommodate the main board components (Figure 7.17h).

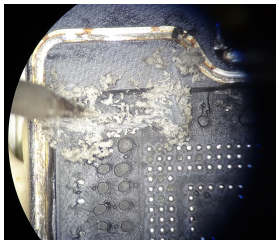
(a) Protecting the Donor Board With Adhesive



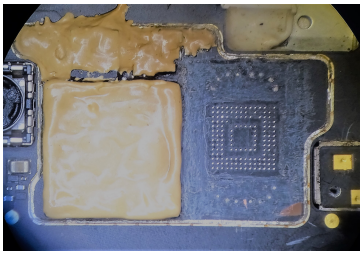
(b) Memory Unsoldering



(c) Memory Underfill

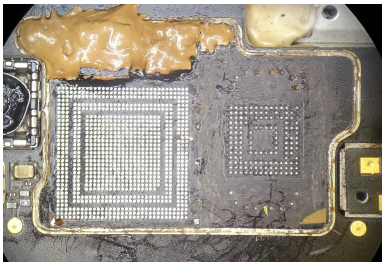


(d) Underfill Removed With Micro-Pane



(e) Donor Board Memory Cleaned Location

(f) Processor Unsoldering



(g) Donor Board Processor Cleaned Location

(h) Donor Board Is Ready to Accommodate the Main Board's Components

Figure 7.17: Step 7: Donor Board Preparation

Step 8: The CPU (from the damaged board) and the memory (from the damaged board) are re-soldered onto the donor board, with a BGA station at low temperature (150 °C).

Step 9: An X-ray is performed to check that the CPU and the memory have been properly soldered (Figure 7.18). Once the X-ray shows that there are no false contacts

between the balls and that they are properly soldered, the investigator can then move onto the next step.

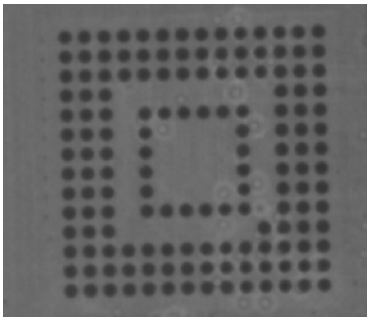


Figure 7.18: Checking eMMC Positioning and Good Soldering on X-ray

If there is any false contact, the component must be unsoldered, cleaned, re-balled at low temperature and then re-soldered on the donor board. A check radiograph should then be taken. If radiography does not reveal any issues, then we can move to the next step. Alternatively, the process must be repeated. The investigator has now completed the first part of the transplantation.

Step 10: All previous steps should have been correctly performed. So after having inserted the different components of the phone (keyboard, screen, battery, etc.), it can be switched on without risk. The phone should start properly on its boot sector (Figure 7.19). The forensic transplant is successful on this BlackBerry 9900 PGP phone.

Figure 7.19: Turn on Transplanted BlackBerry 9900 PGP

If Step 10 is correct, the investigator will be able to use the traditional forensic techniques (used to extract data from undamaged mobile devices) and can perform a traditional forensic mobile analysis.

7.5 Discussion

7.5.1 Choice of Adhesives and Limits

At present, mobile phone manufacturers use glue which degrades at about 300 °C. The key point of this chapter was to use an adhesive with a degradation temperature that was higher than 300 °C. The aim of the manufacturers, using this adhesive, is to facilitate the thermal diffusion of their components. It provides a means of protection against heat and thus increases the robustness of the components and their long-term reliability.

At the time of writing this thesis, manufacturers were not using glue with a degradation temperature of more than 300 °C, as it did not provide an additional thermal advantage for them. Thus, it would be possible for us to investigate the limits of our method if high temperature adhesives were to be used. We could quite logically find an initial solution by putting between the dies of the Package-On-Package component an even higher-temperature glue. However, it would be necessary not to forget the effect of the increased high temperature of unsoldering, which would surely lead to the destruction of the PoP component. It would then be of interest to consider covering the component with a powerful heat skin and to heat the copper layers of the PCB at a high temperature.

7.5.2 Which Electronic Components Should Be Transplanted?

Transplantation can be a long step, depending on the number of components to be moved. It is not a question of transplanting all the components, but of making a rigorous selection. To do this, the reverse engineering of the components must be performed, prior to the transplant test, to understand the security mechanisms inside the phone.

On mobile phones that have a low level of security (no encryption), the only component to be transplanted may be the memory. However, as we have just seen, on phones that embed encryption, such as the BlackBerry, the presence of hardware encryption keys on the processor make it necessary to transplant the memory and the processor.

Finally, some manufacturers, like Apple, make it necessary to identify the components that are crypto-paired together on the phone. Thus, we first have to understand the components that need to be transplanted: memory, processor and the ID crypto-components that are used.

7.5.3 Transplantation's Limit

We see that transplantation is impossible if the memory component and/or the CPU and/or crypto chip are affected (fractured silicon). Thus, before any transplantation, the essential components to be transplanted must be X-ray tomographed to diagnose the origin of the problem and proceed to the component's repair, like a bonding repair (chapter 6), board repair (chapter 5), micro-read (chapter 3), etc.

7.6 Conclusion of the Experiments

Unlike BlackBerry or Samsung devices, the transplantation of Apple devices (iPhone 6S, iPhone 7, iPhone 8 and iPhone X) is quite challenging because crypto-chips are paired inside the phone's board (not only memory and CPU). Hence it would be necessary to learn more about the crypto-chips paired inside Apple systems (BaseBand processor, Touch ID, EEPROM anti-RollBack and BaseBand flash), and then use the adhesives techniques to de-solder and re-solder the Apple Package on Package (PoP) electronic components in order to succeed in iPhones 6S, 7, 8, and X transplantation. This is the area of our current research.

The transplantation of the latest mobile phones is actually a very complex operation which entails a risk of PoP components destruction. The goal of this chapter was to propose a new method called "**PoP chip-off/TCA Technique**" which allows the unsoldering of PoP components without damaging them. More generally, the method assures a more successful transplantation in new mobile phones. We described the current methods used for transplanting phones and their limitations. Finally, the new method was developed and successfully applied to the forensic transplantation of a cryptographically protected mobile phone: BlackBerry 9900 PGP.

8

Conclusion

We have shown in the first lines of this thesis how French law defines and authorises the use of cryptography and its unrestricted circulation. Cryptography is now part of our daily lives and each person, conscious of it or not, uses cryptographic means. We were principally interested in the generalisation of cryptography in a context of judicial investigations conducted on mobile phones, because they provide a wealth of information to investigators, ranging from some simple localised facts to a summary of life.

However, as we have seen in recent tragic cases, such as the Germanwings air-crash or the mass shooting during the San Bernardino attack, investigators have been finding it increasingly difficult to extract data, due to widespread use of encrypted mobile phones. This is also true in the context of international narcotics trafficking, assassination, armed robbery, money laundering or terrorism, in which the leaders of networks or organised groups deploy batches of ultra-secure mobile phones (for example BlackBerry PGP). Criminals rely on the encryption to thwart surveillance from the authorities. A triple layer encryption provides criminals with secure communication and different encryption standards are used to encrypt, such as 512-bit ECC, 256-bit AES and 2048-bit SSL. Accessing requires authentication and an established secure server connection (protected by end-to-end encryption). These companies' servers are often found abroad and French investigators cannot access their contents. Indeed, laws of the states in which the servers are located provide for the impossibility for investigators to examine the contents. Moreover, there are no international agreements in place and companies such as Ennetcom even specialise in hardware pre-configuration of phones, charging about 1,500 euros per device, with the camera and microphone removed; this service offers a full end-to-end messaging encryption. To find evidence, the investigator's only solution is to bypass the security of the encrypted phone.

Nevertheless, the information contained in a mobile phone is crucial to investigations and could be used to thwart terrorist attacks, recreate the itinerary of a murderer (dynamic networking), prove the presence of a phone in the perimeter of a crime scene (static networking, triggering of mobile phone cells), explain the circumstances of a plane crash, or allow the identification of victims.

In 2017, 93% of the world's inhabitants used a mobile phone, compared to 70% in 2010

(source DigitalSOBE). Seventy-two percent of people use their phone to shop online (Apple pay, Android pay) and 23 billion SMS messages are sent per minute. The encrypted messager application WhatsApp alone accounts for 42 billion messages, 1.6 billion photos and 250 million videos sent per day, and the encrypted application Telegram, 15 billion messages per day. Mobile phones are becoming more prevalent and the data they contain are often decisive for investigators.

In addition, investigators find that the more secure a phone is, the more criminals exchange information freely between members of the criminal organisation. Thus, over generations of unsecured phones, messages were often sent in deliberately non-explicit language. The contacts list on the phone did not correspond to real names of the members of the network, but to pseudonyms. Similarly, in the context of drug trafficking, the places, the quantities of the exchanges of narcotic substances and the itineraries of the go-fast (gangs of drug traffickers crossing the country in convoys of powerful cars, breaking roadblocks to deliver their drugs) were exchanged in a language difficult to exploit by the investigators. Now, investigators remark that on the latest generations of secure phones, this mode of operation is radically different. Thus, as criminal networks feel impunity by openly exchanging information via encrypted phones, the information retrieved by investigators on these types of phones is irrefutable evidence. Such evidence includes places of exchange, itineraries of go-fast, real names of the members in the contacts list, and lots of other information besides (amounts of narcotics, etc.). Investigators' interest in accessing the information contained in these ultra-secure phones is becoming a necessity for producing evidence in court.

As we have presented here, the Penal Code of the French Republic is clearly severe when cryptographic means have been used to commit criminal offences. Indeed, the Code of Criminal Procedure allows the investigator, in a well-regulated legal framework, to perform all operations that may lead to the recovery of data that can serve as evidence in court. In order to achieve these operations, the investigator may, with the authorisation of the judge or the public prosecutor, carry out the reverse-engineering of the cryptographic system used for criminal offences. But that will need the development of new techniques to confront the exponential evolution of the embedded cryptography market.

The aims of this thesis have been to find new technical means of reverse-engineering in order to provide investigators with new wherewithal for coping with the general use of cryptography. It is in a context of strong international cooperation between the United Kingdom and the French Republic that this thesis was carried out. Thanks to this co-operation we were able to share cutting-edge knowledge and implement new techniques for extracting and circumventing cryptographic operations for forensic purposes. The Information Security Group of the Ecole Normale Supérieure of Paris, the Information Security Group of Royal Holloway, University of London, the Smart Card and IoT Secu-

rity Centre of Royal Holloway, University of London, the School of Biological Sciences of Royal Holloway, University of London, and the Computer Laboratory of the University of Cambridge were at one stage of the thesis involved in joint research with the IT Forensic Department of the Forensic Sciences Institute of the French Gendarmerie (IRCGN). Sharing of information made it possible to carry out this thesis and will constitute a solid basis for future research projects.

The research steered during this thesis was often conducted with an analogy to the Duncker's Candle Problem. This cognitive test, carried out for the first time in 1945, measured the influence of the functional fixity (inability to think differently) of a subject on his ability to solve problems. This test consisted of bringing the subject into a room in which there was a table on which was placed a candle, a box of matches and a box of drawing pins. The experimenter asked the subject to fix the candle to the wall without the candle's wax falling on the table below. The typical solution of the problem was to empty the box of pins, to place the candle on the box and to use the pins to pin the box onto the wall. Most people, without success, tried to pin the candle to the wall or melt the candle to fix it with wax. It was only after several minutes that the participants understood that the box was not only a container but could also serve to hold the candle on the wall. The spirit of this thesis has been similar in many aspects to this problem. Therefore, we have repeatedly bypassed the principle of the main use of devices routinely used by investigators and made use of those operations, put together end-to-end, to solve more complex cryptographical problems inside modern mobile phones. Developing the new use of laser ablation, SEM, FIB, chemical etching and the BGA station has been made possible through a variety of research in fundamental physics, mathematics, computer science, basic chemistry and electronics.

Thus, the fundamental studies conducted in chapter 2 allowed us to understand the functioning and architecture of modern electronic components. We then presented the current investigator's technical possibilities on mobile phones to extract the data. We quantified the present limitations of traditional extraction techniques, which were often insufficient in cases of encryption algorithms. In this way, we have developed new techniques which are closer to the current investigators' needs. By doing so, we implemented in chapter 4 new chip-off methods using the physico-chemical properties of the 42Sn/58Bi eutectic alloy. This chapter was conducted thanks to the understanding of the X-ray–matter and electron–matter interactions. In chapter 5, we diverted the traditional use of industrial adhesives for forensic repairs of damaged components. We have used their different properties to produce several man-in-the-middle prototypes enabling the reverse engineering of modern security systems. In chapter 6, we relied on the properties of laser–matter interactions and their uses in the advanced forensic repair of damaged components as a result of attacks, accidents or air disasters. Finally, we demonstrated in chapter 7 that all the methods developed in this thesis led us to the realisation of

complex forensic operations, which we have successfully applied to the transplantation of modern encrypted and damaged mobile phones. This international thesis leads us to future strong collaborations with national and international universities. With the constant evolution of cryptography, the sharing of knowledge and skills is the key element allowing the investigator to counteract the use of cryptography for criminal purposes.

For the time being, investigators still have to deal with systems using classical mechanical implementations, and physics of materials. However, a new encryption system will create a much larger technological breakthrough in the years to come than we are currently experiencing. We are thinking here of quantum system of ciphering and transmission that will use the light pulses of photons and the wave–particle duality defined in the works of Albert Einstein and Louis de Broglie. These systems will take many forms: photons, coherent states, or pairs of entangled photons. The information will be transmitted by a quantum channel, which will allow the transit of the light pulses (optical fibre, propagation medium allowing the transmission of light pulses, or quantum satellite). Today, the issue that dominates our society is that of guaranteeing the privacy of communications. This securing of personal data, or sensitive data (confidential data of states or companies), is also now essential in the eyes of the general public to guarantee freedom of thought, speech and exchange. This security becomes essential now also because our mobile phone allows direct access to our bank, our car, our housing. Payments by phone become current currency. Medical consultations by videoconference on our emerging phone technology and the issue of confidentiality of medical data is a real issue. Cyber security is no longer the prerogative of states or private companies: it is becoming progressively indispensable for all of us. Even with classical approaches, mathematics and security tools, the history of cryptography shows us that it will always be bypassed at one time or another.

This is why quantum techniques have gradually made their appearance in the academic world. This new quantum security mode theoretically makes it possible to formally establish a link between the quantity of anomalies and the quantity of information intercepted, thanks to a combination of the laws of quantum physics and information theory. This kind of system is still in the development phase, although some indications lead us to think faster development can be expected. Thus, the first practical tests have already given good results for sending photons over 1,200 kilometres to 3 stations in Tibet, from the world's first quantum communication satellite, Mozi, in September 2017. This work was developed by Professor Pan Jian-Wei and his team from Hefei University, China, as part of Quantum Experiments at Space Scale program (QUESS) developed by the Chinese Space Agency [Yin and Cao, 2017]. China has thus succeeded in establishing a quantum link from space and has taken a decisive step towards an inviolable internet.

Quantum cryptography would require more theoretical skills but experts will have

to face the current encryption technologies using the properties of classical physics and mathematics. This connected world, which will be even faster with the arrival of the 5G (and then the X-G), and which will ship quantum cryptography, will gradually give way to internet browsing and highly secure exchanges. This system will be the next path in the evolution of the hyper-connectivity of our society. While this development will in many ways be beneficial for data security, it will also allow for new impunity for perpetrators, with difficulty for investigators to collect evidence or monitor website visits of a terrorist nature or child pornography. In addition, the increase in the need for the security of society, as well as for individuals (as a result of spying cases revealed by WikiLeaks, Edward Joseph Snowden, or other whistleblowers) are in line with this development. Thus, private companies that develop equipment for the general public and nation states spend astronomical sums to keep their data and security processes secret.

Forensic investigators are designated to identify disaster victims, to combat international child-pornography networks, to assemble evidence against crime perpetrators, to combat international drug trafficking and to help dismantle terrorist networks. All these missions are not exhaustive and investigators must expect new threats in an ultra-connected world. In the short run, we could mention the remote control of a pacemaker, a vehicle or an airliner that will create certain difficulties in post-disaster expertise.

If it is possible for a doctor to remotely program a pacemaker, how could we ensure that the death of a person wearing a pacemaker is natural, or a digital homicide committed by an attacker taking remote control of pacemeker via a security breach? There are 5 million people around the world who are wearing pacemakers and 400,000 in France (and 60,000 new patients every year in France). Should the investigator consider each death of a person carrying a pacemaker as suspect? In the United States, the Food and Drug Administration (FDA), on August 29th 2017, launched a call for emergency update of 465,000 pacemakers manufactured by St. Jude Medical and Abbott following a security breach on some of the most prevalent models such as: Accent SR RF, Accent ST, Assurity MRI, Accent DR RF, Anthem RF, Quadra Allure MP RF and Quadra Allure MP.

The same principle of remote programming or hacking will apply more and more to "connected" cars or airplanes with the potential for causing an accident or crash. As evidence, 1,200 patents were filed between 2012 and 2016 concerning connected and autonomous vehicles. In 2017 Audi released its first autonomous commercial vehicle driving up to 60 km/h, the Audi A8 (the first car in the world to ship level 3 autonomous driving systems). Among the most advanced brands in the field are Google, Daimler, BMW, General Motors, Apple, Facebook, Microsoft, Amazon and Uber. Will it be necessary to consider and assess numerically each vehicle relating to a fatal car accident? How to prove that a remote control take-over has been done?

In the years to come, could digital crime be a perfectly undetectable crime, the perfect crime? The question will arise quickly and investigators must already anticipate it.

Bibliography

[Adler et al., 1995] Adler, E., DeBrosse, J. K., Geissler, S. F., Holmes, S. J., Jaffe, M. D., Johnson, J. B., Koburger, C., Lasky, J. B., and Lloyd, B. (1995). The evolution of IBM CMOS DRAM technology. *IBM Journal of Research and Development*, 39(1.2):167–188.

[Aller et al., 1996] Aller, L., Appenzeller, I., Baschek, B., Butler, K., De Loore, C., Duerbeck, H., El Eid, M., Fink, H., Herczeg, T., and Richtler, T. (1996). Landolt-Börnstein: Numerical data and functional relationships in science and technology–new series, Springer Verlag Berlin, 458 pages.

[Androulidakis, 2016] Androulidakis, I. (2016). Mobile phone security and forensics. Springer International Publishing Switzerland, 103 pages.

[Ashenden, 2010] Ashenden, P. J. (2010). *The designer's guide to VHDL*, volume 3. Morgan Kaufmann, 909 pages.

[Asif et al., 2005] Asif, A., Shi, W., Shen, X., and Nie, K. (2005). Physical and thermal properties of UV curable waterborne polyurethane dispersions incorporating hyperbranched aliphatic polyester of varying generation number. *Polymer*, 46(24):11066–11078.

[Ay et al., 2012] Ay, F., Wörhoff, K., de Ridder, R. M., and Pollnau, M. (2012). Focused-ion-beam nanostructuring of dielectric layers for photonic applications. *Journal of Micromechanics and Microengineering*, 22(10):105008–105013.

[Ayers et al., 2014] Ayers, R., Brothers, S., and Jansen, W. (2014). In *Guidelines on Mobile Device Forensics, NIST,* 85 pages.

[Ayers et al., 2007] Ayers, R., Jansen, W., Delaitre, A. M., and Moenner, L. (2007). Cell phone forensic tools: an overview and analysis update. *NIST Interagency/Internal Report-7387, 165 pages.*

[BBC-News, 2015a] BBC-News (2015a). Bus crash kills at least 42 in Gironde region of France. [online] `http://www.bbc.com/news/world-europe-34612720`.

[BBC-News, 2015b] BBC-News (2015b). Paris attacks: Key questions after Abaaoud killed. [online] `http://www.bbc.com/news/world-europe-34866144`.

[BBC-News, 2016] BBC-News (2016). Nice attack: At least 84 killed by lorry at Bastille Day celebrations. [online] `http://www.bbc.com/news/world-europe-36800730`.

[BBC-News, 2017] BBC-News (2017). Germanwings crash leaves unanswered questions. [online] http://www.bbc.com/news/world-europe-32084956.

[Bez et al., 2003] Bez, R., Camerlenghi, E., Modelli, A., and Visconti, A. (2003). Introduction to flash memory. *Proceedings of the IEEE*, 91(4):489–502.

[Bhavnagarwala et al., 2001] Bhavnagarwala, A. J., Tang, X., and Meindl, J. D. (2001). The impact of intrinsic device fluctuations on CMOS SRAM cell stability. *IEEE journal of Solid-state circuits*, 36(4):658–665.

[BlackBerry-Entreprise-Server, 2014] BlackBerry-Entreprise-Server (2014). Blackberry Enterprise Server for Microsoft Exchange, Security Technical Overview, 189 pages. [online] https://help.blackberry.com/en/bes5-for-exchange/current/sto-pdf/BlackBerry_Enterprise_Server_for_Microsoft_Exchange-Security_Technical_Overview-1330547681607-5.0.4-en.pdf.

[Boeuf, 2017] Boeuf, F. (2017). Logic devices challenges and opportunities in the nano era. *Nanoelectronics: Materials, Devices, Applications, Volume 2,* 640 pages.

[Born, 1927] Born, M. (1927). Born–Oppenheimer approximation. *Ann. Physik*, 84:457–484.

[Braga et al., 2009] Braga, M., Oliveira, J., Malheiros, L., and Ferreira, J. (2009). Phase field simulations in miscibility gaps. *Calphad*, 33(1):237–243.

[Breeuwsma, 2006] Breeuwsma, I. M. (2006). Forensic imaging of embedded systems using JTAG (boundary-scan). *Digital Investigation*, 3(1):32–42.

[Breeuwsma et al., 2007] Breeuwsma, M., De Jongh, M., Klaver, C., Van Der Knijff, R., and Roeloffs, M. (2007). Forensic data recovery from flash memory. *Small Scale Digital Device Forensics Journal*, 1(1):1–17.

[Brian, 2005] Brian, C. (2005). *File System Forensic Analysis.* Pearson Education, Inc., 382 pages.

[Brian, 1990] Brian, S. (1990). The generation and characterization of electron and hole traps created by hole injection during low gate voltage hot-carrier stressing of n-MOS transistors. *IEEE Transactions on Electron Devices*, 37(8):1869–1876.

[Brothers, 2009] Brothers, S. (2009). How cell phone "forensic" tools actually work–proposed leveling system. In *Mobile Forensics World Conference, Chicago, Illinois.*

[Bushnell and Agrawal, 2004] Bushnell, M. and Agrawal, V. (2004). *Essentials of electronic testing for digital, memory and mixed-signal* VLSI *circuits*, volume 17. Springer Science & Business Media, 690 pages.

[Buxbaum, 2008] Buxbaum, G. (2008). *Industrial inorganic pigments*. John Wiley & Sons, 289 pages.

[Campardo et al., 2010] Campardo, G., Micheloni, R., and Novosel, D. (2010). VLSI-*Design of Non-Volatile Memories*. Springer, 582 pages.

[Carley, 1973] Carley, A. (1973). Method for electronic lithography. US Patent 3,741,118.

[Chen et al., 2009] Chen, B., Tuntasood, P., and Fan, D.-T. (2009). NOR flash memory. US Patent 7,598,561.

[Chen, 1990] Chen, J. Y. (1990). CMOS devices and technology for VLSI. *Englewood Cliffs, NJ, Prentice Hall,* 362 pages.

[Cohen-Tannoudji et al., 1973] Cohen-Tannoudji, C., Diu, B., and Laloë, F. (1973). *Mécanique quantique*. Number vol. 1 in Collection Enseignement des Sciences. Masson, 1493 pages.

[Courbon et al., 2016] Courbon, F., Skorobogatov, S., and Woods, C. (2016). Reverse engineering flash EEPROM memories using scanning electron microscopy. In *International Conference on Smart Card Research and Advanced Applications*. Springer, 1(1):57–72.

[Crocker et al., 1973] Crocker, M., Fidler, R., and Smith, R. (1973). The characterization of eutectic structures. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 335(1600):15–37.

[Cui et al., 2014] Cui, H.-W., Fan, Q., and Li, D.-S. (2014). Surface functionalization of micro silver flakes and their application in electrically conductive adhesives for electronic package. *International Journal of Adhesion and Adhesives*, 48:177–182.

[Cui et al., 2013] Cui, H.-W., Li, D.-S., Fan, Q., and Lai, H.-X. (2013). Electrical and mechanical properties of electrically conductive adhesives from epoxy, micro-silver flakes, and nano-hexagonal boron nitride particles after humid and thermal aging. *International Journal of Adhesion and Adhesives*, 44:232–236.

[Czochralski, 1918] Czochralski, J. (1918). A new method for the measurement of the crystallization rate of metals. *Zeitschrift Für Physikalische Chemie*, 92:219–221.

[Falat et al., 2007] Falat, T., Artur, W., and Jana, K. (2007). Numerical approach to characterization of thermally conductive adhesives. *Microelectronics Reliability*, 47(2-3):342–346.

[Feigelson, 2004] Feigelson, R. (2004). *50 Years Progress in Crystal Growth: A Reprint Collection*. Elsevier, 268 pages.

[Felba, 2011] Felba, J. (2011). Thermally conductive adhesives in electronics. *Advanced Adhesives in Electronics: Materials, Properties and Applications,* Elsevier, 1:15–52.

[Fiorillo, 2009] Fiorillo, S. (2009). Theory and practice of flash memory mobile forensics. In *Australian Digital Forensics Conference,* 37 pages.

[Flannery et al., 1987] Flannery, B. P., Deckman, H. W., Roberge, W. G., and d'Amico, K. L. (1987). Three-dimensional X-ray microtomography. *Science*, 237(4821):1439–1444.

[Fock, 1930] Fock, V. (1930). Näherungsmethode zur lösung des quantenmechanischen mehrkörperproblems. *Zeitschrift Für Physik, A Hadrons and Nuclei*, 61(1):126–148.

[Freyssinet, 2012] Freyssinet, E. (2012). *La cybercriminalité en mouvement.* Lavoisier, 232 pages.

[Gamow, 1985] Gamow, G. (1985). *Thirty years that shook physics: The story of quantum theory.* Courier Corporation, 272 pages.

[Gordon et al., 1970] Gordon, R., Bender, R., and Herman, G. T. (1970). Algebraic Reconstruction Techniques (ART) for three-dimensional electron microscopy and X-ray photography. *Journal of Theoretical Biology*, 29(3):471–481.

[Guido et al., 2016] Guido, M., Buttner, J., and Grover, J. (2016). Rapid differential forensic imaging of mobile devices. *Digital Investigation*, 18:46–54.

[Guzmán-Verri and Lew Yan Voon, 2007] Guzmán-Verri, G. and Lew Yan Voon, L. (2007). Electronic structure of silicon-based nanostructures. *Physical Review B*, 76:1311–1321.

[Ha et al., 2017] Ha, D., Yang, C., Lee, J., Lee, S., Lee, S., Seo, K.-I., Oh, H., Hwang, E., Do, S., and Park, S. (2017). Highly manufacturable 7nm finfet technology featuring EUV lithography for low power and high performance applications. In *VLSI Technology, 2017 Symposium on IEEE, T68-T69.*

[Halderman et al., 2009] Halderman, A., Schoen, S. D., Heninger, and Clarkson, W. (2009). Lest we remember: Cold boot attacks on encryption keys. In *USENIX Association 17th USENIX Security 45*, volume 38, 91–98.

[Harari et al., 1994] Harari, E., Norman, R. D., and Mehrotra, S. (1994). Flash EEPROM system. US Patent 5,297,148.

[Hartree, 1928] Hartree, D. R. (1928). The wave mechanics of an atom with a non-coulomb central field. part I. theory and methods. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24. Cambridge University Press, 89–110.

[He et al., 2000] He, Y., Moreira, B. E., Overson, A., Nakamura, S. H., Bider, C., and Briscoe, J. F. (2000). Thermal characterization of an epoxy-based underfill material for flip chip packaging. *Thermochimica Acta*, 357-358(Supplement C):1–8.

[Heckmann et al., 2016] Heckmann, T., Souvignet, T., Lepeer, S., and Naccache, D. (2016). Low-temperature low-cost 58 bismuth –42 tin alloy forensic chip re-balling and re-soldering. *Digital Investigation*, 19:60–68.

[Henins, 1964] Henins, I. (1964). Precision density measurement of silicon. *Journal of Research of the National Bureau of Standards*, 68A(5):529–533.

[Herbts and Hunger, 2004] Herbts, W. and Hunger, K. (2004). *Industrial Organic Pigments, Production, Properties, Applications*. Wiley-VCH, 3rd edition, 659 pages.

[Ho, 2005] Ho, I.-M. (2005). Electronic memory with tri-level cell pair. US Patent App. 11/049,236.

[Hoog, 2011] Hoog, A. (2011). In *Android Forensics: Investigation, Analysis and Mobile Security for Google Android.* Syngress Publishing, 432 pages.

[Horowitz and Hill, 1989] Horowitz, P. and Hill, W. (1989). *The art of electronics*. Cambridge Univ. Press, 1125 pages.

[House, 2017] House, J. E. (2017). *Fundamentals of quantum mechanics*. Academic Press, 372 pages.

[JEDEC, 2010] JEDEC (2010). JEDEC standard, Embededed MultiMediaCard(e-MMC) e-MMC/card product standard, high capacity, including reliable write, boot, sleep modes, dual data rate, multiple partitions supports, security enhancement, background operation and high priority interrupt (MMCA, 4.41). JEDEC Solid State Technology Association, 166 pages.

[Jongh, 2014] Jongh, M. D. (2014). Embedded update 2014. Prague ENFSI FITWG annual meeting, 18 pages.

[Jonkers, 2010] Jonkers, K. (2010). The forensic use of mobile phone flasher boxes. *Digital Investigation*, 6(3):168–178.

[Kaipio et al., 2000] Kaipio, J. P., Kolehmainen, V., Somersalo, E., and Vauhkonen, M. (2000). Statistical inversion and monte carlo sampling methods in electrical impedance tomography. *Inverse problems*, 16(5):1487.

[Kanda et al., 2008] Kanda, K., Koyanagi, M., Yamamura, T., Hosono, K., Yoshihara, M., Miwa, T., Kato, Y., Mak, A., Chan, S. L., and Tsai, F. (2008). A 120mm 2 16Gb 4-MLC NAND flash memory with 43nm CMOS technology. In *Solid-State*

Circuits Conference, 2008. ISSCC 2008. Digest of Technical Papers. IEEE International, 430–625.

[Keizer, 2016] Keizer, G. (2016). FBI chief shoots down theory that NAND mirroring will be used to crack terrorist's iPhone. [online] `https://www.computerworld.com/article/3048243/apple-ios/`.

[Kerisit et al., 2014a] Kerisit, F., Domenges, B., and Obein, M. (2014a). Comparative study on decapsulation for copper and silver wire-bonded devices. In *40th International Symposium for Testing and Failure Analysis, ASM International,* 87-93.

[Kerisit et al., 2014b] Kerisit, F., Domenges, B., and Obein, M. (2014b). Comparative study on decapsulation for copper and silver wire-bonded devices. In *40th International Symposium for Testing and Failure Analysis, ASM International*, page 87–93.

[Kim et al., 2002] Kim, J. K., Kim, W. H., and Lee, D. H. (2002). Adhesion properties of UV crosslinked polystyrene-block-polybutadiene-block-polystyrene copolymer and tackifier mixture. *Polymer*, 43(18):5005–5010.

[Kim et al., 2007] Kim, K., Hong, D., Chung, K., and Ryou, J.-C. (2007). Data acquisition from cell phone using logical approach. *Proceedings of the World Academy of Science, Engineering and Technology*, 26:29–32.

[Kishi et al., 2016] Kishi, H., Tanaka, S., Nakashima, Y., and Saruwatari, T. (2016). Self-assembled three-dimensional structure of epoxy/polyethersulphone/silver adhesives with electrical conductivity. *Polymer*, 82:93–99.

[Kor et al., 2014] Kor, H., Liu, Q., Siah, Y. W., and Gan, C. L. (2014). Laser focus adaptation for decapsulation of copper wirebonded devices. In *40th International Symposium for Testing and Failure Analysis, ASM International,* 94–99.

[Leona et al., 2004] Leona, M., Casadio, F., Bacci, M., and Picollo, M. (2004). Identification of the pre-columbian pigment mayablue on works of art by noninvasive uv-vis and raman spectroscopic techniques. *Journal of the American Institute for Conservation*, 43(1):39–54.

[Li and Wong, 2006] Li, Y. and Wong, C. (2006). Recent advances of conductive adhesives as a lead-free alternative in electronic packaging: materials, processing, reliability and applications. *Materials Science and Engineering: R: Reports*, 51(1):1–35.

[Liu et al., 2013] Liu, Q., Vinet, M., Gimbert, J., Loubet, N., Wacquez, R., Grenouillet, L., Le Tiec, Y., Khakifirooz, A., Nagumo, T., and Cheng, K. (2013). High performance UTBB FDSOI devices featuring 20nm gate length for 14nm node and beyond. In *Electron Devices Meeting, IEEE International,* 2–9.

[Liu, 1998] Liu, W. (1998). *Handbook of III-V heterojunction bipolar transistors*. Wiley, 1284 pages.

[Lundstrom, 1997] Lundstrom, M. (1997). Elementary scattering theory of the Si MOSFET. *IEEE Electron Device Letters*, 18(7):361–363.

[Luo et al., 2016] Luo, J., Cheng, Z., Li, C., Wang, L., Yu, C., Zhao, Y., Chen, M., Li, Q., and Yao, Y. (2016). Electrically conductive adhesives based on thermoplastic polyurethane filled with silver flakes and carbon nanotubes. *Composites Science and Technology*, 129:191–197.

[Ma and Kan, 2017] Ma, Y. and Kan, E. (2017). Bipolar transistors in logic CMOS processes. In *Non-logic Devices in Logic Processes*. Springer, 125–132.

[Mathieu and Fanet, 2009] Mathieu, H. and Fanet, H. (2009). *Physique des semiconducteurs et des composants électroniques–6ème édition: Cours et exercices corrigés*. Dunod, 830 pages.

[Maunder, 1993] Maunder, C. (1993). Standard test access port and boundary-scan architecture. *IEEE Standards Board New York 1149.1,* 212 pages.

[Menezes et al., 1997] Menezes, A. J., Vanoorschot, P. C., and Vanstone, S. A. (1997). *Handbook of Applied Cryptography*. The CRC Press series on discrete mathematics and its applications, 816 pages.

[Miao and Duh, 2001] Miao, H.-W. and Duh, J.-G. (2001). Microstructure evolution in Sn–Bi and Sn–Bi–Cu solder joints under thermal aging. *Materials chemistry and physics*, 71(3):255–271.

[Mishiro et al., 2002] Mishiro, K., Ishikawa, S., Abe, M., Kumai, T., Higashiguchi, Y., and Tsubone, K. (2002). Effect of the drop impact on BGA/CSP package reliability. *Microelectronics Reliability*, 42(1):77–82.

[Moon et al., 2004] Moon, S.-H., Swearingen, S., and Foster, M. D. (2004). Scanning probe microscopy study of dynamic adhesion behavior of polymer adhesive blends. *Polymer*, 45(17):5951–5959.

[Morancho, 2004] Morancho, L. M. (2004). *Nouvelle méthode de test en rétention de données de mémoires non volatiles*. PhD thesis, Institut National Polytechnique de Toulouse, 105 pages.

[Ngô, 2005] Ngô, C. (2005). *Physique quantique*. DUNOD, 448 pages.

[Nielsen and Girgis, 2000] Nielsen, C. and Girgis, S. (2000). *WPI 0.5mm* CMOS *Standard cell Library Databook*. Microelectronics Group, 90 pages.

[Ning et al., 1981] Ning, T. H., Isaac, R., Solomon, P., Tang, D.-L., Yu, H.-N., Feth, G., and Wiedmann, S. (1981). Self-aligned bipolar transistors for high-performance and low-power-delay VLSI. *IEEE Transactions on Electron Devices*, 28(9):1010–1013.

[Okamoto, 2000] Okamoto, H. (2000). Desk handbook: Phase diagrams for binary alloys. *Materials Park, OH: ASM International,* 828 pages.

[Okura et al., 2000] Okura, J., Shetty, S., Ramakrishnan, B., Dasgupta, A., Caers, J., and Reinikainen, T. (2000). Guidelines to select underfills for flip chip on board assemblies and compliant interposers for chip scale package assemblies. *Microelectronics Reliability*, 40(7):1173–1180.

[Pelgrom et al., 1989] Pelgrom, M. J., Duinmaijer, A. C., and Welbers, A. P. (1989). Matching properties of MOS transistors. *IEEE Journal of solid-state circuits*, 24(5):1433–1439.

[Planes et al., 2012] Planes, N., Weber, O., Barral, V., Haendler, S., Noblet, D., Croain, D., Bocat, M., Sassoulas, P., Federspiel, X., and Cros, A. (2012). 28nm FDSOI technology platform for high-speed low-voltage digital applications. In *VLSI Technology (VLSIT), Symposium on,* 133–134.

[Predel, 1991] Predel, B. (1991). Phase equilibria, crystallographic and thermodynamic data of binary alloys. Landolt-börstein, new series. *Group IV*, Volume 5, 254 pages.

[Puttlitz and Stalter, 2004] Puttlitz, K. J. and Stalter, K. A. (2004). *Handbook of lead-free solder technology for microelectronic assemblies.* CRC Press, 1048 pages.

[Rechchach, 2011] Rechchach, M. (2011). *Etude thermodynamique des systèmes ternaires* I*n*-S*n*-Z*n,* B*i*-S*n*-C*u et* B*i*-S*n*-N*i comme matériaux pour soudre sans plomb.* PhD thesis, Université Mohammed V-Agdal, Faculté des Sciences, Rabat, 206 pages.

[Regnery and Souvignet, 2013] Regnery, M. and Souvignet, T. (2013). La récupération de données sur SSD : un défi. In *MISC n066,* 19 pages.

[Rodríguez et al., 2009] Rodríguez, A. H., Trallero-Giner, C., Duque, C., and Vázquez, G. (2009). Optical transition in self-assembled InAs/GaAs quantum lens under high hydrostatic pressure. *Journal of Applied Physics*, 105(4):83–88.

[Samsung, 2015] Samsung (2015). eMMC to UFS: How NAND memory for mobile products is evolving, 5 pages. [online] `https://news.samsung.com/global/emmc-to-ufs-how-nand-memory-for-mobile-products-is-evolving`.

[Schrödinger, 1926] Schrödinger, E. (1926). Quantisierung als eigenwertproblem. *Annalen der physik*, 385(13):437–490.

[Seeger, 2013] Seeger, K. (2013). *Semiconductor physics.* Springer Science & Business Media, 504 pages.

[Shiner et al., 1980] Shiner, R., Caywood, J., and Euzent, B. (1980). Data retention in EPROMs. In *Reliability Physics Symposium. IEEE 18th Annual,* 238–243.

[Skorobogatov, 2016] Skorobogatov, S. (2016). The bumpy road towards iPhone 5c NAND mirroring. *CoRR*, abs/1609.04327, 10 pages.

[Song et al., 2004] Song, J.-M., Chang, Y.-L., Lui, T.-S., and Chen, L.-H. (2004). Vibration fracture behavior of Sn–Bi solder alloys with various Bi contents. *Materials Transactions*, 45(3):666–672.

[Staller, 2010] Staller, K. D. (2010). Low temperature plasma decapsulation of copper-wire-bonded and exposed copper metallization devices. In *36th International Symposium for Testing and Failure Analysis (ISTFA),* 127–132.

[Stulen and Sweeney, 1999] Stulen, R. H. and Sweeney, D. W. (1999). Extreme ultraviolet lithography. *IEEE Journal of Quantum Electronics*, 35(5):694–699.

[Terao et al., 1991] Terao, A., Flandre, D., Lora-Tamayo, E., and Van de Wiele, F. (1991). Measurement of threshold voltages of thin-film accumulation-mode PMOS/SOI transistors. *IEEE Electron Device Letters*, 12(12):682–684.

[Thomas and Moorby, 2008] Thomas, D. and Moorby, P. (2008). *The Verilog® Hardware Description Language.* Springer Science & Business Media, 386 pages.

[Time, 2015] Time (2015). Inside apple CEO Tim Cook's fight with the FBI. [online] `http://time.com/4262480/tim-cook-apple-fbi-2/`.

[Torres et al., 2012] Torres, A., Hernández, L., and Domínguez, O. (2012). Effect of antimony additions on corrosion and mechanical properties of Sn–Bi eutectic lead-free solder alloy. *Scientific Research Publishing*, 3(6):355–362.

[Ueno and Uchiumi, 2010] Ueno, K. and Uchiumi, S. (2010). Nand flash memory. US Patent App. 12/723,112.

[Van Der Knijff, 2002] Van Der Knijff, R. (2002). Embedded systems analysis. In *Chapter 11 of "Handbook of Computer Crime Investigation: Forensic Tools and Technology", E. Casey (Ed.),* 448 pages.

[Vianco et al., 2004] Vianco, P., Rejent, J., and Grant, R. (2004). Development of Sn-based, low melting temperature Pb-free solder alloys. *Materials Transactions*, 45(3):765–775.

[Vidas, 2010] Vidas, T. (2010). Volatile memory acquisition via warm boot memory survivability. In *2010 43rd Hawaii International Conference on System Sciences,* 1–6.

[Volinsky et al., 2004] Volinsky, A. A., Rice, L., Qin, W., and Theodore, N. D. (2004). FIB failure analysis of memory arrays. *Microelectronic Engineering*, 75(1):3–11.

[Ward and Dutton, 1978] Ward, D. E. and Dutton, R. W. (1978). A charge-oriented model for MOS transistor capacitances. *IEEE Journal of Solid-State Circuits*, 13(5):703–708.

[Waring and Hallas, 2013] Waring, D. R. and Hallas, G. (2013). *The chemistry and application of dyes*. Springer Science & Business Media, 430 pages.

[Weste et al., 2005] Weste, N., Harris, D., and Banerjee, A. (2005). CMOS VLSI design. *A circuits and systems perspective*, Volume 11, 867 pages.

[Yin and Cao, 2017] Yin, J. and Cao, Y. (2017). Satellite-based entanglement distribution over 1200 kilometers. *Science*, 356(6343):1140–1144.

[Zeng and Tu, 2002] Zeng, K. and Tu, K.-N. (2002). Six cases of reliability study of Pb-free solder joints in electronic packaging technology. *Materials science and engineering: R: reports*, 38(2):55–105.

[Zhang et al., 2014] Zhang, J., Zivic, N., Dumur, F., Xiao, P., Graff, B., Fouassier, J. P., and Gigmes, D. (2014). UV-violet-blue LED induced polymerizations: Specific photoinitiating systems at 365, 385, 395 and 405nm. *Polymer*, 55(26):6641–6648.

[Zhang et al., 2013] Zhang, Z., Luo, S., and Wong, C. (2013). Flip chip underfill: Materials, process, and reliability. In *Advanced Flip Chip Packaging, Springer,* 155–199.

Appendices

A

Calculating Methods: Silicon Energy Bands' Structure

The Schrödinger equation, developed for the first time by the Austrian physicist Erwin Schrödinger in 1926 [Schrödinger, 1926], is the fundamental equation of quantum mechanics that describes the evolution over time of a non-relativistic mass particle, and has the same role than the fundamental equation of dynamics in classical mechanics.

A.1 Schrödinger Equation

Stationary energy states and system wave functions are given by the solutions of the Schrödinger equation [Schrödinger, 1926] (Equation A.1):

$$H\Psi_{nlm}(r,\Theta,\Phi) = E_{nl}\Psi_{nlm}(r,\Theta,\Phi) \tag{A.1}$$

with $\Psi_{nlm}(r,\Theta,\Phi)$ the spherically symmetric potential, (n) the main quantum number, (l) the orbital quantum number and (m) the magnetic quantum number.

A silicon crystal is composed of a very large number of interacting particles: electrons and atomic nuclei. So if we take into account all the interactions found in the silicon crystal we can write the Hamiltonian of the system in the form (Equation A.2):

$$H_{silicon} = E_{c_n} + E_{c_e} + V_{e-e} + V_{e-n} + V_{n-n} \tag{A.2}$$

with E_{c_e} and E_{c_n} the kinetic energies of electrons and nuclei and V_{e-e}, V_{e-n} and V_{n-n} the electron–electron, electron–nucleus and nucleus–nucleus interaction energies.

The stationary states of energy and the wave functions of the system are given by the solutions of the silicon's Schrödinger equation (Equation A.3):

$$[E_{c_n} + E_{c_e} + V_{e-e} + V_{e-n} + V_{n-n}]\Psi_{nlm}(r,\Theta,\Phi) = E_{nl}\Psi_{nlm}(r,\Theta,\Phi) \tag{A.3}$$

From this complex equation, several approximations can be made. The first is to limit the interaction between particles by taking only the most important term: the Coulomb interaction. But physicists often use the Born–Oppenheimer approximation to help solve the Schrödinger equation.

A.2 Born–Oppenheimer and Adiabatic Approximations

The Born–Oppenheimer approximation was developed for the first time in 1927 [Born, 1927] and allows us to separate the equation of the eigenvalues of the nuclei from that of the electrons (Equation A.4).

$$H\Psi(R_n, r_e)) = E_{nl}\Psi(R_n, r_e) \tag{A.4}$$

with R_n the coordinates of the nuclei and r_e represents those of the electrons.

The Born–Oppenheimer approximation shows that the eigenstates of the global system are wave functions which are the products of the electronic wave function and the nuclear wave function. Using this approximation in the case of silicon allows us to rewrite the Schrödinger equation as (Equation A.5):

$$[E_{c_n} + E_{c_e} + V_{e-e} + V_{e-n} + V_{n-n}]\Psi_n(R_n)\Psi_e(R_n, r_e) = E.\Psi_n(R_n).\Psi_e(R_n, r_e) \tag{A.5}$$

The adiabatic approximation finally allows an independent equation (Equation A.6):

$$[E_{c_e} + V_{e-e} + V_{e-n}]\Psi_e(R_n, r_e) = E.\Psi_n(R_n).\Psi_e(R_n, r_e) \tag{A.6}$$

A.3 Hartree–Fock Approximation

The Hartree–Fock [Hartree, 1928, Fock, 1930] methods are methods of approximate resolution of the Schrödinger equation of a multi-body quantum system. This approximation consists of globalising individual electron–electron interactions and writing that each electron evolves in the average potential of all other electrons of silicon. Thereby the Hamiltonian of Hartree–Fock is defined by (Equation A.7):

$$H_{Hartree-Fock} = -\frac{\hbar^2}{2m}\nabla^2 - \sum_l \frac{Z_l.e^2}{|r_e - R_{nl}|} + P_{Coulomb} + P_{e-e} \tag{A.7}$$

with $\sum_l \frac{Z_l.e^2}{|r_e - R_{nl}|}$ the potential energy of electron–nuclei interactions, $P_{Coulomb}$ and P_{e-e} the electron–electron interactions.

Finally, in the context of the Born–Oppenheimer approximation and the Hartree–Fock

approximation we can write the crystalline potential as (Equation A.8):

$$V_{c_{Born-Oppenheimer-Hartree-Fock}}(\vec{r}) = -\sum_{l} \frac{Z_l.e^2}{|r_e - R_{nl}|} + P_{Coulomb} + P_{e-e} \tag{A.8}$$

The Schrödinger equation, in the case of silicon crystal, is a function of the crystalline potential of the Born–Oppenheimer and Hartree–Fock approximations (Equation A.9):

$$[-\frac{\hbar^2}{2m}\nabla^2 + V_c(\vec{r})]\Psi_n(\vec{k},\vec{r}) = E_n(\vec{k})\Psi_n(\vec{k},\vec{r}) \tag{A.9}$$

with $V_c(\vec{r})$ the crystalline potential seen by the electrons.

B

Main Forensic File Signatures Table

File	Header signature	Footer signature
jpeg	FF D8	FF D9
gif	47 49 46 38	00 3B
png	89 50 4E 47 0D 0A 1A 0A	49 45 4E 44
html	3C 48 54 4D 4C 3E	3C 2F 68 74 6D 6C 3E
pdf	25 50 44 46 2D 31 2E	25 25 45 4F 46
doc	D0 CF 11 E0 A1 B1 1A E1	57 6F 72 64 2E 44 6F 63 75 6D 65 6E 74 2E
zip	50 4B 03 04 14	50 4B 05 06 00

Table B.1: File Signatures: Headers and Footers

File type	start
mp3	49 44 33 2E
excel	D0 CF 11 E0
tar	1F 8B 08 00
ppt	D0 CF 11 E0
rar	52 61 72 2
wmv	30 26 B2 7
pgd	50 47 50 64 4D 41 49 4E
png	89 50 4E 47 0D 0A 1A 0A
pptx	50 4B 03 04
mls	4D 4C 53 57

Table B.2: File Starts: File Identifications

Résumé

Avec l'arrivée des dernières générations de téléphones chiffrés (BlackBerry PGP, iPhone), l'extraction des données par les experts est une tâche de plus en plus complexe et devient un véritable défi notamment après une catastrophe aérienne ou une attaque terroriste. Dans cette thèse, nous avons développé des attaques physiques sur systèmes cryptographiques à des fins d'expertises judiciaires.

Une nouvelle technique de re-brasage à basse température des composants électroniques endommagés, utilisant un mélange eutectique 42Sn/58Bi, a été développée. Nous avons exploité les propriétés physico-chimiques de colles polymères et les avons utilisées dans l'extraction de données chiffrées. Une nouvelle technique a été développée pour faciliter l'injection et la modification à haute-fréquence des données. Le prototype permet des analyses en temps réel des échanges processeur-mémoire en attaque par le milieu. Ces deux techniques sont maintenant utilisées dans des dispositifs d'attaques plus complexes de systèmes cryptographiques.

Nos travaux nous ont mené à sensibiliser les colles polymères aux attaques laser par pigmentation. Ce processus permet des réparations complexes avec une précision laser de l'ordre de 15 micromètres. Cette technique est utilisable en réparations judiciaires avancées des crypto-processeurs et des mémoires.

Ainsi, les techniques développées, mises bout à bout et couplées avec des dispositifs physiques (tomographie 3D aux rayons X, MEB, laser, acide fumant) ont permis de réussir des transplantations judiciaires de systèmes chiffrés en conditions dégradées et appliquées pour la première fois avec succès sur les téléphones BlackBerry chiffrés à l'aide de PGP.

Mots-clés

Retro-Conception Matérielle, Extractions Physiques, Attaques par le Milieu, Tomographie 3D aux Rayons X, Interactions Laser-Matière et Electron-Matière, Attaques Chimiques, Transplantations Judiciaires.

Abstract

When considering the latest generation of encrypted mobile devices (BlackBerry's PGP, Apple's iPhone), data extraction by experts is an increasingly complex task. Forensic analyses even become a real challenge following an air crash or a terrorist attack. In this thesis, we have developed physical attacks on encrypted systems for the purpose of forensic analysis.

A new low-temperature re-soldering technique of damaged electronic components, using a 42Sn/58Bi eutectic mixture, has been developed. Then we have exploited the physico-chemical properties of polymer adhesives and have used them for the extraction of encrypted data. A new technique has been developed to facilitate injection and high-frequency data modification. By a man-in-the-middle attack, the prototype allows analysing, in real-time, the data exchanges between the processor and the memory. Both techniques are now used in more complex attacks of cryptographic systems.

Our research has led us to successfully sensitise polymer adhesives to laser attacks by pigmentation. This process allowed complex repairs with a laser with 15 micrometres precision and has been used in advanced forensic repair of crypto-processors and memory chips.

Finally, the techniques developed in this thesis, put end-to-end and coupled with physical devices (X-ray 3D tomography, laser, SEM, fuming acids), have made it possible to have successful forensic transplants of encrypted systems in degraded conditions. We have successfully applied them, for the first time, on PGP-encrypted BlackBerry mobile phone.

Keywords

Hardware Reverse Engineering, Physical Extractions, Man-In-The-Middle Attacks, X-ray 3D Tomography, Laser-Matter and Electron-Matter Interactions, Chemical Attacks, Forensic Transplantations.