



BRNO UNIVERSITY OF TECHNOLOGY
VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ



FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF INTELLIGENT SYSTEMS

FAKULTA INFORMAČNÍCH TECHNOLOGIÍ
ÚSTAV INTELIGENTNÍCH SYSTÉMŮ

RECONSTRUCTION OF DATA FROM DAMAGED CD/DVD

REKONSTRUKCE DAT Z POŠKOZENÉHO CD/DVD

BACHELOR'S THESIS

BAKALÁŘSKÁ PRÁCE

AUTHOR

AUTOR PRÁCE

JORDÁN JAROLÍM

SUPERVISOR

VEDOUČÍ PRÁCE

doc. Ing. Dipl.-Ing **MARTIN DRAHANSKÝ, Ph.D.**

BRNO 2016

Vysoké učení technické v Brně - Fakulta informačních technologií

Ústav inteligentních systémů

Akademický rok 2015/2016

Zadání bakalářské práce

Řešitel: **Jarolím Jordán**

Obor: Informační technologie

Téma: **Rekonstrukce dat z poškozeného CD/DVD**
Reconstruction of Data from Damaged CD/DVD

Kategorie: Umělá inteligence

Pokyny:

1. Prostudujte literaturu týkající se způsobu záznamu a uchovávání informací na CD a DVD. Zaměřte se i na práce, které se zabývají samoopravnými kódy a opravou poškozených částí záznamů.
2. Připravte si vzorky datových nosičů, u nichž si před řízeným poškozením uložíte jejich obraz.
3. Pomocí mikroskopu (optického či elektronového) nasnímejte kusy poškozeného média (CD/DVD) a uložte je v obrazové podobě pro následné spojení do jednoho celku.
4. Využijte stávající, příp. vlastní, aplikaci pro složení nasnímaných částí datového nosiče do jednoho celku. Nasnímané části pomocí této aplikace spojte do jednoho celku.
5. Navrhněte a implementujte aplikaci pro analýzu dílčích částí poškozeného datového nosiče za účelem přečtení dat z tohoto nosiče.
6. Proveďte experimenty a dosažené výsledky zhodnoťte.

Literatura:

- ŠMEJDA, Martin. *Redundance dat na datových nosičích pomocí Reed-Solomonových kódů*. 2008.
- ŠICNER, Jiří. *Srovnání algoritmů dekódování Reed-Solomonova kódu*. 2011. PhD Thesis. Vysoké učení technické v Brně. Fakulta elektrotechniky a komunikačních technologií.
- KOLAČANSKÝ, Pavel. *Obnovení a rekonstrukce smazaných dat a dat z poškozených disků*. 2009.

Pro udělení zápočtu za první semestr je požadováno:

- Body 1 až 3.

Podrobné závazné pokyny pro vypracování bakalářské práce naleznete na adrese

<http://www.fit.vutbr.cz/info/szz/>

Technická zpráva bakalářské práce musí obsahovat formulaci cíle, charakteristiku současného stavu, teoretická a odborná východiska řešených problémů a specifikaci etap (20 až 30% celkového rozsahu technické zprávy).

Student odevzdá v jednom výtisku technickou zprávu a v elektronické podobě zdrojový text technické zprávy, úplnou programovou dokumentaci a zdrojové texty programů. Informace v elektronické podobě budou uloženy na standardním nepřepisovatelném paměťovém médiu (CD-R, DVD-R, apod.), které bude vloženo do písemné zprávy tak, aby nemohlo dojít k jeho ztrátě při běžné manipulaci.

Vedoucí: **Drahanský Martin, doc. Ing., Dipl.-Ing., Ph.D., UITS FIT VUT**

Datum zadání: 1. listopadu 2015

Datum odevzdání: 18. května 2016

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
Fakulta informačních technologií
Ústav inteligentních systémů
612 06 Brno, Božetěchova 2

doc. Dr. Ing. Petr Hanáček
vedoucí ústavu

Abstract

This bachelor thesis deals with the reconstruction of data from a damaged CD/DVD. The aim is to create an application which will be able to get a bitstream from an image (scanned fragments of a disc) acquired by an optical or electron microscope. This thesis also examines a data representation on the disc, process of preparation of specimens of CD and DVD and scanning of these specimens. The algorithm of the created software, its testing and results are described in this thesis as well.

Abstrakt

Tato práce se zabývá rekonstrukcí dat z poškozených CD/DVD. Cílem práce je napsat aplikaci, která bude schopna z obrazového vstupu (naskenované fragmenty disku) získaného optickým/elektronovým mikroskopem přečíst bitový stream. Práce se také věnuje způsobu uložení dat na disku, procesu přípravy jednotlivých vzorků CD a DVD a jejich skenování. Dále je v práci popsán algoritmus vytvořeného software, jeho testování a výsledky.

Keywords

CD, DVD, disc, disk, OpenCV, damage, reconstruction, data, bitstream, track, microscope

Klíčová slova

CD, DVD, disc, disk, OpenCV, poškození, rekonstrukce, data, bitstream, stopa, mikroskop

Reference

JAROLÍM, Jordán. *Reconstruction of Data from Damaged CD/DVD*. Brno, 2016. Bachelor's thesis. Brno University of Technology, Faculty of Information Technology. Supervisor Drahanský Martin.

Reconstruction of Data from Damaged CD/DVD

Declaration

Hereby I declare that this paper is my original authorial work which I have worked out by my own under the guidance of Mr. doc. Ing. Dipl.-Ing Martin Dražanský. All sources, references and literature used or excerpted during elaboration of this work are properly cited and listed in complete reference to the due source.

.....
Jordán Jarolím
May 12, 2016

Acknowledgements

The completion of this work could not have been possible without a participation and assistance of experts from FEI Czech Republic, s. r. o. and without material resources provided by this company. These contributions are sincerely appreciated and gratefully acknowledged.

© Jordán Jarolím, 2016.

This thesis was created as a school work at the Brno University of Technology, Faculty of Information Technology. The thesis is protected by copyright law and its use without author's explicit consent is illegal, except for cases defined by law.

Contents

1	Introduction	3
2	Mechanical construction of CDs/DVDs	4
2.1	CD	4
2.1.1	Basic parameters	4
2.1.2	Surface	5
2.1.3	Layers	5
2.1.4	Track	6
2.1.5	Pits	6
2.1.6	Laser beam	6
2.2	DVD	7
2.3	Used CD/DVD types	8
3	Writing and reading data on/from compact disks and SW data reconstruction	10
3.1	Data representation and addressing	10
3.1.1	Low-level data encoding	10
3.1.2	Addressing and track analysis	11
3.1.3	Types of data on track	12
3.2	Reading of data and error corrections	13
3.2.1	Introduction to error correction	13
3.2.2	Reading data and encoding of information	15
4	Scanning of surface with electron microscope	16
4.1	Electron microscope	16
4.2	Specimen preparation and scanning	18
4.2.1	General notes	18
4.2.2	Specimen preparation	18
4.2.3	Specimen scanning	19
4.3	Images post-processing	20
5	Software to get bitstream from images	22
5.1	Used technologies	22
5.2	Inputs and outputs	22
5.3	Used algorithm	23
5.4	Graphical mode	26
5.5	Testing	26

6 Conclusion	28
Bibliography	29
Appendices	31
List of Appendices	32
A Images from electron microscope	33
A.1 Data tracks vs heat gun	33
A.2 Data layer after cyklohexane ultrasound bath	34
A.3 Edge of fragment of CD	35
A.4 Scratches	36
A.4.1 Scratch 1	36
A.4.2 Scratch 2	37
A.5 Layers of CD	38
A.5.1 Tilted data layer + metal layer + protective layers	38
A.5.2 Layers of CD (substrate)	39
A.5.3 Layers of CD (coating + ink)	40
B Contents of CD	41

Chapter 1

Introduction

Storing of data has a long history in computer science as it was necessary to keep the source codes for the first mainframes. The very first technology used in computer science to store data were punched cards containing commands or data represented by presence of tiny holes in predefined positions. Later, magnetic tapes were used. Afterwards, first hard drives and floppy disks were invented, and finally CD and later DVD technology was introduced. Nowadays, many other technologies are used as USB drives, SSD, Blu-Ray, cloud services and the new ones are still coming.

All of the mentioned technologies, according to contemporary possibilities, tried to write and read data as fast as possible and tried to keep data readable as long as possible. An ability of data recovery from damages and distortions is very important as well. Correcting codes with data redundancy are used for the data recovery from CD/DVD disk. These codes work on a software level with a bitstream acquired from the standard CD/DVD drive using laser beam and analysing refracted light. Nevertheless, the alternative way of reading data from damaged disk is described in this thesis. An electron microscope was used to get images of the data layer of CD instead of analysing pits and lands with standard laser beam. It is necessary to combine these images together and post-process them while changing of hue/saturation and contrast levels. Afterwards, a specialised software described in this thesis can be used to get its bitstream by recognising pits and lands from fragments of data layer.

In Chapter 2, the technology, composition, and specifications of CD and DVD is described. Chapter 3 deals with the storage of the data and the data encoding and error corrections. Chapter 4 describes the technology of electron microscope, specimen preparation, specimen scanning and possible problems with scanning. Software to get the stream is described in Chapter 5. Experiments and results are shown in Chapter 5 as well.

Chapter 2

Mechanical construction of CDs/DVDs

One of the first Compact Discs (audio disk) was introduced to public in 1982 in Japan [14, 19]. The cooperation between Sony and Phillips brought the 12 cm disk which served to save and to reproduce a high quality audio signals. This became the base for the world CD standards. The IEC-908, usually called as Red book, from 1980 [20] is recognised as the world first standard describing the technology of the CD [15]. Whereas it was more than 30 years ago, a compact disk went through many improvements and evolved to many forms, for example well known DVDs and Blu-Ray disks. Although the physical construction of modern disks might be different from the construction of the first disks, they are based on the same principle.

2.1 CD

However, the evolution happened in small steps and it was a long way to today's Blu-Ray disks. The CD also has evolved in time. Nowadays, the most common and the most used types of CD are:

- CD-ROM - Read only disk
- CD-R - Recordable (Write once only)
- CD-RW - Rewritable (Write, erase, and re-write)

2.1.1 Basic parameters

Although some parameters of layers of the disk may differ, the composition is usually very similar. Cylinder disk with 120 mm diameter (can be less), 15 mm width of the centre hole and 1.2 mm thickness contains multiple layers. From the bottom of the disk (CD-ROM), it is a polycarbonate (plastic) substrate layer, a metal reflective layer and a protective lacquer. An optional protective label can be found on most disks as well [1]. Polycarbonate layer on the bottom of the disk mostly serves as protection of the data on the disk. Some big disruptions or dirt on the surface of this layer can cause misreading of data from the data layer. Data are written to the data layer as pits and lands, whereby the pits diffract the scanning light from the laser beam and the optical power, that is reflected back, is analysed and decoded [19].

2.1.2 Surface

The surface of each disk is divided into multiple parts. From the middle it is: centre hole, first transition area, clamping area, second transition area, information area, rim area. However, the most interesting part of the disk is the information area, in context of data reconstruction. This area lies between 44mm and 118mm and contains user data. The information area is further divided into the Lead-in area, User-data area and the Lead-out area. All the material specifications specify the materials only in this informational area. The rest of the disk can be made of any materials whereby total weight of the disk must be between 14g and 33g [1].

2.1.3 Layers

The thickness of layers of CD-ROM is shown on figure 2.1 [19]. From the bottom, it is the thickest disk substrate, usually a kind of polycarbonate with the refraction index of the used material 1.55. On this cover layer is a very thin metalized pit surface with thickness between 0.05 μm and 0.1 μm . Pits and lands are molded into the surface of the disk substrate and metalized pit surface serves to light reflection. Above, there is a protective cover lacquer coating with the thickness about 10 μm - 30 μm . And finally, the printing ink or protective label can be found on the top of the disk. Its thickness is about 5 μm . As we can see, CD-ROM does not have any separated data layer so this type of disk is not suitable for scanning on electron microscope (more in Chapter 4).

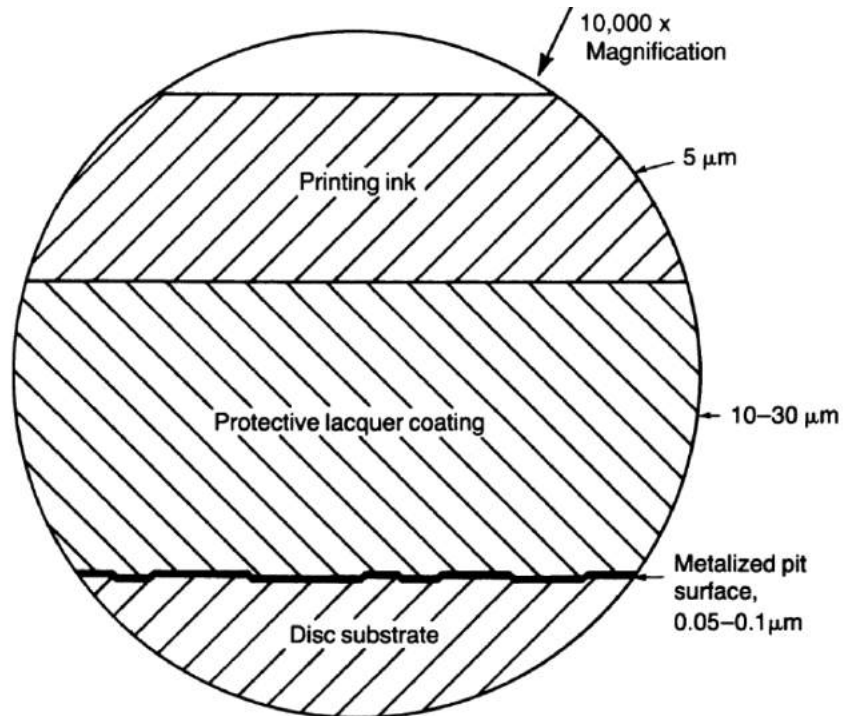


Figure 2.1: CD Layers[19]

CD-R contains one more layer between the polycarbonate substrate and the metal layer. This writable data layer is made by an organic dye which is photosensitive. Pits and lands are written to the dye by a laser beam which causes a chemical change on this layer. This

chemical change causes different diffractions of laser beam and can be recognised as pits and lands. Laser beam is also reflected by metal layer [4].

CD-RW has different type of data-recording layer. It also lies between the polycarbonate substrate and the metal reflective layer but it is phase-changing metal alloy film. A laser beam creates pits on this layer by heating the film above the melting temperature. The dielectric layers on both sides of the phase-changing film cause rapid cooling and this helps to created bits to remain in amorphous state. Reheating the phase-changing film to a specific temperature (below the melting temperature) reverts the film back to the crystalline state and erases previous written bits. The erasing and writing can be done together in a single pass when rewriting a disk [4]

2.1.4 Track

The most important part of the disk is its spiral track on the data layer. It contains pits and lands and goes from the centre of the disk outward when the disk rotates counter-clockwise as seen by laser beam. The pitch of the spiral track is $1.6 \mu\text{m} \pm 0.1 \mu\text{m}$. A pit track can contain about 3 billions of pits and the total length of unspiraled track is about 3,5 miles [19]. As CD must rotate with CLV (constant linear velocity) and track is spiral (each outer revolution contains more pits than inner revolution), the rotation speed of disk must vary according to the position of the pickup. The initial speed is 500 rpm, the rotational speed decreases as the pickup moves to the outward down to 200 rpm [19].

2.1.5 Pits

The information on the track is stored as pits and lands. Minimal length of pits is between $0.833 \mu\text{m}$ and $0.972 \mu\text{m}$, maximal length is between $3.05 \mu\text{m}$ and $3.56 \mu\text{m}$ and their width is approx $0.5 \mu\text{m}$ or $0.6 \mu\text{m}$. The length of the pits depends on recording velocity which can be between 1.20 m s^{-1} and 1.40 m s^{-1} . But this velocity must be constant with the maximum variation of 0.01 m/s . All pits are about $0.11 \mu\text{m}$ deep and are seen as bumps on the data layer surface. The size of pits enables relatively high information density, thus used laser must be able to recognise them. [19, 16].

2.1.6 Laser beam

The wavelength of used laser for reading a CD is 780 nm in the air (refractive index 1). However, the laser beam travels through the polycarbonate substrate on the disk with refractive index of 1.55. This causes change of its wavelength and its velocity and bends the laser beam. The wavelength in the polycarbonate substrate is about 500 nm . The velocity of a light decreases when passes from one environment to other with higher refractive index. The velocity of light slows from 3×10^5 kilometres/second to 1.9×10^5 kilometres/second. The deceleration causes bending of the beam and focusing occurs. These changes of the laser beam, thickness of the polycarbonate substrate and the numerical aperture (0.45) of the laser's lens make the laser beam focused approximately on $1 \mu\text{m}$ (sometimes said $1.4 \mu\text{m}$ [10]) on the data layer surface (as shown on figure 2.2 [19]). This is about two times bigger area than the width of the pit. The effects of scratches or disruptions are thus minimised and damages smaller then 0.5 mm do not cause any error in the readout [19].

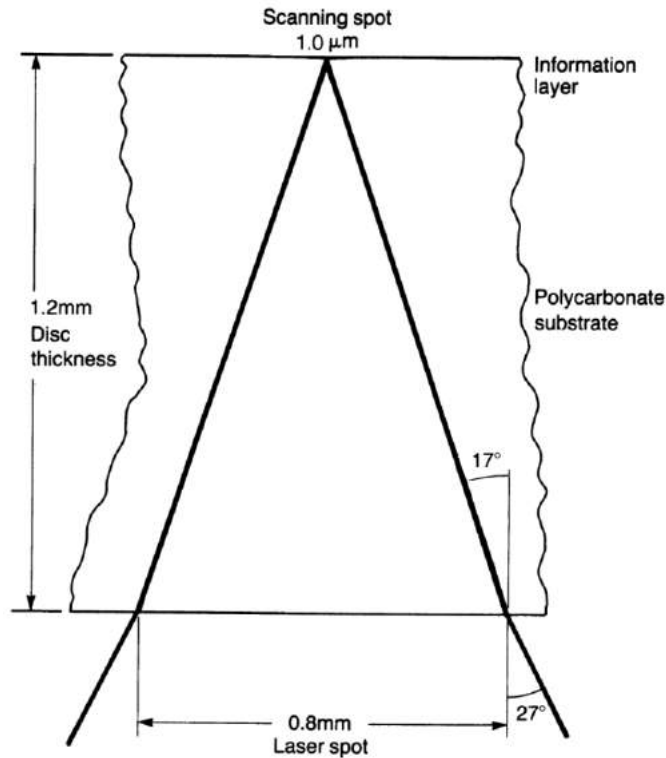


Figure 2.2: Laser beam[19]

2.2 DVD

Although the DVD (digital versatile disk) is based on a CD principle and both using pits and lands to store the data, they are both manufactured differently. We can imagine a DVD as two CDs glued together. DVD can be read/written on one or both sides. Another difference between CD and DVD is their track density, size of pits and lands, order of the layers and the count of recordable layers. DVD is manufactured in different versions nowadays [4, 10]:

- DVD-ROM: manufactured with one or two recorded layers on each side, read only
- DVD-video: manufactured with one or two recorded layers on each side, video
- DVD-audio: manufactured with one or two recorded layers on each side, multi-channel
- DVD-R: recordable DVDs with only one recording layer on each side, recordable
- DVD+-RW: recordable DVDs with only one recording layer on each side, random access
- DVD-RAM: recordable DVDs with only one recording layer on each side, rewritable

The disk diameter and disk height is the same as in case of CD, therefore it is 120 mm or 80 mm diameter, 15 mm centre hole, 1.2 mm thickness. However, the composition of layers is different. The DVD consists of two 0.6 mm thick substrates bonded together with the data layer(s) in between, see figure 2.3.

As we can see on figure 2.3 [10], some kinds of DVD use dual-layer technology. Layer 1 must be read through layer 0, thus layer 0 must be semi-transparent. Layer 0 reflects

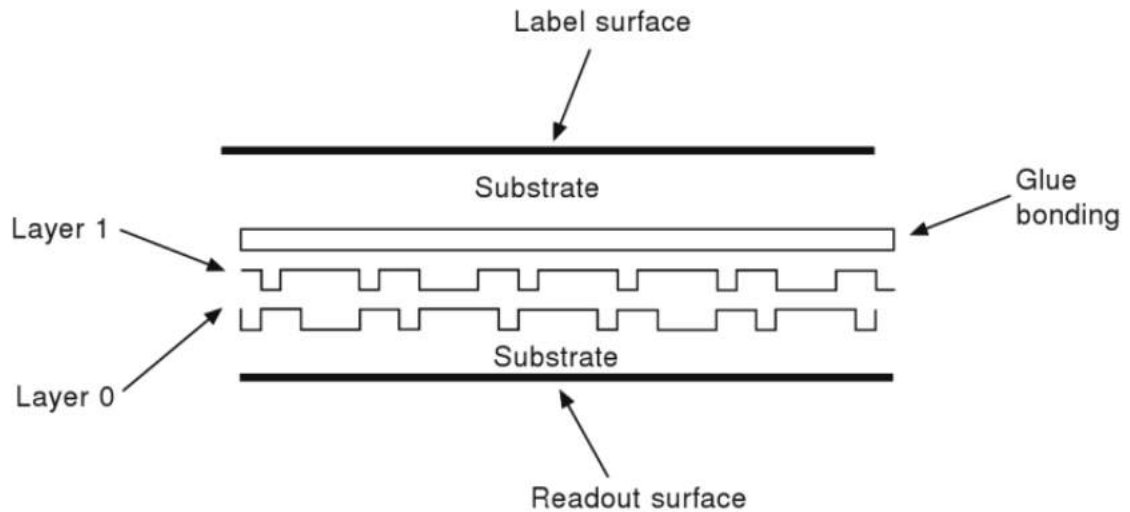


Figure 2.3: DVD Layers [10]

about 20% of the light, while the layer 1 reflects about 70%. Furthermore, data must be read from both sides for double-sided disks. Altogether, the disk can contain up to 4 data layers (2 layers from both sides).

The wavelength of a laser used for DVDs is 650 nm or 635 nm. Since the wavelength of the used laser is smaller than the wavelength of laser used for the CDs, the size of the beam spot, pits and lands can be smaller. The density of pits can be bigger. Thus the capacity of the disk is bigger. The beam spot is less than 1 μm . The track pitch is 740 nm. The minimal width of pits is 320 nm, their minimal length is 400 nm. Therefore, DVD can hold over 25 times more data than CD [10]. Comparison of CD and DVD can be seen on figure 2.4

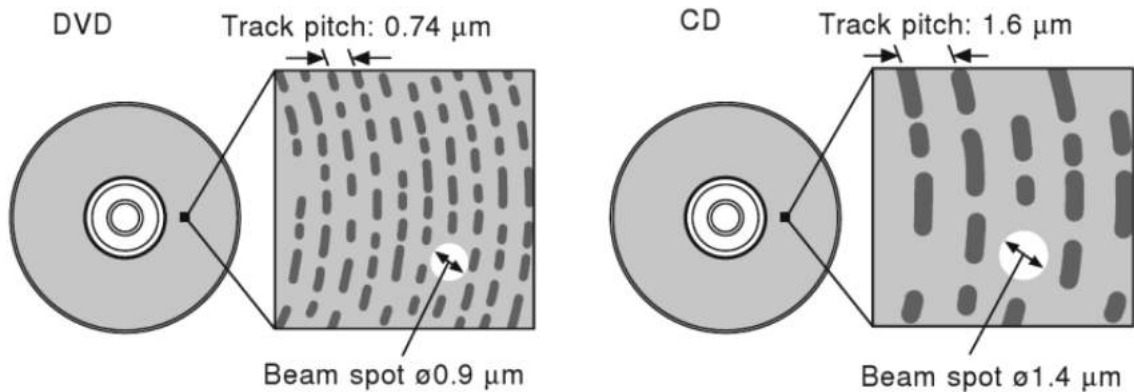


Figure 2.4: CD vs DVD [10]

2.3 Used CD/DVD types

Since the electron microscope was used for reading the data, it was necessary to separate the layers of the disk and read the data directly from the surface of the data layer with the electron beam (further explanation in Chapter 4). As one of the basic ideas of CD/DVD

is to preserve data for a long time undamaged, disks are not supposed to be split back into the layers. Unfortunately, it is not possible to separate the layers of every kind of disk.

It seems that the easier way is to scan the basic type of the disk - CD-ROM. However, as explained in Chapter 2.1, CD-ROM does not contain any separate data layer and pits and lands are molded to the surface of the substrate. Thus, the most suitable type of the disk is CD-R. This type of the disk is more suitable than CD-RW or any kind of DVD, because its construction is more simple and it has less layers. Therefore, it is easier to separate layers and the probability of preserving layers undamaged, while separating them, is higher.

During the preparation of CD/DVD samples for scanning was experimentally realised that the most suitable CDs for separating layers are TDK CDs. Other tested brands of CDs were Maxell, Verbatim, Philips and Sony but without satisfactory results. The preparation process and process of scanning are both described in Section 4.

Chapter 3

Writing and reading data on/from compact disks and SW data reconstruction

This Chapter deals with encoding and decoding information on the disks. It describes how the data are organised on the spiral track, how they are read and explains correcting mechanisms which are able to correct huge damages in the data stream.

3.1 Data representation and addressing

Whereas the used type of disk for scanning is CD-R, the following section is focused on storing data according to the standards for CD-R or older CD-ROM. Although the CD-R was introduced later, it is fully compatible with CD-ROM. Cited standards are ECMA-130 [1], ISO/IEC 10149:1995 [11] and ECMA-394 [2].

3.1.1 Low-level data encoding

As mentioned above, the data are physically stored on CD as pits and lands on spiral track. The information on the spiral track is represented by 8-bit bytes and it is grouped to sectors, whereas each sector contains 2352 bytes and it is the smallest addressable area of a Digital Data Track that can be accessed independently. The number of the sectors on CD can vary depending on the amount of a data stored. Sector can have various layouts depending on the setting of the Sector Mode byte and each sector is divided to multiple parts and each part stores multiple types of information. However, the first 12 bytes are always the sync bytes. The rest of the sector contains 2340 bytes and is scrambled and mapped into frames afterwards. These frames are called F_1 -Frames and each F_1 -Frame contains 24 8-bit bytes. These F_1 -Frames are encoded by Cross Interleaved Reed-Solomon encoder into 32 8-bit bytes since 8 8-bit bytes carrying parity information are added. These 32-bit frames are called F_2 -Frames. After addition of a Control byte, we will get F_3 -Frames containing 33 bytes. This Control byte is generated as one from 98 possible combinations and is mainly used for addressing purposes. The 98 control bytes are added to 98 consecutive F_2 -Frames and form groups of F_3 -Frames called Sections. There is no relation between these Sections and Sectors (the smallest addressable area) defined and a mapping of the Sections to the Sectors depends on a concrete implementation. However,

each frame in section contains 192 data bits ($24 * 8$), altogether the size of section is 2352 bytes ($192 \text{ databits} * 98 \text{ frames}$) - the same size as the size of the sector.

The F_3 -Frames are further encoded in order to be recorded on CD. All of the 8-bit bytes of 33 bytes in F_3 -Frame are encoded by 8-to-14 modulation (EFM) and each of the 33 bytes now consists of 14 bits. These 14 bits are called Channel bits and can contain at least two and at most ten ZEROs between two ONES. Furthermore, F_3 -Frame is represented by a Channel Frame which contains Sync Header (24 Channel bits), Merging bits (3 Channel bits) and 33 14-Channel bit bytes each followed by 3 Merging Channel bits. Altogether, the Channel Frame contains 588 Channel bits decoded from original 192 bits (24 8-bit bytes). The Channel bits from the Channel frame are physically recorded to the track as pits and lands while a ONE Channel bit is represented as change from pit to land or vice versa and ZERO Channel bit is represented as no change on the track. This is called non-return-to-zero, inverted (NRZI). Illustrative flow of encoding is shown on figure 3.1.

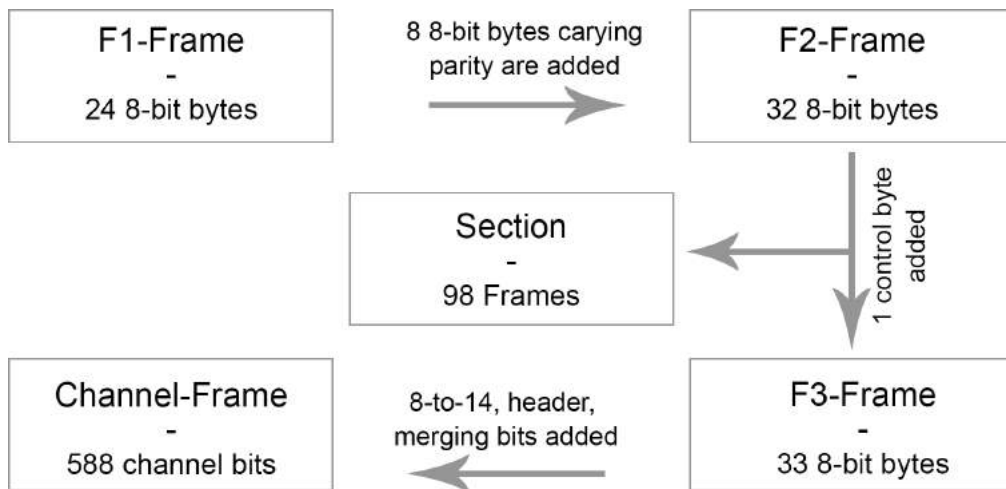


Figure 3.1: Data organisation

3.1.2 Addressing and track analysis

Using the presented scheme, it is possible to address individual sectors. Although a sector is defined as smallest addressable area, they are mapped to sections. The address of the section is defined as the time from the start of the user data area to that section. This time is stored in control bytes of each section and it is called an absolute time. Since the rate of Channel bits is defined as 4 321 800 Channel bits per second and each section consists of 98 frames and each frame consists of 588 Channel bits, the total amount section which can be read is 75 sections per second. Thus, sections are addressed with $1/75$ s accuracy.

Furthermore, the Channel bit is also called the smallest time unit used on a CD [13]. Nevertheless, the data are stored as pits and lands with specific sizes, while different sizes mean different information. It follows therefore that the time of Channel bit must be possible to convert to the size units following the sizes of pits and lands. The size of the Channel bit is called the Channel bit length and can be computed, using the writing/reading velocity, as follows:

$$\frac{1.20 \text{ m s}^{-1}}{4\,321\,800 \text{ channelbit s}^{-1}} = 277.662 \text{ nm} \quad (3.1)$$

It is 277.662 nm for the writing/reading velocity of 1.20 m s^{-1} and 323.939 nm for the velocity of 1.40 m s^{-1} . If we combine the information about the channel bit length with maximum length of ZERO string in channel bit word (2-10) we will get possible lengths of pits as shown on figure 3.2 for the velocity of 1.20 m s^{-1} [13].

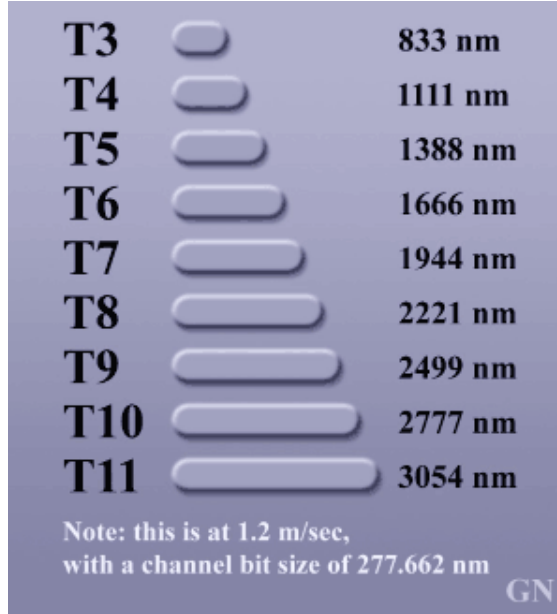


Figure 3.2: Pit sizes [13]

The pit T3 on the figure 3.2 [13] represents the sequence of *1001* and T11 represents *10000000001*. As mentioned above, the minimal and maximal sizes of pits and their density can vary according to the velocity. It means that the length of the track containing the 4 321 800 channelbit can be between 1.20 m and 1.40 m. These channel bits can contain 1 s of audio for example. The relation between pits and channel bits is shown on figure 3.3 [13].

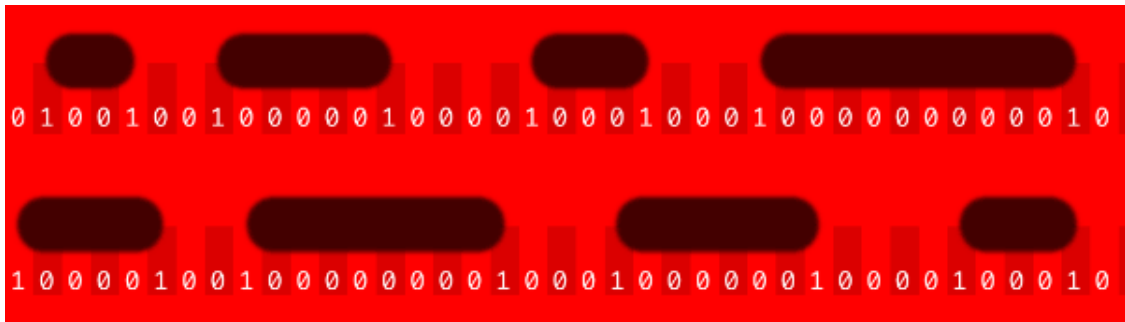


Figure 3.3: Channel bits [13]

3.1.3 Types of data on track

Moreover, 2 types of information that can be encoded by explained mechanisms and stored on the disk. These are *digital audio* and *data* [18]. This division has historical reasons as the first standard [20] defined CD as an audio medium. Furthermore, the spiral track

which contain pits and lands can be divided into entities called tracks. The track entity is more above than low-level encoding or spiral data track in the hierarchy of CD and it is the biggest entity on a CD. Thus, 2 types of tracks (entities) exist - the CD-DA track for digital audio and the data track. The CD can contain both of these types combined, but data track must come first [18].

Taking the low-level data encoding into account, we can calculate the *data* bit rate for the data track, where 192 is number of data information bits in frame, 98 is number of frames in section and 75 section per second can be read:

$$192 * 98 * 75 = 1\,411\,200 \text{ bit/s} \quad (3.2)$$

Compared with required CD digital audio data rate with 44.1 kHz sampling frequency, 16 bits per sample on 2 channels:

$$44100 * 16 * 2 = 1\,411\,200 \text{ bit/s} \quad (3.3)$$

One can see that the track type does not affect the low-level encoding nor the output data bitrate and the representation on the low level uses the same mechanisms for data or digital audio.

3.2 Reading of data and error corrections

It was explained how the data are organised on a CD. The encoding and reading of data will be further explained in this section with emphasis on the error correction.

3.2.1 Introduction to error correction

Although the CD is very resistant to any damages and it has very long lifetime, some errors can arise from production defects or from the damages on the surface of cover layer. One of the characteristics of the damages is that they usually appear in a long burst. So that, the data reconstruction is secured by interleaving mechanism which break up these long burst to the short errors and by system which is able correct these errors. This mechanism is called CIRC - Cross Interleave Reed-Solomon Coding [18].

The 2-step Reed-Solomon code is used in CIRC for the error correction while reading data from a CD. These two levels are specified as:

- $C1 : m = 8, n = 32, k = 28$
- $C2 : m = 8, n = 28, k = 24$

where m is one unit of information, n is number of units after the redundant data (units) are added and k is number of data carrying units. This concrete configuration represents that each unit consists of 8 bits and 4 units which can be used to for error data correction exist [12].

Very simple illustrative example of the configuration can be $m = 1, n = 7, k = 4$. This means that we need to encode 16 possible combinations of 4 symbols (2^4), but we have 128 (2^7) combinations available. Thus, only 16 of the most suitable combinations from 128 combinations is picked up. These combinations are picked up in order to be different from each other by concrete number of symbols. This difference is called the distance d . In this case, the distance between two illustrative words $X = 1001101$ and $Y = 1010001$ is $d = 3$

because these 2 words differ in 3 symbols. The distance $d = 3$ is the same for all 16 picked combinations from 128 possible combinations. If we change one symbol in X , we can get for example $X' = 1001001$. Now, the distance between X' and X is 1, the distance between X' and Y is 2. As the original distance between X and Y was 3, we can assume that now the distance between X' and any other picked symbol is at least 2. Therefore, X' is an erroneous form of X and X' can be reconstructed back to its original form. However, if we change 2 symbols to $X'' = 1011001$, the distance between X'' and X is equal to 2 but the distance between X'' and Y is only 1. Therefore, it is no longer possible to see X'' as an erroneous form of X . For t_1 errors, we can deduce the following[12]:

$$d \geq 2t_1 + 1 \tag{3.4}$$

This means that for distance $d = 3$ only one error symbol can be corrected. However, up to $t_2 = 2$ errors can be detected according to the following:

$$t_2 = d - (2t_1 + 1) \tag{3.5}$$

where $t_2 \geq 0$. This implicates that if no error is corrected, up to 2 errors can be detected, if one error is corrected, 0 other errors can be detected. Lets apply this to the distance of $d = 5$, as used on CD. We will get[12]:

$$0 \text{ symbol correction} : t_1 = 0 \rightarrow t_2 = 4 \tag{3.6}$$

$$1 \text{ symbol correction} : t_1 = 1 \rightarrow t_2 = 2 \tag{3.7}$$

$$2 \text{ symbol correction} : t_1 = 2 \rightarrow t_2 = 0 \tag{3.8}$$

Then, the maximum number of detected errors is 4 with correction of 0 symbols. It is 2 with correction of 1 symbol and 0 with correction of 2 symbols. If 3 or more errors are occurred, it is not possible to reconstruct the data reliably because the word can be considered as some other data word. However, there is one more concept used called erasure. Even if it was defined that for $d = 5$ the maximum number of corrected symbols is 2, up to 4 symbols can be corrected using this concept. Lets have a data word X of length 16 bits. Lets suppose that 4 of its symbols were replaced by a different symbol and are erroneous. If we know, by some other way, the position of erroneous symbols and compare this new data word X' with the original X , the distance is $d = 4$ but the distance to any other symbol is still $d = 5$. Therefore, we can identify X' as an erroneous form of X [12].

As mentioned above, these correcting mechanisms are a base for Reed-Solomon correcting codes that are used for error correction on data stream acquired from CD. The RS codes belong to the category of Bose-Ray-Chaudhuri codes so they are correction, cyclic and linear codes and are one of the most widely used correcting mechanisms [21]. Cyclic codes have the attribute that a cyclic shift of a valid codeword produces another valid codeword [17]. Linear code expresses a code with an alphabet that is finite field F and a set of codewords is a subspace of this alphabet [5]. The RS codes are usually defined by (n, k) [21] as explained above.

However, the RS codes are not the only one mechanism used to prevent the data loss. The others are interleaving, merging bits and EFM, as mentioned in Section 3.2. These all must be used and decoded while reading.

3.2.2 Reading data and encoding of information

The reading of data is the opposite process to encoding them. One can see, that the actions to encode data must be taken in a reverse order. First of all, the data from cross interleave component are decoded back from 14-bit units to 8-bit bytes. This is opposite process to EFM. The 3 merging bits are ignored as they do not carry any information and the control byte is processed. Secondly, these 32 bytes are sent to the first Reed-Solomon decoder (C1). This decoder is able to correct most of the random errors using 4 parity bytes from 32. Then, remaining 28 bytes are passed through unequal delays to interleave and spread long burst errors into smaller one before they are passed into C2 decoder which corrects the rest of the burst errors. The de-interleaving process takes place to restore the correct byte order. These correction mechanisms are extremely powerful - up to 200 errors per second are allowed [18].

The technique of storing and encoding of data was explained in this section as well as the addressing on the disks, error correction mechanisms and process of reading.

Chapter 4

Scanning of surface with electron microscope

One of the most important parts of this bachelor thesis is scanning of the CD layers with an electron microscope. Acquired images from the electron microscope are used as source for recognising software. Therefore, it is necessary to understand the basics of electron microscopy, to understand the specimen preparation and to describe possible problems with scanning of a specimen.

4.1 Electron microscope

Electron microscopy has a long history and it has developed to many forms. The most known types of electron microscopes are SEM, TEM and AEM [7]. All of these basic types exist in more versions, as for example ESEM. This type is explained in Section 4.2.2. In this thesis, the SEM or ESEM were used, especially FEI Helios NanoLab G3 CX¹.

SEM is Scanning Electron Microscope, based on scattered electrons and provides information about the surface of the specimen. As we can see on figure 4.1, SEM consists of 4 main parts. The tube with the source of electron beam and electron beam deflected by deflection coils, the vacuum chamber with a movable stage with a specimen, the detector and the software with display device. A thermal emission on wolfram thread is usually used as the source of the electron beam [9]. Afterwards, the electron beam is deflected by deflection coils / condenser lenses to scan whole field of view. Movable stage gives a possibility to scan larger areas of the specimen surface. The mainly used types of detectors are [6]:

- Secondary Electron (SED)
- Backscattered (BSD)
- Large Field (LFD)
- Gaseous Secondary Electron Detector (GSED)
- Charge-coupled Device (CCD)

The SED in backscattered mode was used in this thesis for scanning final images of the data layer. The difference between secondary and backscattered electrons is in their

¹<http://www.fei.com/products/dualbeam/helios-nanolab/>

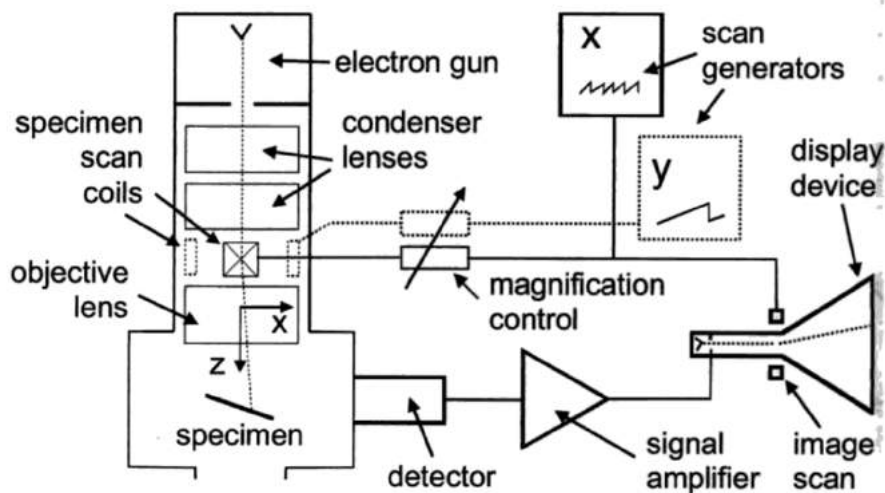


Figure 4.1: SEM principle [7]

energy and the depth from which they are emitted. The energy of the secondary electrons is, by convention, less than 50 eV. The energy of backscattered electrons is more than 50 eV [6]. When the electron beam is shot out to the specimen, other electrons are emitted from the specimen surface and captured by the detector. If the detector is in secondary mode, the voltage on detector is positive and catches almost all emitted electrons from the surface of the specimen. In this mode, the surface of the specimen is usually scanned. If the detector is in backscatter mode, the voltage on the detector is negative and detector repels electrons except those that have been emitted directly to the detector and their energy is high enough. This mode is used to scan the structure of the specimen (deeper structure of the surface, not the structure of whole specimen as in case of TEM). See the figure 4.2 [3].

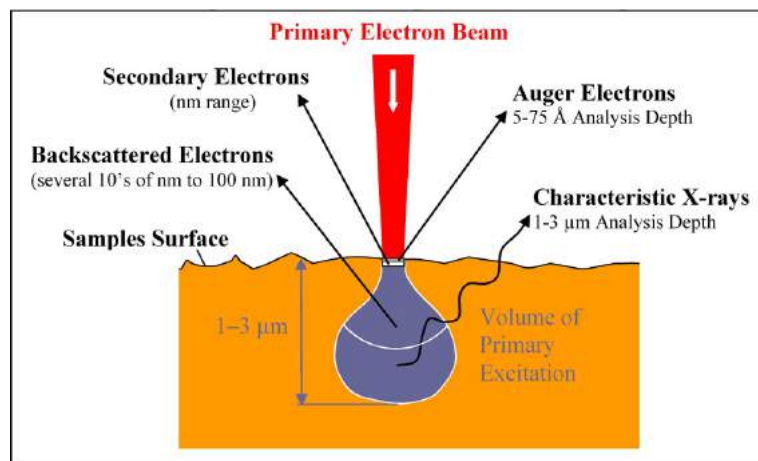


Figure 4.2: Secondary vs backscattered electrons [3]

The depth from which the electrons are emitted depends on the material of the specimen. The electrons captured by the detector are accelerated and transformed by scintillator to the light and then processed to the final image.

4.2 Specimen preparation and scanning

4.2.1 General notes

To scan a surface of data layer of CD and see the pits and lands is necessary to solve a few problems. First of all, the thickness of cover lacquer and metalized layer is more than $10.1\ \mu\text{m}$. The thickness of polycarbonate cover layer is $1.2\ \text{mm}$. However, the limit depth of electron emission is several *nanometres* for secondary electrons and only 10's or 100's of *nanometres* for backscattered electrons. One can see that scanning of the pits and lands from the depth of $10\ \mu\text{m}$ (as seen from the top of the CD without label ink) is on (over) the limit of the electron microscope without a preparation of the specimen.

Another complication is a charging of the specimen. Polycarbonate and lacquer/ink are both nonconductive - a waste charge cannot be grounded to the side of the vacuum chamber. So that, the waste electrons remain on the surface of the specimen and repeal the new electrons shot out from the electron beam. Then the emission of the electrons from the surface of specimen is very small and the observation of the specimen is distorted and becoming worse and worse in time.

Finally, the size of the specimen brings another complication - the vacuum chamber must be big enough and the range of motion of movable stage has to meet the size of the specimen. Lets say that the specimen is one quarter of the CD. In this case, the range of the movable stage must be at least $60\ \text{mm}$ to each direction. Furthermore, the depth of field while scanning is about $1\ \mu\text{m}$ with $10000x$ magnification [8]. This means tolerance only of 0.00191° for the specimen with the length of its side $60\ \text{mm}$. So that, even very tiny deviation from an absolute horizontal position of the specimen causes a blurring of the image. Last but not least, scanning on electron microscope takes a very long time.

4.2.2 Specimen preparation

The natural way of scanning pieces of CD is to scan whole specimens through all layers. Although, the depth of the data layer in the structure of the CD is on (over) the limit of electron microscope, the depth of emission depends on the material. This would be the easiest way how to get images of pits and lands. So that, the first problem to solve was a charging of the specimen while scanning through cover layers.

One of the possible methods how to divert the charge from the specimen is to metallise the surface of the specimen with thin layer of metal. Commonly used metals are gold or platinum. Two layers of gold were tried, one $5\ \mu\text{m}$ thin and second $10\ \mu\text{m}$ thin. Even if these layers help to ground the negative charge on the surface of the specimen, the data layer is under $15\ \mu\text{m}$ of material (gold + cover lacquer layer) so it is impossible to recognise any pits or lands on the surface of data layer. The data layer is just too deep. Another possible option is to use different type of electron microscope, called ESEM. ESEM means Environmental Scanning Electron Microscope. In this bachelor thesis, the used one is FEI Quanta FEG. It differs from the SEM by the pressure in the chamber and generally said, wet and uncoated specimen can be observed in the chamber. ESEM is able to raise the relative humidity in the chamber up to 100%. Demineralised hot water is injected into the chamber and thanks to the low pressure in the chamber, the water changes to the steam which helps to divert the charge from the specimen. Unfortunately, specimen stayed still charged and the quality of the images was not as good as needed. Furthermore, ESEM is primary intended as microscope for biological specimens [6].

Considering unsuccessful attempts with grounding the charge and scanning through

cover layers, the only possible option seems to be a separation of data layer from the others. Due to the high requirements of quality, the CD was made from high quality materials and is very resistant to any kind of damage. It is not so difficult to broke or damage the CD, but it is necessary to keep the data layer undamaged or in a readable state at least. Finally, CD was not made to be split back into a layers.

As defined in the international standard [1], operational temperature of a CD is between -25°C to 70°C and storage conditions says -20°C to 70°C . Trying to separate layers, the CD was exposed to temperature below and above the mentioned limits. The CD was sunk into liquid nitrogen having a temperature below -196°C but without any signs of damage or separation of layers. Sinking CD into boiling water for more than one hour also did not bring needed results. Another attempt was made with a heat gun, temperature was set between 100°C and 250°C . However, the heat gun damaged the CD more than separate the layers (shown on figure A.1 in the appendix of this document). Some attempts with the electric iron were also unsuccessful, with no effects on CD. Another possible way how to separate the layers is a chemical reaction of polycarbonate with a kind of dissolvent, for example isopropyl, acetone or cyclohexane. Even though these dissolvents caused changes on the surface of the polycarbonate cover layer, any of them was not able to melt the polycarbonate layer completely. However, the combination of cyclohexane and ultrasound bath was successful. After the ultrasound bath, the disk substrate was melted completely as well as cover lacquer. The remaining layer is metal reflective layer accompanied by a data layer made from organic dye. As shown on figure A.2 placed in the appendix of this document, this data layer was hardly damaged. Therefore, it is not possible to dissolve the disk substrate and keep the data layer readable with tested dissolvents. After these unsuccessful attempts with separation of layers, one method brought satisfactory results. It was found out that the best way to separate the layers of the CD is to take a broken CD and put it into hot water steam for a few minutes. The broken edge of the CD (or its part) shall face directly into the steam flow. Then, the molecules of the water steam separate layers step by step and it is possible to peel off the cover lacquer, metal reflective layer and the data layer from the cover polycarbonate layer. However, it is very hard to peel off a large parts and the layer is usually tore into smaller pieces. This negatively affect the amount of reconstructed data. So it is necessary to do it very carefully and it is highly recommended to do this part of preparation in a clean laboratory, because any dust particle on the data layer (bottom of the metal reflective layer) will devaluate the image acquired form the electron microscope later.

4.2.3 Specimen scanning

Lets acquire the right specimen by following the described process. Now, it is possible to start with the scanning of the specimen. As mentioned in Chapter 4, the microscope used for scanning is FEI Helios NanoLab G3 CX in backscattered mode. As we have broken cd into small pieces and the data layer as well, it is necessary to scan the surface of the layer piece by piece. However, breaks causes large damages on data layer. This is shown on figure A.3 placed in the appendix. The scanning is also very time consuming. The area of size 1 mm^2 takes about 240 minutes to scan. Therefore, about 1500 days would be needed theoretically to scan the surface of the CD containing the data track (between 24.8 mm and 59 mm) + time needed for preparation of specimen + time to change specimens in microscope chamber + post-processing of acquired images. Another difficulty is that only very small damage on the surface of data layer destroys a huge amount of pits and lands -

huge amount of data. As can be seen on figure A.4 and A.5.

Despite all the described problems, it has been achieved to satisfactory results as small fragments of CD were successfully scanned. As shown on figure 4.3 pits and lands can be both successfully recognised by a human eye and the quality of the picture is good enough to be post-processed.

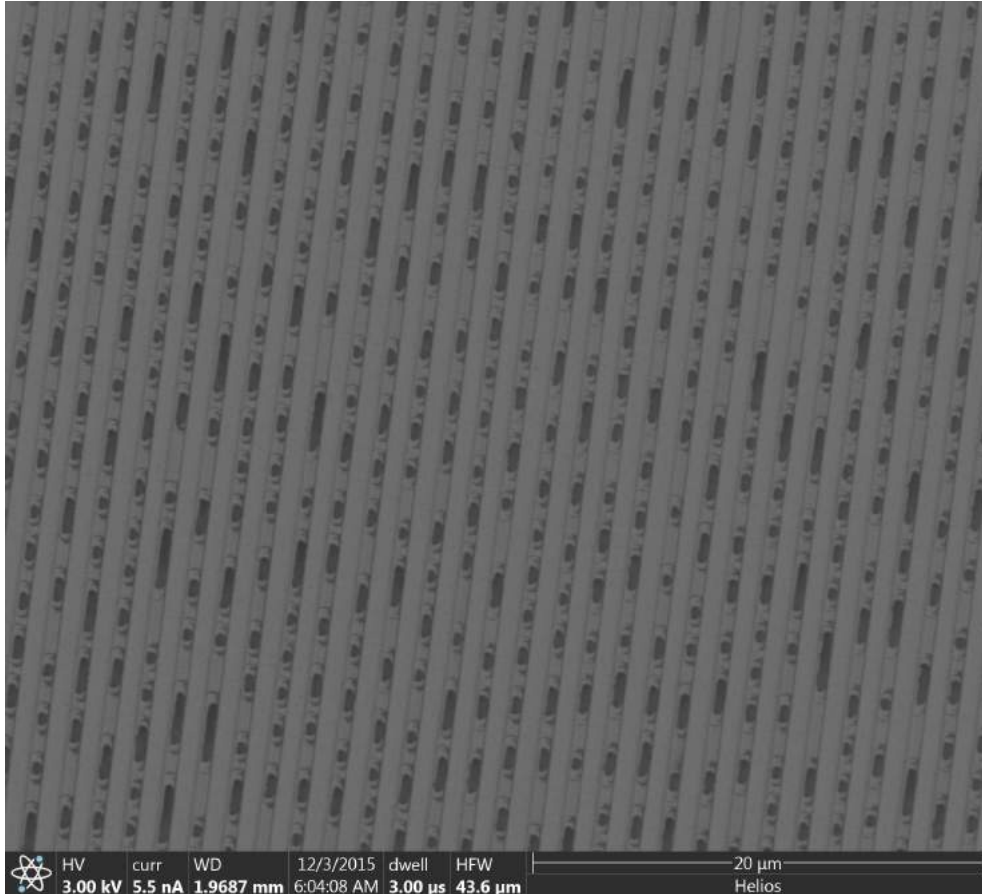


Figure 4.3: Pits and lands as seen on electron microscope

4.3 Images post-processing

The image post-processing can be divided into 3 main sections. Firstly, it is necessary to compose acquired images to one big image containing large area of the data layer surface. Secondly, the grayscale images from the electron microscope must be transformed to B/W images in order to process by specialised SW later. While transforming images from grayscale to B/W, hue/saturation, contrast and brightness levels can be changed in order to get the most suitable image for pit recognition. Thirdly, the specialised software is used to process the image and get bit stream according to the pit and lands depicted in the B/W image.

Two of the three steps can be done in software called Maps (FEI). First of all, very large amount of the images must be composed into one ensemble. However, individual images may not fit to each other. Thus, it is necessary to stitch images to prevent a discontinuity of the image as shown on figure 4.4.

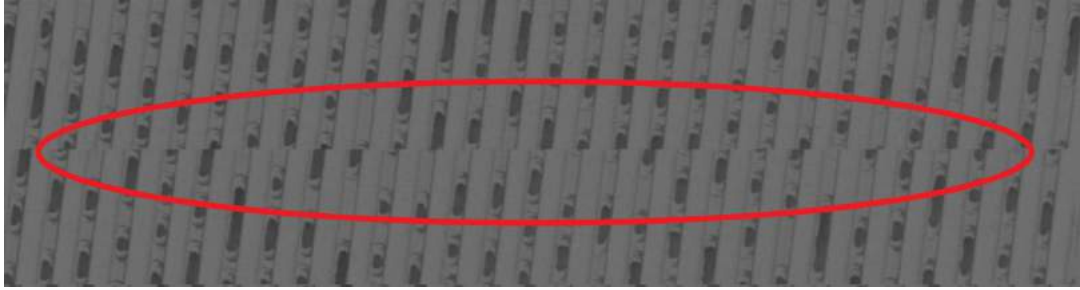


Figure 4.4: Unstitched images

After stitching the images, the image can be converted to B/W. Contrast, brightness, hue and saturation levels can be changes in order to get the highest possible contrast between lands and pits but without the data loss at the same time. Different ways of post-processing of image are shown on figure 4.5 where the last image represents the best result. Since every image from the electron microscope may differ, according to actual settings, quality of specimen, charge of specimen, etc, it is necessary to set the level individually for every single image and it is not possible to define universal values in order to get the best results.

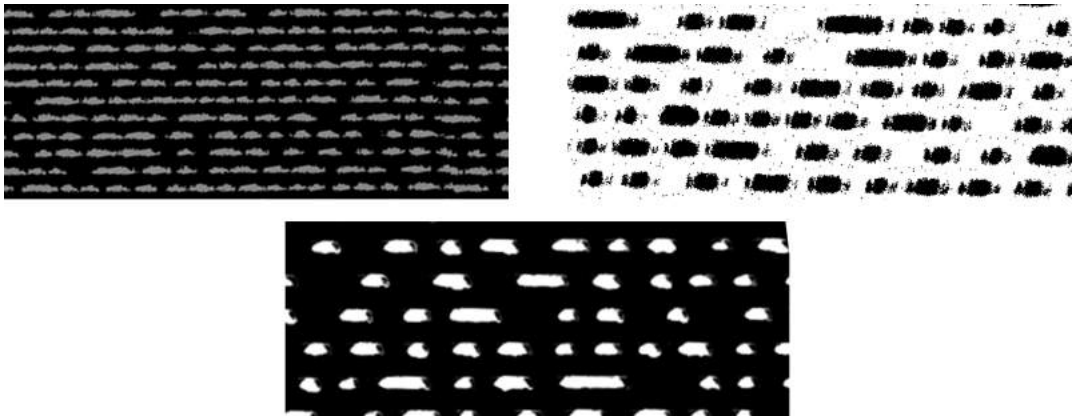


Figure 4.5: Post-processing of images

After the preparation of the specimen, scanning with electron microscope and post-processing image, it is possible to use the software in order to get a bitstream. Other examples of images of CD acquired from the electron microscope are shown in the appendix of this document.

Chapter 5

Software to get bitstream from images

The main purpose of this project is to write a software which is able to recognise pits and lands from the images and to get a bitstream. Created software can deal with *basic* damages of the data layer and image distortions and enables to transform images in its graphical mode. This Chapter describes used technologies, algorithm, inputs, outputs and testing.

5.1 Used technologies

The software to get a bitstream from images is primarily written in C++ and uses OpenCV library. It also supports graphical user interface which is written in QT. CMake is used to create a makefile and build the runnable application. Thus, minimal requirements are:

- CMake 3.4.2 or higher
- OpenCV 3.1.0 or higher
- QT 4.8.7 or higher
- GCC 5.3.0 or higher

OpenCV is a library for image and video processing, recognition, stitching etc. . . In this project, OpenCV is used for recognition of data tracks, pits and lands and for filtering damages and distortions of the images. QT is library which serves mainly to creation of user graphic interfaces. Therefore, it is used to create graphical interface providing basic image transformations and read/save operations.

5.2 Inputs and outputs

Whereas the software can be run in two modes, the requirements may differ. In text mode, the requirements for the input image are:

- Image shall be in grayscale or B/W
- Pits and lands shall be distinguishable by human eye

- Image shall contain typical surface of CD/DVD (pits and lands are organised into tracks and covers most of the image)
- Damages cover only a small part of the image
- Image is rotated as the tracks are in horizontal position with minimal deflections
- White pits on black background

The image opened in graphical mode can be coloured and rotated arbitrary because these parameters can be changed in graphical mode. However, the rest of requirements is the same as in text mode. Furthermore, all of the requirements defined for the text mode must be met before the start of image analysis and data recognition.

Output is saved as .txt file in the build folder of the project and contains bit values expressed as chars 0 and 1 without any separators. Directory can be change in graphical mode and output is not saved automatically in this mode.

5.3 Used algorithm

Lets have an input image, which is preprocessed, its colour levels were changed according to figure 4.5 and its tracks are in perfect horizontal position. Exact input requirements are defined in Section 5.2. The algorithm of getting data stream from images can be described as algorithm 1 (opening operations and graphic mode with image transformations are omitted):

```

Data: Source image
Result: Bit stream
if open image then
    find and save contours;
    determine limit values for damages;
    handle damages;
    find and save central point of each contour;
    sort contours by y-axis;
    split contours into tracks;
    foreach track do
        foreach image point in track do
            if image point lies in contour from the track then
                | save 1;
            else
                | save 0;
            end
        end
    end
    save output;
end
else
    show error;
end
end

```

Algorithm 1: Processing of image

Even though that the algorithm is quite straightforward, let's explain some of its parts with exact values. All of the values are defined by CD/DVD standards or they are calculated from the size of contours or from the distances of tracks.

First of all, it is necessary to analyse the distortions and damages on the image and handle them. The average height and the average area of all contours are computed. Then, the size and height of every contour (pit) is compared with limit values for size and height. If the contour is bigger it is deleted from the list of contours and the averages are actualised. Thus, the erroneous area will not affect further computations. This is repeated until all of the over-limited contours are deleted. The limit values are set as:

$$size_{max} = size_{avg} * \frac{3560}{833} \quad (5.1)$$

$$height_{max} = height_{avg} * \frac{1.6}{0.5} \quad (5.2)$$

where $3560 \mu\text{m}$ is the maximum length, $833 \mu\text{m}$ is the minimum length of the pit, $1.6 \mu\text{m}$ is maximal theoretical possible height of pit and $0.5 \mu\text{m}$ is defined size of pit. Whereas the faulty contour was deleted from the list, it will seem as land in later processing. Thus it will produce $0's$.

Afterwards, contours can be split into the tracks. Whereas the scanned areas acquired from electron microscope are very small (up to 0.5 mm), the curvature of the track can be ignored and all operations are tailored to straight lines. As defined in Chapter 2, the height of pits is up to $h = 0.6 \mu\text{m}$ and the pitch of the track is $d = 1.6 \mu\text{m}$. The average height h_1 of contours (pits) in the image can be determined in pixels without any problems. So that, we can determine average height of track d_1 in pixels as:

$$\frac{1.6 \mu\text{m}}{0.6 \mu\text{m}} = \frac{d_1}{h_1} \quad (5.3)$$

Now, the tolerance for the maximum vertical distance of two central points (point in the centre of a contour) on one track can be determined. In the beginning, it is y coordinate of the first contour $+1/2 * d_1$. Later, it is the average of y coordinates from one track $+1/2 * d_1$ where this average is actualised with every controlled contour. It is $1/2 * d_1$ because the contours are sorted by coordinates of $y - axis$ and software shall check contours below the actual track level only. Now, it is possible to go through contours sorted by y coordinate and check if its central point is above the tolerance level. If not, this contour is recognise as the beginning of the new track. This is shown on figure 5.1. Tracks with defined sizes are shown in the left part of the image. Splitting into the tracks is demonstrated as different tracks has different colours of pits in the right part of the image.

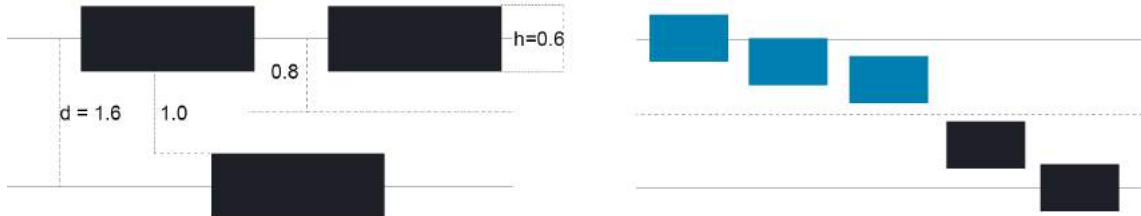


Figure 5.1: Tracks and pits (heights)

Now, the tracks can be analysed and the presence of pits can be tested. The average position p of all central points on track is computed for each track. This serves as line-marker of horizontal position (y -axis) for further analysis. The searching algorithm moves horizontally from left to right on this line-marker. For every testing horizontal position, all of the contours are tested and if any of them contains any point from the range of track, the 1 is saved, otherwise the 0 is saved. This is repeated for all tracks. The range of track is determined as every value between $p + 1/2 * d_1 * 0.85$ and $p - 1/2 * d_1 * 0.85$. The different stages of tracks analysis are shown on figure 5.2. The line-marker is visible for each track and the white areas represents tested pixels. All these pixels have been tested as they are/are not present in some contour from adjacent track.

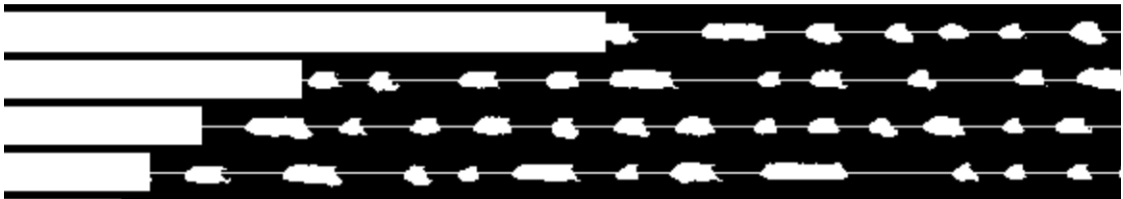


Figure 5.2: Analysis of tracks

Next logical step would be to analyse the output string of 0's and 1's and use NRZI. However, the channel bit length must have been known in order to do this. As explained above, the length/the number of channel bits determines the number of 0's and 1's in final stream which has to be further encoded. Nevertheless, the length of the channel bit is determined by velocity and the velocity cannot be determined from the source images which contains only small fragments of data layer. Thus, the raw bit stream cannot be further analysed.

One of the requirements for the input image is that its tracks must be in perfect horizontal position. During the implementation, there were attempts to do this automatically. The algorithm was based on determination of original deflection of tracks and rotation to horizontal position. However, the determination of the original angle, based on rotation of contours, was problematic. The angle of contours (pits) is not recognised with a high accuracy and this causes large deflections as seen on figure 5.3. The line in the middle of the image follows the average angle of contours. The tracks on the picture are obviously rotated but the pits are recognised as rectangles with almost 0° rotation. So that, the exact angle of track could not be recognised.

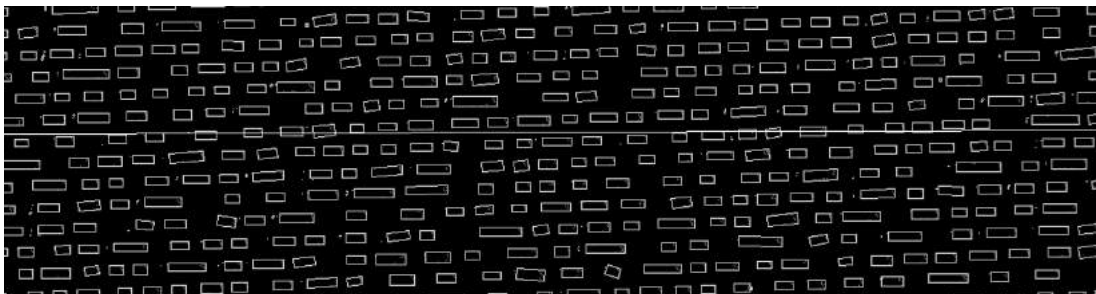


Figure 5.3: Rotation

5.4 Graphical mode

This application also support a graphical mode which enables user to do basic image transformations. Supported operations are: opening of image, saving of image, inverting of colours, zoom, changing of threshold levels, brightness, contrast, rotation, turning on/off all filters, getting of bitstream and saving of bitstream.

5.5 Testing

Software was tested on Unix systems as Mac OSX El Capitan 10.11.3 and Ubuntu 14.04. All tests have been run on various images from really small one up to images which contain about 190 tracks and 25 000 pits. Testing can be divided into 4 main parts. In the first part, handling of damages was tested. In the second part, recognition of tracks was tested. Recognition of pits was tested in the third part and final stream accuracy was analysed in the fourth part.

Whereas the obtaining of the image by electron microscope was very time consuming and technologically challenging, the scanning was focused on undamaged parts of the CD surface. Thus, it was necessary to simulate basic damages. The recognition of simulated damages is successful and damaged parts are ignored as shown on figure 5.4.



Figure 5.4: Filtering of damages

The system was also tested multiple times for correct recognising of tracks. As mentioned, the only premise for successful recognising is that tracks must be in horizontal position and the curvature is ignored. Testing of recognising tracks is shown on figure 5.5. There you can see that the position of track is determined with high accuracy exactly in the middle of the pits on the track.

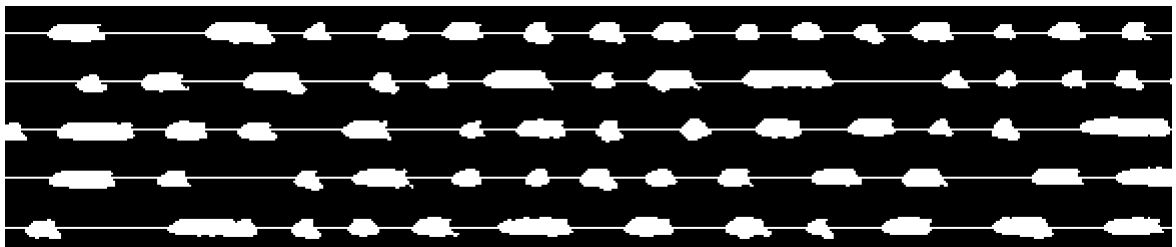


Figure 5.5: Track analysis

The next step is testing of recognition of pits and testing of the accuracy of the final stream. Lets have a testing image which contains one pit 5 pixels far from the edge of the image. The pit is 10 pixels long, then the track contains 5 pixels long gap again. One of the testing images can be seen on figure 5.6. This shall produce string 00000111111111100000.

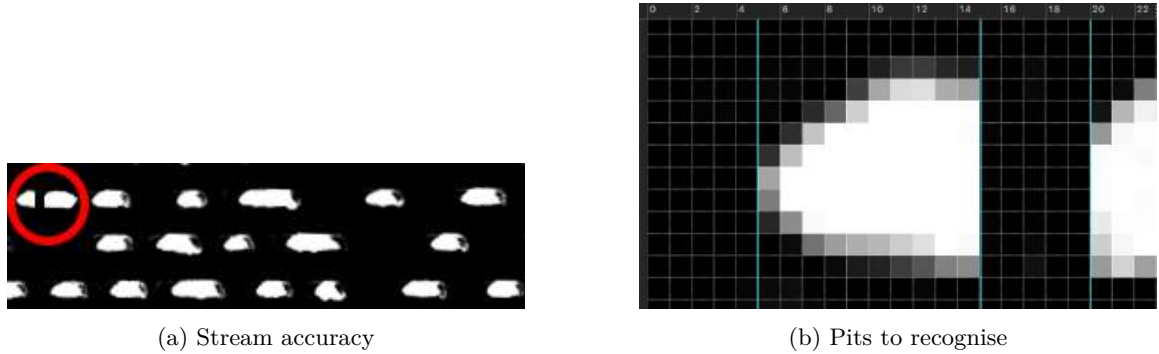


Figure 5.6: Stream accuracy

If we check the program output, we can see something similar (output is shortened). Thus, the steam is very accurate, number of 0's and 1's reflects the number of pixels. Furthermore, the number of recognised pits in output reflects the number of the pits in the image on each track.

```
1. track from 3 tracks:::::7 pits found
2. track from 3 tracks:::::5 pits found
3. track from 3 tracks:::::8 pits found
RAW STREAM:
000001111111111000001111111111111111100000000...
```

All types of described tests were run on various images with various sizes, various number of tracks and various number of pits. The success rate depends on the quality of the input image. However, if the minimal requirements are met, as in case of figure 4.5, all tests produce correct outputs. Another input images for further testing can be found on attached CD as described in appendix B.

General algorithm was described in this chapter. Some of the important methods were also further explained and basic methodology of testing was also described as well as some simple examples of testing inputs/outputs.

Chapter 6

Conclusion

This bachelor thesis dealt with a bit stream reconstruction from broken CDs/DVDs. The necessary theoretical basics were explained in this thesis as well as the procedure of scanning and the software for analysis of specimens.

Different types of disks were compared with emphasis on the composition of disks. Materials of different disks and thickness of layers were contrasted. The size of pits and lands on the data layer, laser beam and track pitch were examined. Furthermore, the organisation of data on the disk was described. Terms as F_1 , F_2 , F_3 *frame*, section, sector, EFM, NRZI, CIRC were explained as well as the data encoding and error corrections with Reed-Solomon correcting codes. Moreover, necessary basics of electron microscopy were described.

Scanning with electron microscope and preparation of specimens were also described. It was pointed to possible problems with scanning of the surface of the disk as negative charge of surface, position of data layer on the disk and size of the specimen. The thesis brought possible solutions to these problems and fractions of CD data layer could be successfully scanned and examined. Acquired images were analysed by created software in order to get a bit stream for data reconstruction. The accuracy of reconstructed bit stream, number and size of pits were tested. Basic damages of the data layer can be also handled by created software. Taking into account the size and quality of input images, software analysis and stream reconstruction were successful on every image that meets minimal input requirements defined in this thesis.

However, further development of this thesis, in order to get real data, is still very challenging. First of all, the scanning with the electron microscope did not bring expected results. In some cases, the preparation of the specimen causes damages on the surface and almost all data were lost. Furthermore, only very tiny pieces of data layer could be scanned because the scanning is very time consuming. So that, the effectiveness of image acquisition shall be further examined. Some other type of microscope could be used. Then, it would not be necessary to split the layers of the specimen and the scanning could be faster. This implicates acquisition of better images with larger areas of CD surface. Thus, software could be modified in order to follow the track with curvature. Furthermore, the recognition of damages could be tested on huge amount of real damages scanned by microscope. It could be also possible to acquire all surface of the CD by faster scanning. With all surface of the CD scanned, the output stream could be decoded with NRZI, EFM, Reed-Solomon codes, etc. . . So the real data could be reconstructed.

Bibliography

- [1] European Computer Manufacturers Association. Standard ecma-130: Data interchange on read-only 120 mm optical data disks (cd-rom). Technical report, 1988.
- [2] European Computer Manufacturers Association. Standard ecma-394: Recordable compact disc systems cd-r multi-speed. Technical report, 2010.
- [3] Vrije Universiteit Brussel. Auger electron spectroscopy - aes [online]. <http://www.surfgroup.be/aes>, [cit. 2016-04-05].
- [4] Fred R. Byers. *Care and Handling of CDs and DVDs: A Guide for Librarians and Archivists*. Gaithersburg: National Institute of Standards and Technology, 2003. ISBN 978-1932326048.
- [5] P.M. Cohn. *Further Algebra and Applications*. SpringerLink : Bücher. Springer London, 2011. ISBN 978-1852336677. LCCN 2002026862. URL <https://books.google.is/books?id=j1LmBwAAQBAJ>.
- [6] M.J. Dykstra and L.E. Reuss. *Biological Electron Microscopy: Theory, Techniques, and Troubleshooting*. Number dfl 997 in Biological Electron Microscopy: Theory, Techniques, and Troubleshooting. Springer US, 2003. ISBN 978-1461348566. LCCN 92034102. URL <https://books.google.is/books?id=JULNEc9rmUUC>.
- [7] R. Egerton. *Physical Principles of Electron Microscopy: An Introduction to TEM, SEM, and AEM*. Springer ebook collection / Chemistry and Materials Science 2005-2008. Springer US, 2011. ISBN 978-1441938374. LCCN 2005924717. URL <https://books.google.is/books?id=t6TTgqp305MC>.
- [8] J. Goldstein. *Practical Scanning Electron Microscopy: Electron and Ion Microprobe Analysis*. Springer US, 2012. ISBN 978-1-4613-4422-3. URL <https://books.google.is/books?id=zBEDCAAQBAJ>.
- [9] P.J. Goodhew, J. Humphreys, and R. Beanland. *Electron Microscopy and Analysis, Third Edition*. Taylor & Francis, 2000. ISBN 978-0748409686. LCCN 00037716. URL <https://books.google.is/books?id=zuxPIVmGLGsC>.
- [10] K. F. Ibrahim. *DVD Players and Drives*. Newnes, An imprint of Elsevier, 2003. ISBN 978-0750657365.
- [11] ISO/IEC Information Technology Task Force (ITTF). Iso/iec 10149:1995. Technical report, 1995.

- [12] H. Nakajima and H. Ogawa. *Compact Disc Technology*. Ohmsha, 1992. ISBN 978-4274033476. LCCN lc91077698. URL <https://books.google.is/books?id=9G9Nu9n0DJQC>.
- [13] Giesbert Nijhuis. Optical cd code [online]. http://www.laesieworks.com/digicom/Storage_CD.html, [cit. 2016-03-18].
- [14] Koninklijke Philips N.V. The history of the cd - the beginning [online]. <http://www.research.philips.com/technologies/projects/cd/cd-family.html>, 2015 [cit. 2016-03-18].
- [15] Koninklijke Philips N.V. The history of the cd - the beginning [online]. <http://www.research.philips.com/technologies/projects/cd/>, 2015 [cit. 2016-03-18].
- [16] Koninklijke Philips N.V. The history of the cd - the beginning [online]. <http://www.research.philips.com/technologies/projects/cd/technology.html>, 2015 [cit. 2016-03-18].
- [17] V. Pless. *Introduction to the Theory of Error-Correcting Codes*. Wiley Series in Discrete Mathematics and Optimization. Wiley, 2011. ISBN 978-0471190479. LCCN 97049834. URL https://books.google.is/books?id=4v6_Hcqbm-8C.
- [18] Andy Poggio. Cd-rom technical summary, from plastic pits to fantasia [online]. <http://bat8.inria.fr/~lang/hotlist/cdrom/Documents/tech-summary.html> | <http://www.martinreddy.net/gfx/cd.info>, 1998 [cit. 2016-04-02].
- [19] Ken C. Pohlmann. *The Compact Disc: A Handbook of Theory and Use*. A-R Editions, Inc., 1989. ISBN 978-0895792280.
- [20] Philips Intellectual Property and Standards. Iec-908: Compact disc digital audio - the red book. Technical report, 1980.
- [21] Šicner Jiří. *Srovnání algoritmů dekódování Reed-Solomonova kódu*. PhD thesis, Vysoké učení technické v Brně. Fakulta elektrotechniky a komunikačních technologií, Brno, 2011.

Appendices

List of Appendices

A	Images from electron microscope	33
A.1	Data tracks vs heat gun	33
A.2	Data layer after cyklohexane ultrasound bath	34
A.3	Edge of fragment of CD	35
A.4	Scratches	36
A.4.1	Scratch 1	36
A.4.2	Scratch 2	37
A.5	Layers of CD	38
A.5.1	Tilted data layer + metal layer + protective layers	38
A.5.2	Layers of CD (substrate)	39
A.5.3	Layers of CD (coating + ink)	40
B	Contents of CD	41

Appendix A

Images from electron microscope

A.1 Data tracks vs heat gun

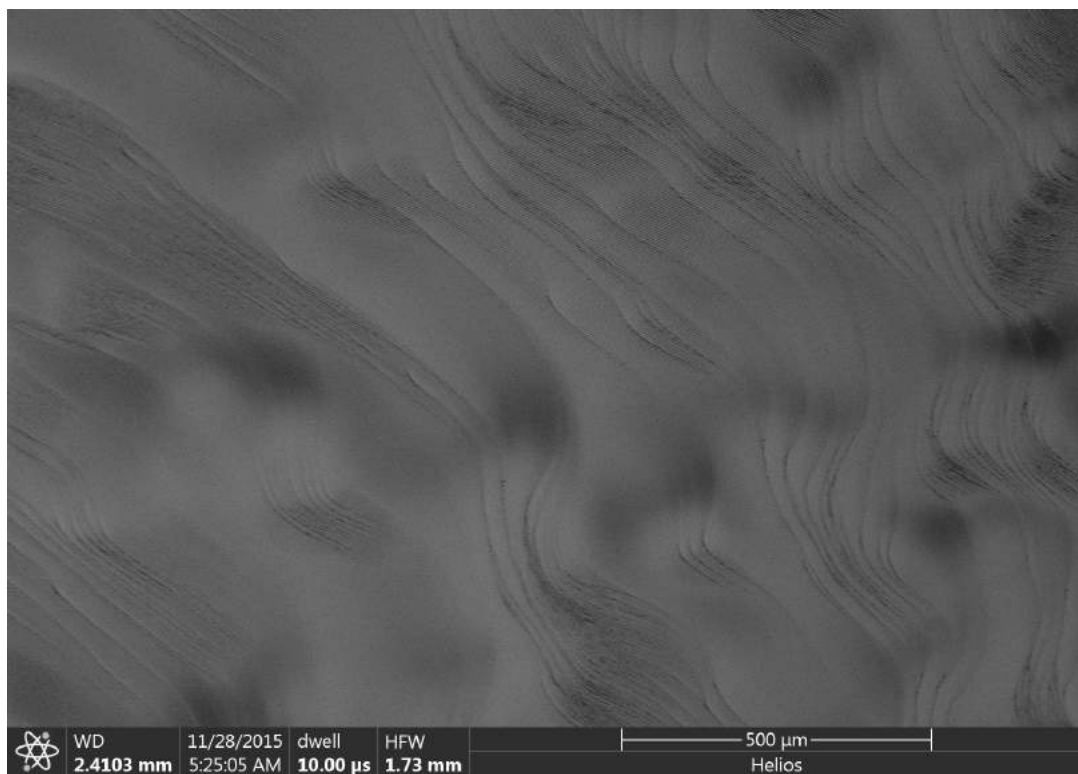


Figure A.1: Data tracks vs heat gun

A.2 Data layer after cyklohexane ultrasound bath

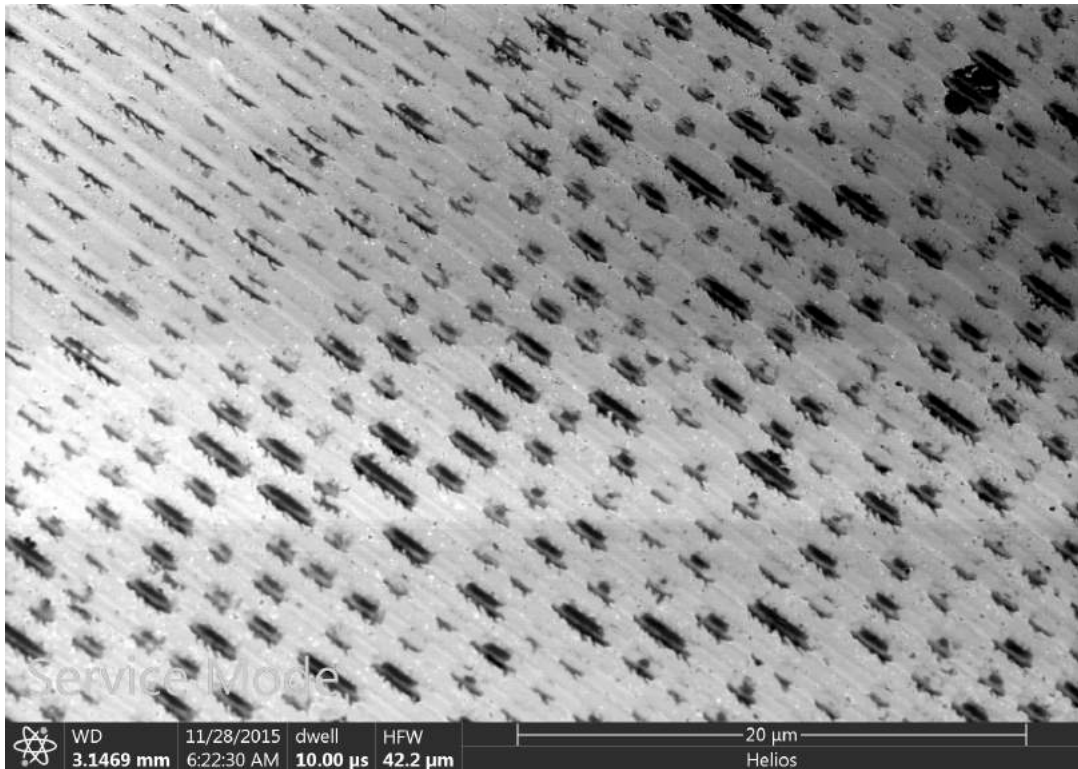


Figure A.2: Data layer after cyklohexane ultrasound bath

A.3 Edge of fragment of CD

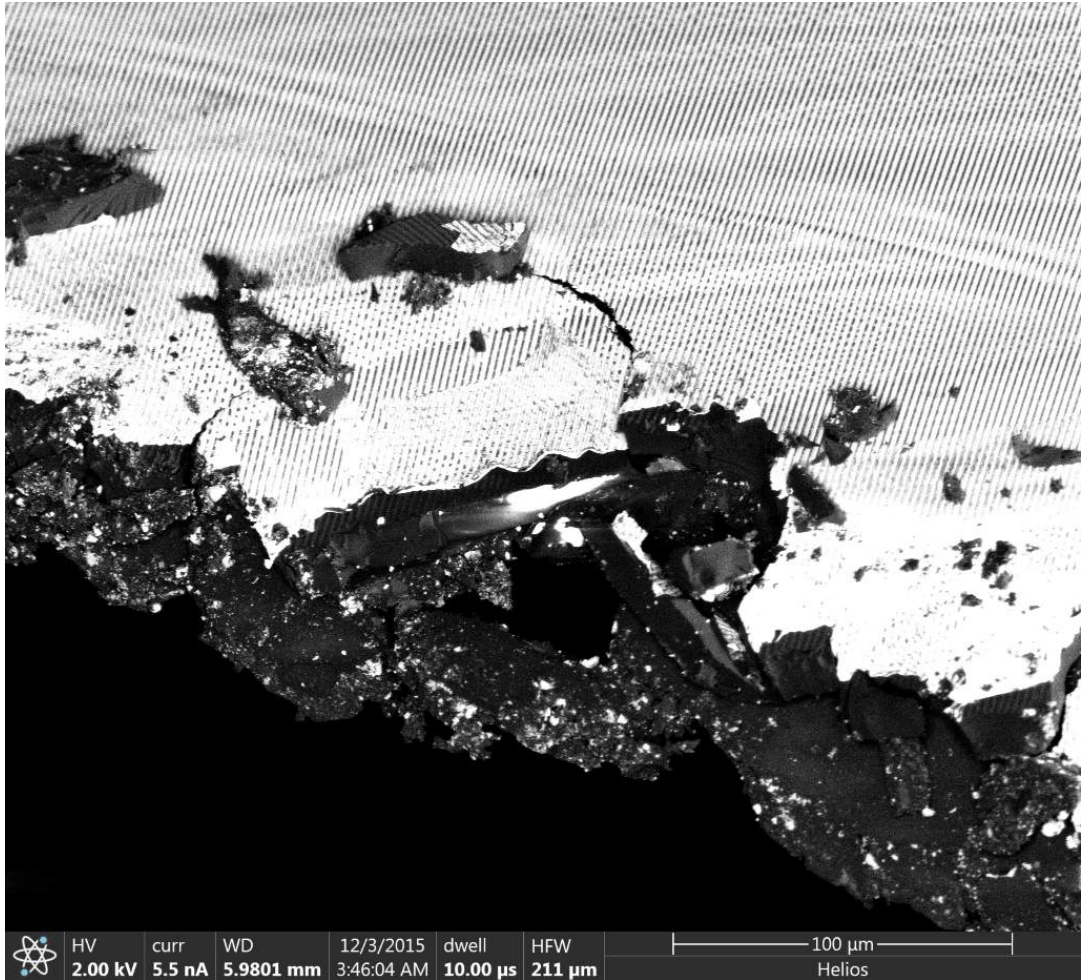


Figure A.3: Edge of fragment of CD

A.4 Scratches

A.4.1 Scratch 1

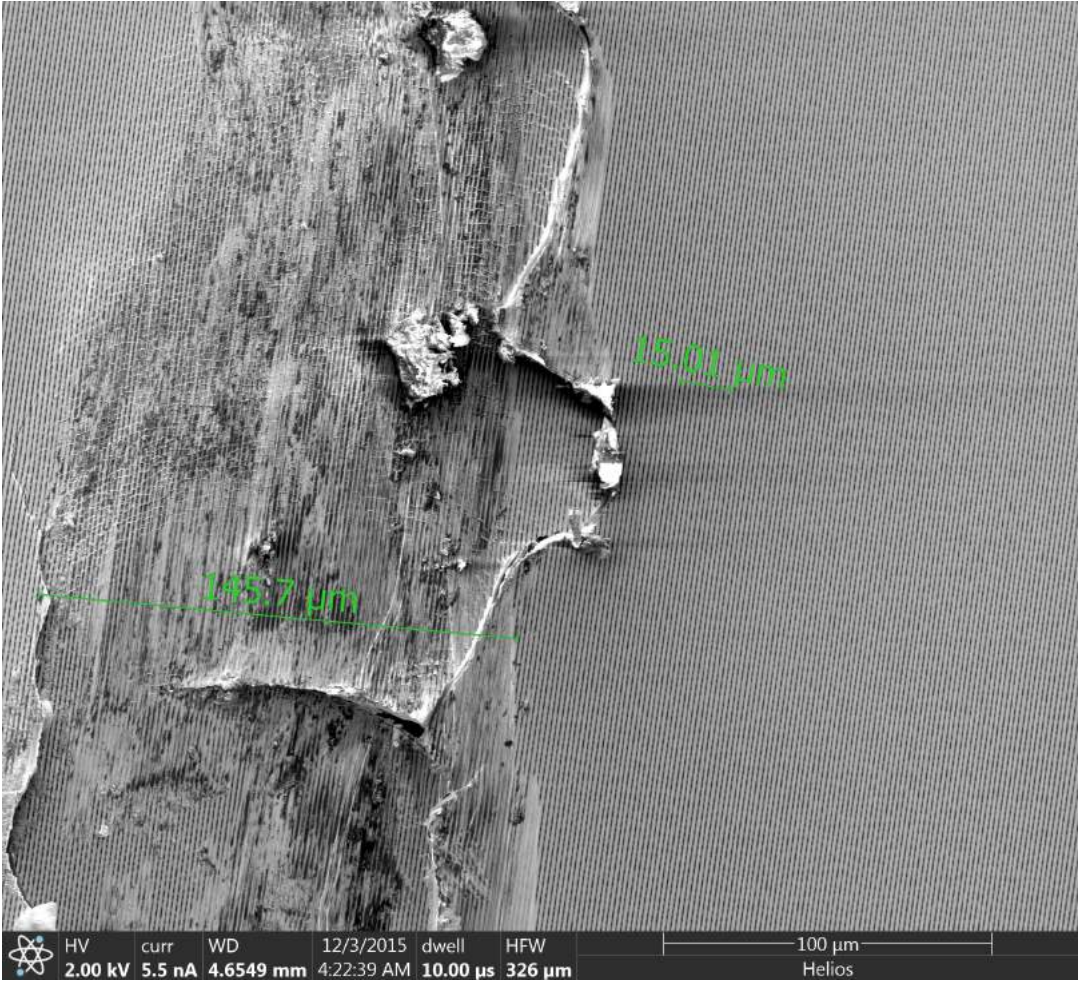


Figure A.4: Scratch 1

A.4.2 Scratch 2

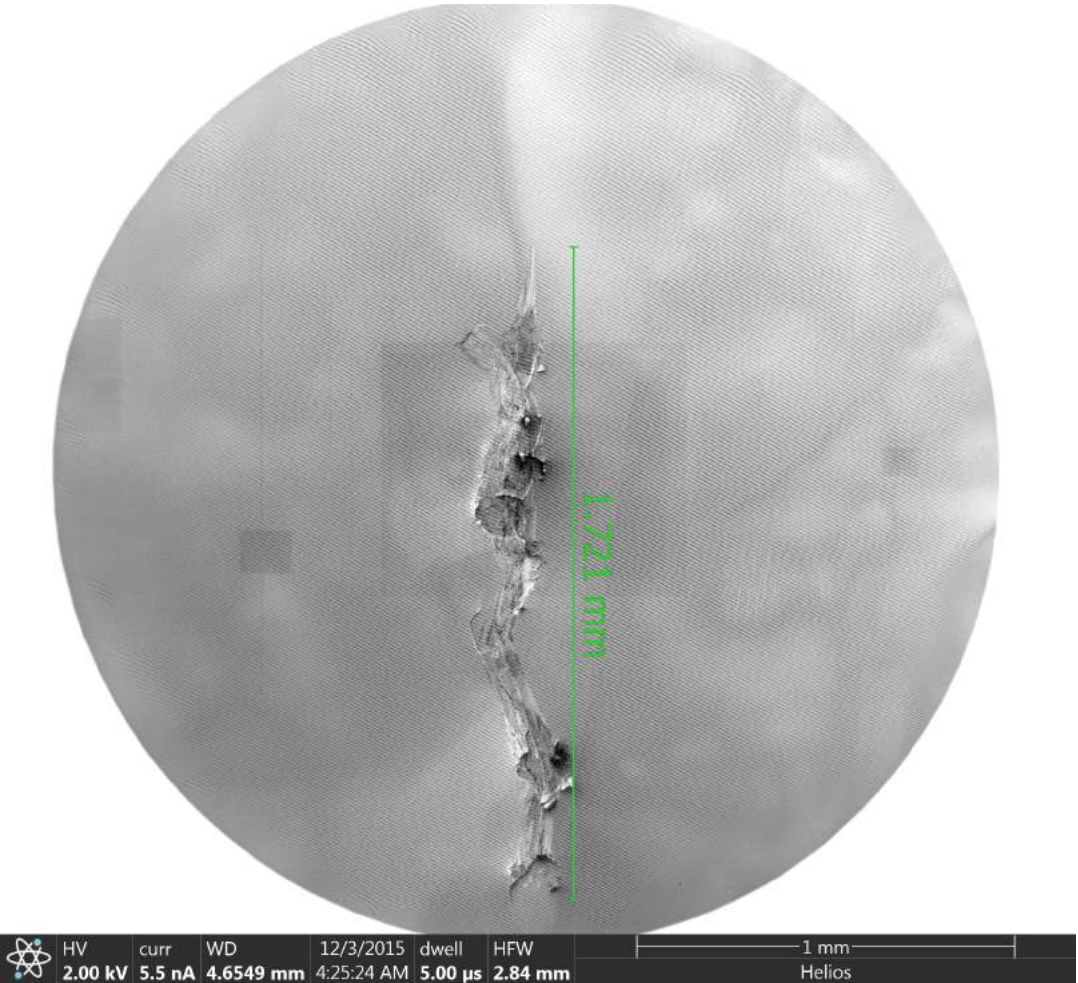


Figure A.5: Scratch 2

A.5 Layers of CD

A.5.1 Tilted data layer + metal layer + protective layers

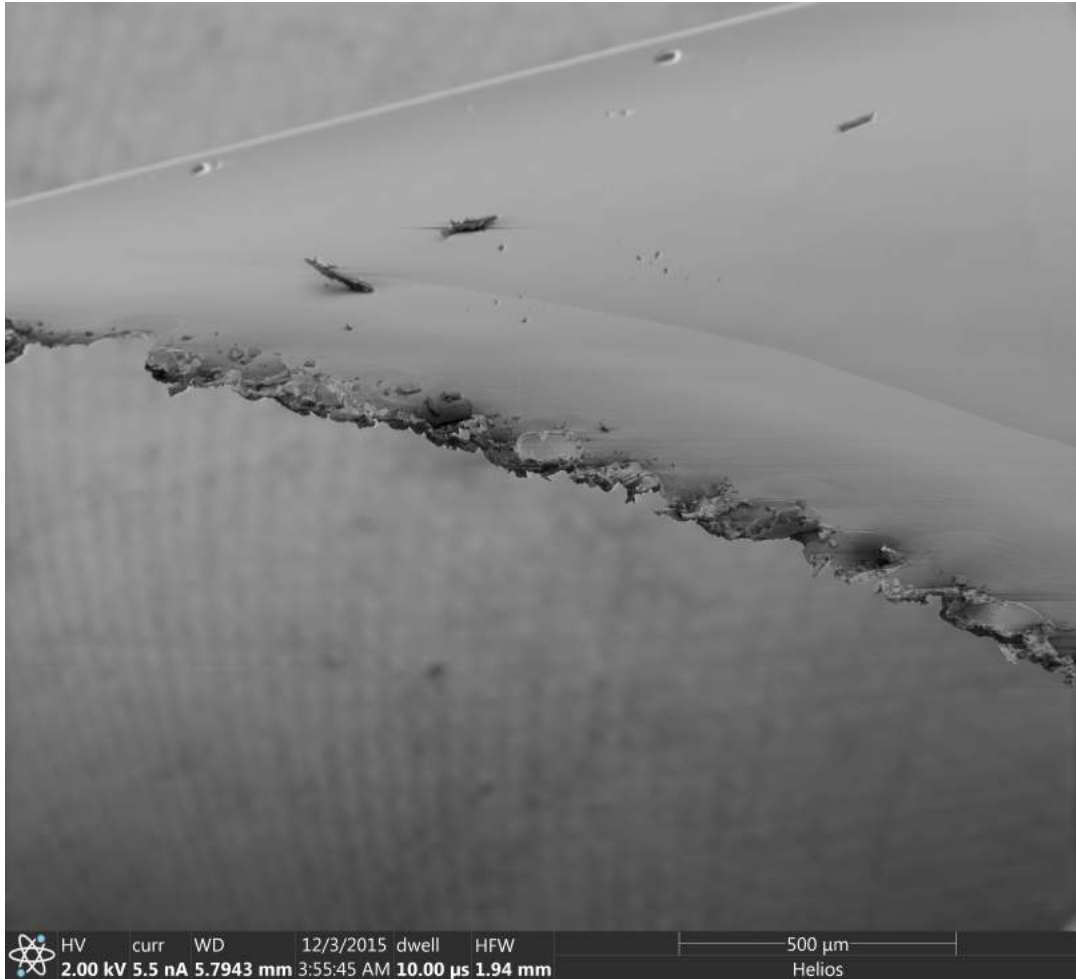


Figure A.6: Tilted data layer + metal layer + protective layers

A.5.2 Layers of CD (substrate)

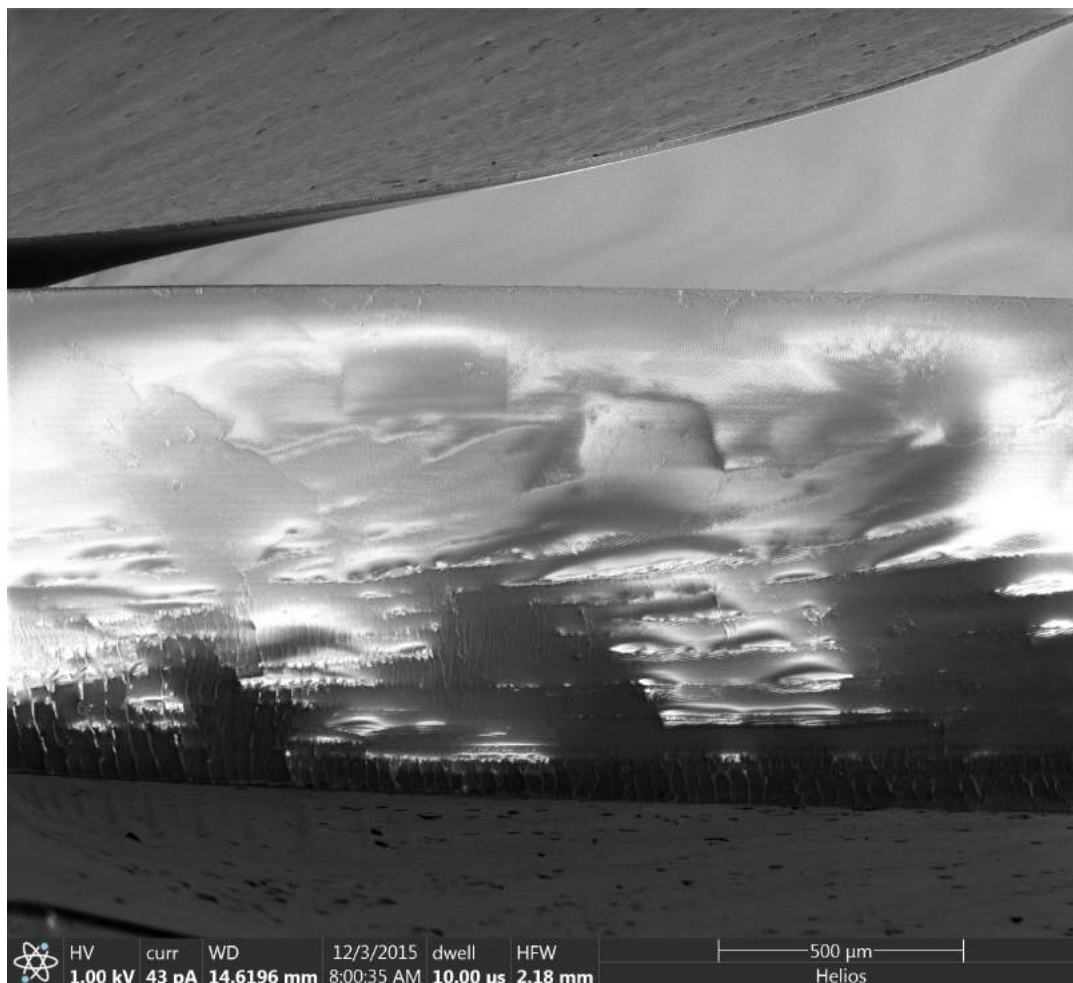


Figure A.7: Layers of CD (substrate)

A.5.3 Layers of CD (coating + ink)



Figure A.8: Layers of CD (coating + ink)

Appendix B

Contents of CD

Attached CD contains:

- Source images for software analysis
- Sample images acquired by electron microscope
- Source codes
- reconstruction.txt file describing a usage of the software
- readme.txt file describing a content of CD
- This document + document uploaded to IS