UNIVERSITÀ DEGLI STUDI DI TRIESTE

Sede Amministrativa del Dottorato di Ricerca

CNR-INFM LABORATORIO NAZIONALE TASC

Sede convenzionata

XIX CICLO DEL
DOTTORATO DI RICERCA IN
NANOTECNOLOGIE

3D LASER SCANNER BASED ON SURFACE SILICON MICROMACHINING TECHNIQUES FOR SHAPE AND SIZE RECONSTRUCTION OF THE HUMAN EAR CANAL

Settore disciplinare FIS/07

DOTTORANDO

Mauro Prasciolu

COORDINATORE DEL
COLLEGIO DEI DOCENTI
Chiar.mo Prof. Fernando Tommasini
Università degli Studi di Trieste

TUTORE E RELATORE
Chiar.mo Prof. Enzo Di Fabrizio
Università degli studi Magna Græcia di Catanzaro

2007

a Giulietta

Abstract

As technology advances, hearing aids can be packaged into increasingly smaller housings. Devices that fit entirely within the deeper portion of the external auditory canal have been developed, called completely-in-the-canal (CIC). These aids are custom moulded and have high cosmetic appeal because they are virtually undetectable. They also have several acoustic advantages: reduced occlusion effect, reduced gain requirements, and preservation of the natural acoustic properties of the pinna and external ear. However, CIC hearing aids require proper fitting of the hearing aid shell to the subject ear canal to achieve satisfactory wearing comfort, reduction in acoustic feedback, and unwanted changes in the electro-acoustic characteristics of the aid. To date, the hearing aid shell manufacturing process is fully manual: the shell is fabricated as a replica of the impression of the subject ear canal. Conventional impression acquisition method is very invasive and imprecise, moreover the typical post-impression processes made on the ear impression leaves room for error and may not accurately represent the structural anatomy of patient's ear canal. There are some laser approaches able to perform a 3D laser scanning of the original ear impression but, the entire shell-making process is completely dependent on the ear impression and often is the sole cause of poor fitting shell. Therefore, direct ear canal scanning is the only way to perform accurate and repeatable measurements without the use of physical ear impression.

The conventional optical elements are not able to enter in the inner part of the ear and perform a scanning of the cavity. This work is devoted to the direct scanning of human external auditory canal by using electromagnetically actuated torsion micromirror fabricated by micromachining technique as scanner. This is the first ever demonstration of actual scanning of human external auditory canal by a single integral Micro-Electro-Mechanical System (MEMS). A novel prototype 3D scanning system is developed together with surface reconstruction algorithm to

obtain an explicit 3D reconstruction of actual human auditory canal. The system is based on acquisition of optical range data by conoscopic holographic laser interferometer using electromagnetically actuated scanning MEMS micromirror. An innovative fabrication process based on poly(methylmethacrylate) (PMMA) sacrificial layer for fabrication of free standing micromirror is used. Micromirror actuation is achieved by using magnetic field generated with an electromagnetic coil stick. Micromirror and electromagnet coil assembly composes the opto-mechanical scanning probe used for entering in ear auditory canal. Based on actual scan map, a 3D reconstructed digital model of the ear canal was built using a surface point distribution approach.

The proposed system allows noninvasive 3D imaging of ear canal with spatial resolution in the 10 μm range. Fabrication of actual shell from in-vivo ear canal scanning is also accomplished. The actual human ear canal measurement techniques presented provide a characterization of the ear canal shape, which help in the design and refining of hearing aids fabrication approaches to patient personalized based.

Contents

1 Introduction

According to the latest scientific statistics [1], the percentage of the population with hearing problems is about 10%: a considerable figure which would mean about 70 million people in Europe and 30 million in North America while, at a global level, this would reach the incredible total of 500 million people (Figure 1.1a). A hearing deficit is a problem that mainly affects elderly people because it is often a result of the natural process of ageing and deterioration of the hearing apparatus; over half of those affected, in fact, are over 65 (Figure 1.1b).

Recent developments, however, indicate a high incidence of problems related to health factors in developing countries, while, in industrialized countries there is an increase in hearing loss among younger people because of the increasing incidence of an excessive noise pollution.

Statistics indicate that only one out of five people who need a hearing aid actually uses one. This strong resistance among patients is linked to certain traditional and almost "historical" fears of psychological nature, such as the fear of social stigma and prejudice connected to deafness and use of hearing aids, which are perceived as not very effective and difficult to use.

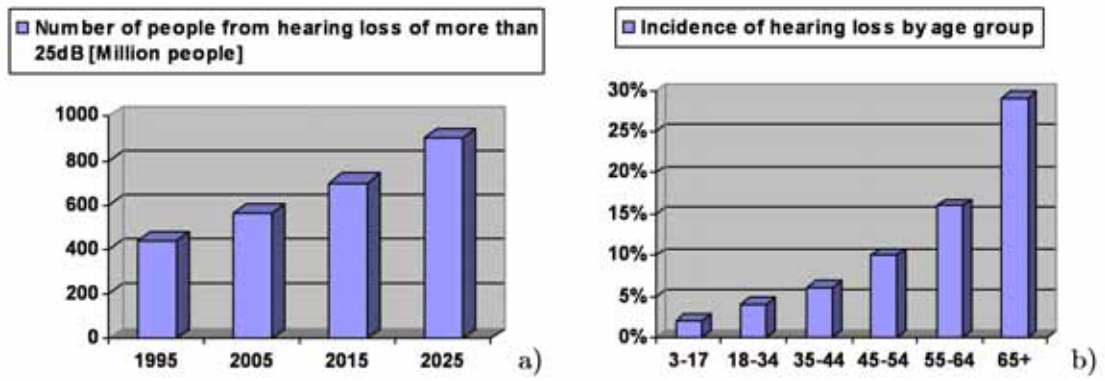


Figure 1.1: Statistics about hearing disorders, and deafness compiled from fact sheets produced by the National Institute on Deafness and Other Communication Disorders (NIDCD): a) forecast of the number of people with hearing loss of more than 25dB; b) incidence of hearing loss by age.

Hearing loss negatively impacts on quality of life, personal relationships and of course, the ability to communicate. Social isolation and inability to pay attention are much more common among hearing impaired adults aged 50 or older than among non-hearing impaired people in the same age group, according to an American study [2]. Communication problems due to hearing loss often lead to various degrees of depression by disrupting relations with other people. The study included more than 2000 people with self reported hearing loss. Even though the researches did not find statistical differences between people with mild hearing loss and those suffering from moderate or worse hearing impairment, data indicate that even milder hearing loss can have a negative impact on quality of life. Therefore it is of great importance that the hearing loss is not suppressed and left untreated, as modern hearing aids and proper guidance often mitigate negative consequences.

The vast majority of hearing problems do not require medical or surgical intervention. Some 90 to 95 percent of all cases of hearing loss can be corrected by hearing aids. A hearing aid is a device used to help hard-of-hearing people hear sounds better.

In the past, a funnel-like amplification cone, called "ear trumpet" or "ear horn", was used (Figure 1.2a). The first hearing aids were large unwieldy boxes which took up as much room as a small suitcase. They were nothing like today's small digital hearing aids (Figure 1.2b) which can be hidden inside the ear and have an almost perfect sound reproduction.

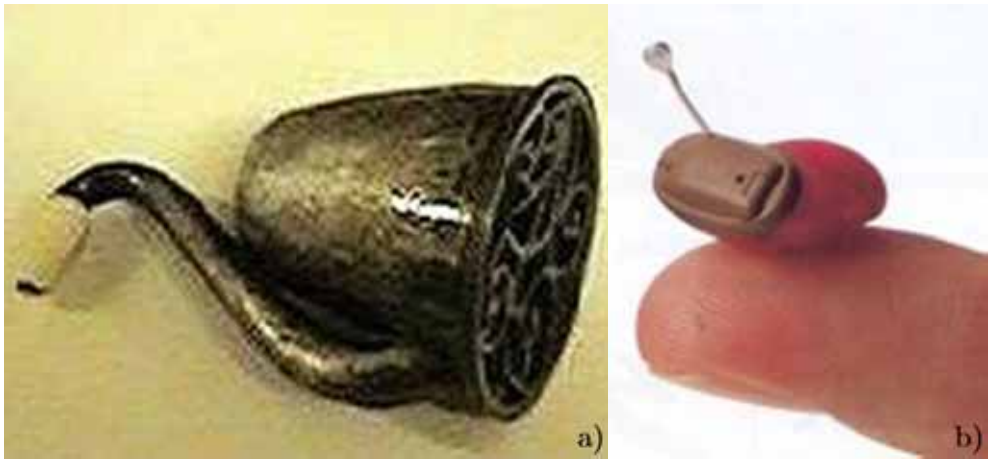


Figure 1.2: Hearing aids history: a) London Dome, hand engraved sterling silver, made by F.C. Rein in London, England about 1850. This is the oldest item in the museum. The dome measures 7.1 cm high by 6.8 cm in diameter; b) Nowadays completely in the canal hearing aid

In the beginning, hearing aids were heavy and definitely not suitable to wear. An early hearing aid typically consisted of a separate microphone, an amplifier, headphones and a bulky battery. As the device worked best when placed on a table and used with a pair of headphones, it was troublesome to use. Although the battery was large, it only lasted for a couple of hours at a time, after which the battery went dead. Furthermore, early hearing aids were expensive, and only few people could afford to buy them.

As early as 1902, three years after the first hearing aids were available, hearing aid became lighter. Hearing-impaired sufferers could save their energy and purchase a smaller portable device which could be worn, and worked better than earlier models. Although the design was smaller, amplifier and batteries had to be hung around the neck and the microphone had to be held in the hand in order to hear properly. People using these hearing aids must have attracted a lot of attention. In addition, the size of the microphone varied in accordance with the extent of the hearing loss, therefore a person with profound hearing loss was compelled to use a large microphone in addition to the hearing aid. Portable, but heavy hearing aids were used, after various improvements, well into the 1950s. However, the invention of the transistor in 1947 revolutionized hearing aid technology. Until then, valves had been used in hearing aids, but the transistor made it possible to build smaller and better ones. The amplifying property of the transistor was superior to that of valves, and as a consequence, batteries allowed to use new hearing aids for longer periods of time.

The first hearing-impaired people to switch over to new transistor-driven hearing aids had to carry the support on their body, but before long, hearing aids became so small that they could be placed directly behind or above the ear. These small devices allow people with hearing problems did not attract so much attention any more. From the 1960s onwards, small, in-the-canal hearing aids were developed. However they were not as reliable as the larger, more visible behind-the-ear hearing aids, which underwent continuous improvement and became more discreet.

Common to all hearing aids was the analogue technology used for treatment of sound. The development of digital sound treatment led to considerable improvements in the effectiveness of hearing aids. Nowadays, digital hearing aids are constructed with a small programmable computer and are capable of amplifying millions of different sound signals very precisely, thus improving hearing ability of hearing-impaired people. During the mid-1980s the first digital hearing aids were launched, but these early models were slightly unpractical. It was not until some ten years later that digital hearing aids really became successful, with small digital devices placed either inside or discreetly behind the ear.

Today, digital hearing aids make a decisive difference to the quality of life of hearing-impaired people. Unfortunately, high costs force many people to choose cheaper, analogue hearing aids instead.

There are many common types of hearing aids, which differ based on size, and location on the patient. Behind the ear aids (BTE) have a small case that fits behind the ear and conducts sound to the ear through an earmold that is custom made. BTEs can be used for mild to profound hearing losses. Recent innovations in BTEs include miniature "invisible" BTEs with thin hair like sound tubes. These are often less visible than in the ear aids (ITE) and some models keep the ear canal more open, so listeners may still utilize their residual natural hearing (most helpful for those with normal hearing in the lower frequencies). Ideal for high frequency losses, these miniature versions are generally used for mild to moderate hearing loss.

ITEs are placed in the outer ear bowl (called concha); sometimes they are visible when standing face to face with someone. ITE hearing aids are custom made to fit each individual's ear. They can be used in mild to severe hearing losses. Feedback, a squealing/whistling caused by sound leaking out of the aid and being amplified again, may be a problem for severe hearing losses. Some modern circuits are able to provide feedback regulation or cancellation to assist with this.

Completely in the canal hearing aids (CIC) are hearing aids which fit entirely within the wearer’s ear canal. These devices are the smallest hearing aids on the

market and are invisible to the average observer. Most major hearing aid brands and manufacturers now offer this inconspicuous alternative to traditional behind the ear hearing aids. CIC hearing aids are custom made to fit deep inside the individual wearers ear canal and are said to mimic the natural auditory process more closely then any other style of hearing aid. They are best suited to people with a mild to moderate hearing loss.

The main issue facing hearing aids is the occlusion effect. Occlusion effect is the unnatural, hollow sound and/or plugged up sensation that some hearing aid users report. An occlusion effect [3] occurs when an object completely fills the outer portion of the ear canal, trapping the bone-conducted sound vibrations of a person's own voice in the space between the tip of the hearing aid and the eardrum. Ordinarily, when people talk or chew, these vibrations escape through the open ear canal and the person is unaware of their existence. But when the ear canal is blocked by an hearing aid, the vibrations are reflected back toward the eardrum and increases the loudness perception of their own voice. Compared to a completely open ear canal, the occlusion effect may boost low frequency (usually below 500 Hz) sound pressure in the ear canal by 20 dB or more. There are basically only two ways to reduce or remove the occlusion effect. The most effective way is to not completely block the ear canal. This permits the bone-conduction sound generated in the ear canal to escape the ear the way it is supposed to. When someone is wearing hearing aids, the only way to do this is to create a vent hole in the earmold. It is a hole drilled completely through the hearing aid, from the outer surface to the inner surface. The amount of sound that escapes, and thus the magnitude of occlusion effect, depends upon the size of the vent. The larger the vent, the more the occlusion effect can be reduced.
Another way of reducing the occlusion effect is to use a very long and tight hearing aid, or to locate the hearing aid far down in the ear canal, very close to the eardrum.

Since CIC aids extend down into the bony portion of the ear canal, the occlusion effect may be reduced or eliminated.

For audiologists, otologists, speech therapists and other hearing professionals, the primary appeal of CIC hearing aids are their acoustic advantages. They are able to closely simulate environmental, and more importantly, speech sounds, patterns, nuances etc. This is imperative to a hearing impaired individual’s competency with the spoken word. Additionally, since CIC hearing aids are worn closer to the eardrum then their BTE counterparts, their microphones are better able to amplify and therefore to give a boost to any residual hearing the wearer may have. Because of the proximity to ear drum, less power is needed, reducing the likelihood of feedback. The advantage that is most often

touted by wearers of CIC devices, is the cosmetic appeal and inconspicuousness. Since the outer portion of the typical CIC aid is well down in the ear canal, it is much less visible, and it is practically immune to wind noises that bother conventional aid wearers. The outer ear is not blocked by any portion of the CIC aid, maintaining the ear's natural directivity for sounds. Finally, telephones can be used normally, without any special techniques. Conversely, it is important to keep in mind that CIC hearing aids also have some drawbacks. CIC hearing aids are more expensive then BTE hearing aids. Another issue that may scale up cost is that increased susceptibility to ear wax build up puts CIC hearing aids at higher risk for damage, therefore necessitating expensive repair or replacement of entire hearing aid.

CIC hearing aids require more precision to fit than conventional aids. They fit deep down in the ear canal, making the impression needs process more exacting. It is important that the ear canal be examined for conditions that would make fitting these aids unwise. Because of the exact fit required to make these aids function properly, the patient can be requested for several visits to complete fitting process.

However, CIC hearing aids require proper fitting of the hearing aid shell to the subject ear canal, in order to achieve satisfactory wearing comfort, reduction in acoustic feedback, and unwanted changes in the electro-acoustic characteristics of the aid. To date, hearing aid shell manufacturing process is fully manual: the shell is fabricated as a replica of the impression of the subject ear canal

All custom made hearing aids and earmolds are made from a "cast" of the ear. The cast is referred to as an ear impression. The hearing professional makes the ear impression in the office. It takes about 10 to 15 minutes. The impression of the patient’s ear is made injecting putty-like material in the external auditory canal (EAC).

Before attempting to take an ear impression, basic anatomy of the ear has to be known. The ear canal should be carefully examined with an otoscope. If the canal shows signs of infection, discharge or unusual amounts of wax build up, the impression does not have to be taken and the individual should be referred directly to a physician.

After otoscope analysis, a cotton dam or foam block of sufficient size to properly occlude the canal is insert. It must be slightly smaller than the opening at the primary seal in order to travel down the canal, usually past the second bend. The block must not be so small as to allow impression material to be pushed further it, which would usually result in an underfilled and distorted impression. The concha of the ear is gently pull out and up, away from the head. This will open

the canal aperture so that the block can be inserted approximately 1.5 - 2.0 cm. down the canal. The impression material is silicone or a monomer/methacrylate combination (such as Audalin). The gel is load into the syringe and the tip is placed into the ear canal and begin pushing in the plunger. As the canal fills, the syringe is withdrawn slowly, backing it out against the pressure of the material. The tip of the syringe is kept in the concha with continuing pressure until the bowl is filled and the tip is then moved up into the helix. This area is filled completely and the outer part of the helix is lightly pressed in order to bring it back to its original shape. At this point, it is allowed the impression to set in the ear for at least 5 minutes. The impression is removed by pulling gently up and slowly out of the ear canal. The impression and block come out in one piece(Figure 1.3a). The impression is carefully examined for voids and if necessary, a new impression is taken immediately using the same procedure. The impressions are then sent to the hearing aid manufacturer and this process generally takes about two weeks.

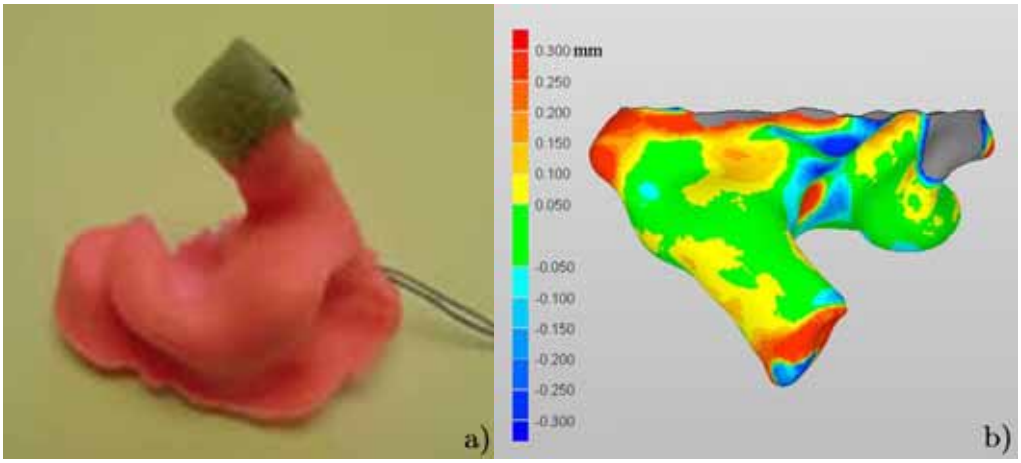


Figure 1.3: Traditional acquisition method: a) ear impression; b) metrological differences between two different impression of the same external ear canal (EAC)

The conventional method is very invasive and imprecise: the silicon gel pressure on the ear surface during hardening modifies physical dimensions and shape of the canal. A poor fitting is the cause of leakage of sound around the earmold. It provides unintentional, uncontrolled and variable venting and often leads to acoustic feedback. When leakage occurs, a portion of the output signal from the hearing instrument receiver radiates back to the microphone. To prevent acoustic feedback and feedback-related sound distortion, the earmold needs to fit exactly with the ear canal.

In Figure 1.3b is shown a digital comparison between two different impressions of the same external ear canal (EAC). In the critical parts of the molds,

the metrology is more than 500 μm different. Moreover, also the traditional post-impression processes made on the ear impression leaves room for error and may not accurately represent the structural anatomy of the patient's ear canal.

The metrological information is contained in the impression. The shipping, handling and processing the impression is slow and expensive. It is also not compatible with the software for the rapid shell modelling (RSM) which digitizes the entire design and production process of hearing aids. RSM software transforms digital metrological information into custom hearing aids that can be produced directly by a rapid prototyping machine. This integration of advanced design automation and rapid prototyping technology can reduce costs in design and production process. The most important hearing aid companies are implementing the rapid prototyping techniques for the earmold manufacture. Rapid prototyping allows automatic construction of physical objects using solid freeform fabrication. It takes the 3D digital scan of the impressions, transforms them into cross sections, still virtual, and then creates each cross section in physical space, one after the next until the hearing aid shell is finished.

The conversion of the metrological information of the impression to a 3D digital metrological information can be done with laser approaches able to perform a 3D laser scanning of the original ear impression. In any case, the entire shell-making process is completely dependent on the ear impression and, this is often the sole cause of poor fitting shell.

Therefore, no-invasive and direct digital ear canal scanning is the only way to perform accurate and repeatable measurements without the use of physical ear impression.

The aim of this work was to develop and validate an innovative 3D laser scanning system for the direct scanning of EAC. The system is able to scan EAC with a no-invasive acquisition and supply a digital output suited for the rapid prototyping techniques. Micromachining technologies are of interest to the biomedical instrumentation industry since the manufacturing technologies offer an avenue to produce miniaturized instrument system with submicron precision. Micromachining enables the realization of biomedical instrumentations previously unattainable using conventional manufacturing techniques. Therefore, micromachining technology can be used to advance biomedical instrumentation development by enabling the development of novel instrumentation and by enabling miniaturization of existing biomedical instrumentation. The key advantages of miniaturization are faster analysis times, highly parallel analysis systems, and the opportunity to manufacture disposable system using micro and nano fabrication technology.

Following the specification, a prototype, called Otoscan3D, based on surface silicon micromachined micromirror scanning probe was developed. To the best of our knowledge, this is the first ever demonstration of actual scanning of EAC by a actuated micromirror scanning system that gives digital metrological information. The first complete CIC hearing aid has been manufactured from the digital data and the patient feedback is also accomplished.

The outline of this dissertation is as follows.

The second chapter presents a review of the lithographic techniques, illustrating the working principles and characteristics of Electron Beam Lithography, X-Ray Lithography and optical lithography. Furthermore some basic aspects of wet- and dry-etching processes and deposition techniques are reported, providing a more detailed description of the Reactive Ion Etching.

The third chapter reports the different approaches followed for design, fabrication, and characterization of the MEMS micromirror. The last section include the MEMS scanning probe design and characterization.

The fourth chapter reports the basic knowledge for understanding the working mechanism of conoscopic system. In particular, the conoscopic holography system is shown.

The fifth chapter reports the prototype Otoscan3D details and the integration with the MEMS scanning probe. The motion system and the software control are described. The last section analyzed the 3D scanning techniques for the ear geometry and the 3D map reconstruction.

The sixth chapter reports the direct ear measurement. The chapter starts with topological information of the human ear anatomy. The second section regards the first in-vitro measurement and the scanning probe calibration on mechanical sample. Subsequently, the first in vivo measurement is reported. In the following section a comparison between standard method and direct digital measurement is presented. The last section describes hearing aid fabrication, including digital scan and patient feedback.

2 Nanofabrication techniques

Nowadays, miniaturization of devices is a relevant subject in many technological fields, like electronics, optics and mechanics. The request for high resolution features (less than 100 nm) is strongly increasing not only in technological research area (and linked industrial applications) but also in fundamental studies area; indeed reduced dimensions of structures give rise to many fascinating phenomena due to confinement effects, like discretization of spin-waves in magnetic nano-structures [4], discrete electron transport in quantum wires [5], quantum Hall effect [6], mesoscopic superconducting devices as qubits (quantum bits) [7, 8]. Beside the shrinking of structures' dimensions, the integration of several components in the same microdevice constitutes another key-point for nanofabrication science; by now the realization of Micro-Electro-Mechanical-Systems (MEMS), in which electronic and mechanical elements are integrated [9-11], is usual task for nanofabrication, whereas LABs-on-Chip and drug-delivery devices [12, 13], in which biological functionalities are integrated with MEMS, are very promising goals for the future. Amongst the most widely exploited nanofabrication techniques we find lithographic processes.

The name Lithography is due to their working mechanism; indeed the main step in a lithographic process is replication of a pattern on the sample. According to the employed technique, it is possible to distinguish among electron beam lithography (EBL), X-ray lithography (XRL), ion beam lithography and optical lithography. Recently a low-cost and promising process has been also proposed: imprint lithography [14]. As suggested by the names of the lithographic techniques, their main difference is the pattern transfer action, respectively achieved by means of an electron beam, X-ray radiation, ion beam, ultra-violet (UV) radiation, and simply pressure force for the imprint lithography; the difference will be explained

in the following sections. During this work, electron beam lithography and optical lithography were largely exploited for the fabrication of MEMS micromirrors.

2.1. Generic lithographic process

It is worth to describe in details the steps of a generic lithographic process, in order to point out the work needed for samples fabrication and either the reader used to some technical concepts and terms. Except for the above mentioned difference in the pattern transfer method, the lithographic processes can be divided in two main families: additive and subtractive processes (Figure 2.1). As shown in figure, first three steps are common to both processes, which are instead differentiated by the last steps: in former process material is added to the sample by deposition techniques, whereas in the latter material is subtracted from the sample by removal processes. In the following, the process is described more in details, with reference to Figure 2.1 which shows an optical lithography scheme as example.

The first step consists in the resist coating of an appropriate substrate; the choice of the substrate is determined by the final application of the sample. Typical substrates are cleaned silicon wafers or silicon coated by insulator layers (SiO_2, Si_3N_4), by metallic films or by multilayers, according to the subsequent process requests. The resist is a sensitive solution which changes its chemical properties if exposed to a suitable radiation; in particular resist solubility is modified in an appropriate solution called developer. The spin coating step is usually ended with a thermal treatment of the sample, in an oven or onto a hot-plate, in order to evaporate the solvent of the resist solution. For the second step, the exposure, a mask is placed between UV source and the sample. According to the desired pattern, some adsorbing and transparent regions are present on the mask; as a consequence only some regions of the sample are exposed by UV radiation. During the exposure, the solubility of the resist is changed only near the incident radiation, that is to say that only the exposed areas are chemically modified.

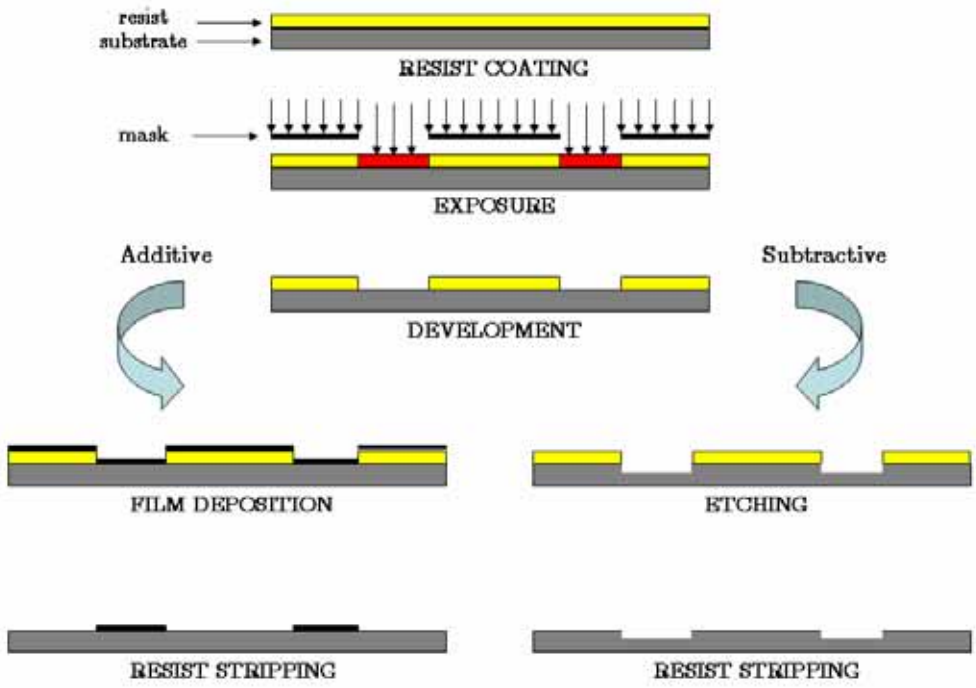


Figure 2.1: Schematic representation of the steps of a generic lithographic process; the first three steps are common to both the additive and subtractive process, the last two steps are different: on the right the additive process, on the left the subtractive one.

Afterwards, in the development step using a positive tone resist, only the exposed resist is removed, thus producing a resist pattern that is a replication of the mask pattern. Nevertheless, resist patterns produced by lithography are not permanent elements of final devices, but only temporary copies of the desired features. Resist structures are subsequently exploited to define the final stable structures. At this point, additive and subtractive processes are differentiated. In the former, a material deposition is performed over all the sample, covering both the resist structures and the substrate; at the end a resist stripping step (by means of a suitable solvent) removes all the resist, thus releasing the final structures. This process is commonly called lift-off. In the subtractive process instead, the resist is used as a mask in a selective removal step, called etching. Etching process can be chemical, in which chemical reactions with the substrate are exploited for its removal, or physical, in which all the sample is ion-bombarded, or a combination of both chemical and physical processes. Finally the resist is stripped away and only the desired permanent structures remain deposited on the sample. The lithographic process as presented above is the simplest one; in order to

fabricate a real device is usually requested a longer procedure. It is necessary to reiterate more times the steps for different patterns, combining additive and subtractive processes. Furthermore, as the lithographic techniques just described intrinsically produce two-dimensional patterns with a finite thickness, more complex three-dimensional structures will require different approaches to be used (see for example [15]). Furthermore, lithographic processes can also be divided in serial and parallel writing techniques (Figure 2.2). The pattern is prepared with a computer aided design (CAD) software and exposed directly on the resist; in the latter the pattern is on a mask that shadows the incoming radiation and is replicated in one single shot on the resist (Figure 2.1). Electron beam and ion beam lithography are serial writing techniques, in which the beam is focused on the sample, and deflected according to the CAD pattern; this means that the structures are exposed sequentially, one by one. X-ray and optical lithography are instead parallel writing processes, in which the mask pattern is replicated with a single shot exposure. Parallel process is faster than the serial one and it allows to replicate a lot of times the same pattern by a mask usually fabricated by means of a serial writing lithography, and in this case the mask pattern can not be modified. On the contrary, serial writing technique is more time consuming, but allows to change easily the pattern by means of CAD software.

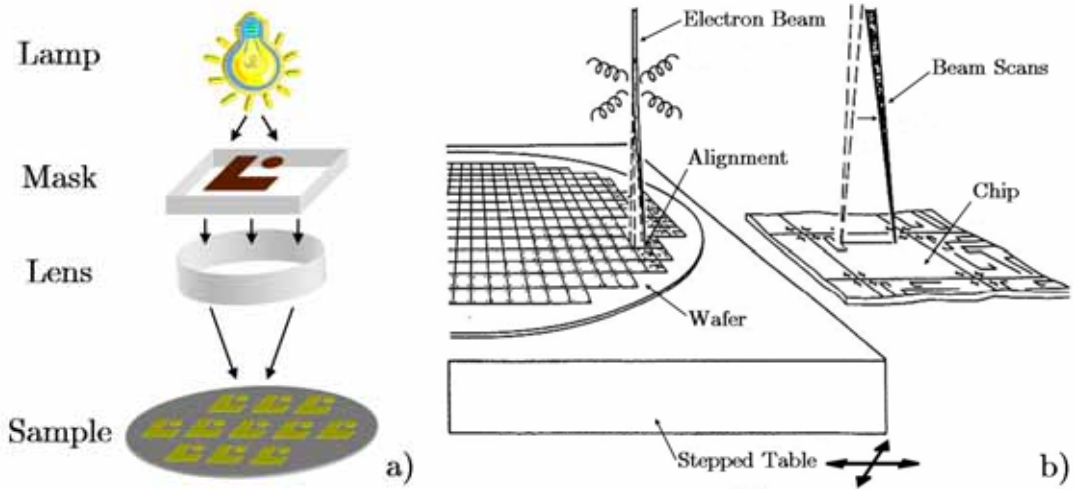


Figure 2.2: Parallel (a) and serial (b) writing techniques.

2.2. Resist

The resists are radiation sensitive compounds, the name is due to the fact that they have to be etching resistant. The principal components of resist are: a polymer (base resin), a sensitizer and a casting solvent. The polymer changes structure when exposed to radiation; the solvent allows spin application and formation of thin layer on the substrate; sensitizer controls the chemical reaction in the polymeric phase. Resists are usually classified in positive or negative, according to the response to radiation. In a positive tone resist, the chemical reaction during the exposure typically weakens the polymer by the rupture or scission of the main and side polymer chains and the exposed resist became more soluble in the developing solutions. In a negative tone resist, the reaction strengthens the polymer by random cross-linkage of many chains or pendant side chains, becoming less soluble.

A well-known family of positive resists is the poly(methymethacrylate) (PMMA) resists (Figure 2.3). PMMA becomes soluble through chain scission under illumination. PMMA is used in electron beam, ion beam and X-ray lithography.

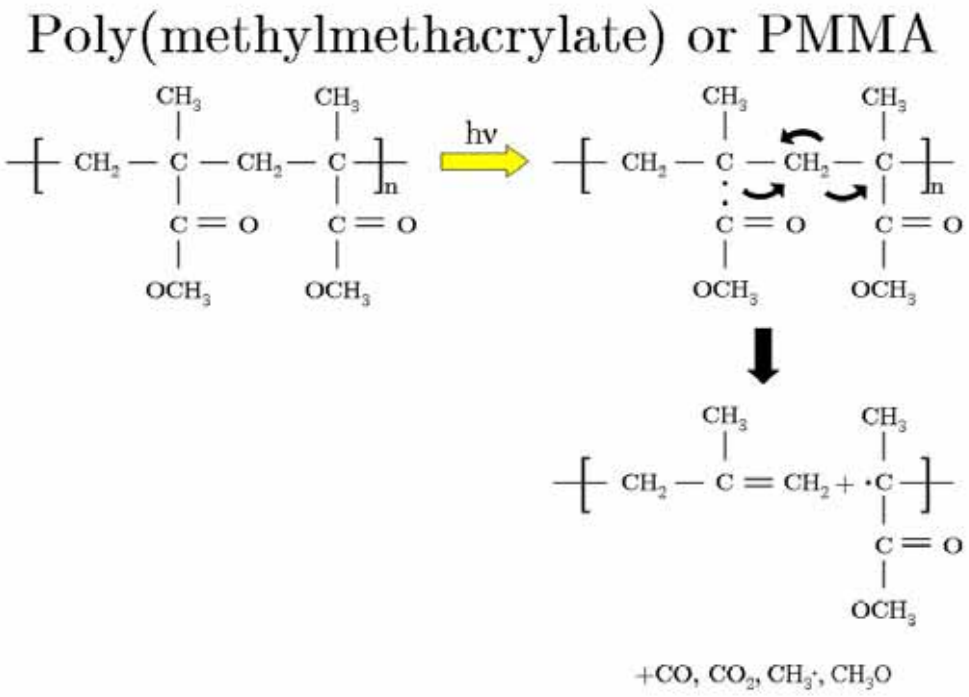


Figure 2.3: Photo-induced chain scissor of PMMA resist

The resolution capability of a resist is directly related to resist contrast γ. For positive resist, the contrast is related to the rate of chain scission and the rate of change of solubility with molecular weight. After development, the thickness of the exposed resist layer decreases until, at a critical dose D_p, the film is completely removed. Lithographic sensitivity D_p and contrast can be obtained from the response curve. A typical response curve for a positive resist is shown in Figure 2.4.

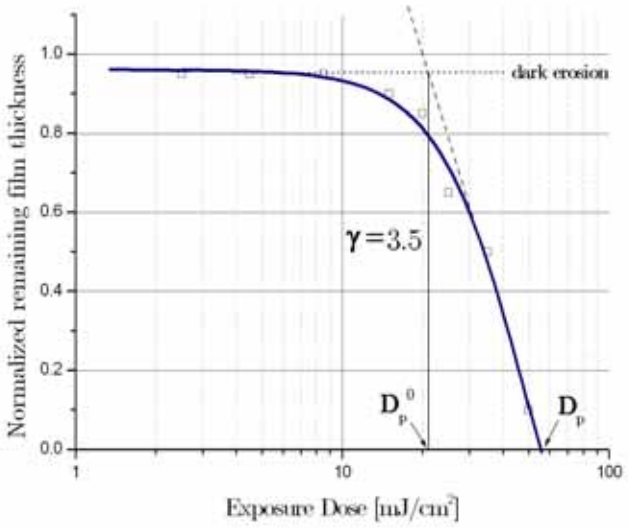


Figure 2.4: Typical exposure response curve for positive resists.

To construct the response curve a series of positive resist pads of known area are subjected to varying doses and developed in a solvent that does not attack the unexposed film. The thickness of the remaining film in the exposed area is then measured and normalized to the original thickness and plotted as a function of cumulative dosage. Contrast γ_p is determined from the slop of this sensitivity or exposure response curve as:

$$\gamma_p = \frac{1}{\left(\log D_p - \log D_p^0\right)} = \left[\log \frac{D_p}{D_p^0}\right]^{-1} \qquad \text{(Eq. 2.1)}$$

and D_p is the x-axis intersection. For given developer, D_p corresponds to the dose required to produce complete solubility in the exposed region while not affecting the unexposed resist. D_p^0 is the dose at which the developer first begins to attack the irradiated film. For dose less than D_p but higher than D_p^0, another developer

could “force develop” the resist. In force developing, the developer attacks or thins the original, unexposed resist. This describes how the profile of a positive resist can be manipulated by the operator.

Resists with higher contrast result in better resolution than those resists with lower contrast. This can be explained as follows. In an exposure, energy is delivered in a diffusive manner due to diffraction and scattering effect. Some areas outside the exposed pattern will receive an unintended dose higher than D_p^0 but lower than D_p. The resultant resist profile will exhibit some slope after the development. The higher contrast of the resist, the more vertical the resist profile. Since line width is measured at a specified height above the resist/substrate surface, as the resist profile becomes less vertical, the resist line width represents the original mask dimension less accurately.

Value of D_p^0 and D_p are figures of merit used only to compare different resists. For lithographic sensitivity numbers to have any value at all, they must be accompanied by a detailed description of the conditions under which they were measured.

2.3. Electron Beam Lithography

The electron beam lithography (EBL) is a serial writing technique, as mentioned above, and the pattern is directly exposed on the resist without any mask; as suggested by its name, an electron beam performs the exposure process. As no mask is required, EBL technique is very flexible and it allows to modify or completely change the desired pattern, according to new requests or restrictions that arise during an experimental work. For this reason it is a suitable technique when implementing new devices or facing exploratory works; furthermore it is the usual technique for fabrication of X-ray masks. On the other hand, EBL is a time-expensive process compared to parallel writing techniques, and it is not a convenient choice when a large production of well established patterns is required. During EBL exposure the interaction between electrons and resist causes the resist to be exposed, and later removed in the development (we consider to the case of a positive tone resist). The mechanism that allows the exposure of a positive resist is the breaking of the polymeric chains due to electron collisions; consequently the average molecular weight of the resist is reduced in the exposed areas, thus increasing their solubility. The result of an exposure is a pattern drawn in terms of solubility in order to produce a contrast as large as possible, where the contrast is the ratio of the exposed areas dissolution rate to the dissolution rate in unexposed

areas. It is clear that the goal of a physical theory of electron lithography is the computation of the three-dimensional molecular weight (or dissolution rate) distribution in the resist.
In the following section first it is presented a schematic representation of an EBL system and later some basic physics of interactions between electrons and matter is reported, keeping in mind that the final task is the dissolution rate distribution.

2.3.1. EBL apparatus

A schematic representation of a typical EBL system is reported in Figure 2.5: main components are an electron source (gun), apertures and electromagnetic lenses for tuning the beam, and a blanking element for a fast switch of the beam between on and off state. The electron gun is usually a thermionic or a field emission element, as these two emission mechanisms produce the brightest electron beams; just after the gun there is the alignment system coils for gun alignment, followed from the first condenser lens. Then we find the blanker plates: an electrostatic potential is applied to the plates in order to deflect the beam out of the fenditures, thus preventing the electrons to reach the sample; hence, switching on and off plates potential allows to close and open the beam over the sample in a very fast way. After the blanker there are zoom electromagnetic lenses and in the last part of the electron column stigmator and focus lenses are placed together with the deflector coils. It is possible to state [16] a close analogy between electron optics and standard light optics; in this framework electron beam is thought as a light beam, while the electron optics as their optical counterparts. Hence most of the working principles of an electron column can be understood with reference to the more common optical situations. The intensity distribution of the produced beam, on a plane perpendicular to the beam axis, is a circular gaussian, whose width is usually referred to as the dimension of the beam spot; this is of course a fundamental parameter for the writing resolution. Furthermore, pattern ability, that is the achieved resolution, is dramatically affected by beam conditions, and a successful exposure requires a well tuned beam; particular care must be taken in aligning all the fenditures of the column to avoid an astigmatic beam, i.e. a gaussian spot that is not circular shaped. At the end of the electron column is placed the stage with the sample; the pattern is exposed over the sample both deflecting the electron beam (for small displacements) and moving the stage (for large displacements); of course during the exposure the beam is blanked while moving from a structure to be exposed to the following one. All this operations are automatic and computer assisted.

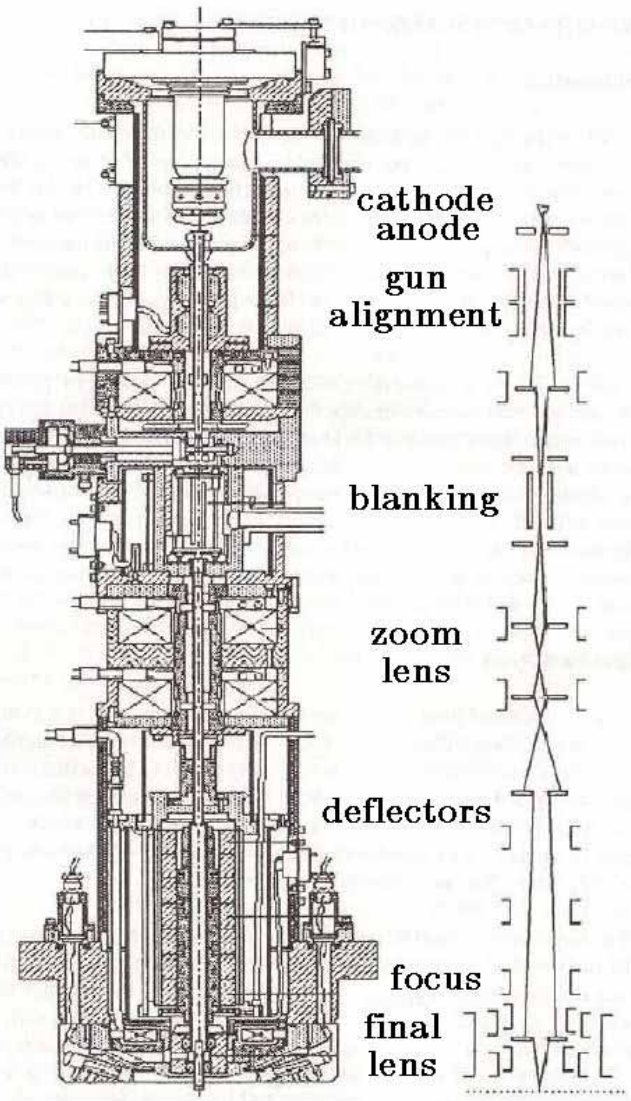


Figure 2.5: Cross-section drawing of an EBL column with the schematic ray-trace for the electrons.

2.3.2. Interactions between electrons and matter

In the introduction of this section it was pointed out that the goal of a physical theory of electron lithography is computation of the three-dimensional solubility distribution in the resist. For this reason, our interest has to be focused on the interaction between electrons and resist, with particular care to those events that cause the breaking of polymer chains. To ease this problem we consider that a constant amount of energy is absorbed when one act of breaking the resist molecules occurs (experimentally 80-100 eV [17]); so our problem reduces to that of

the dissipated energy distribution of electrons. Standard beam energies in electron lithography range from 10 up to 50 and sometimes 100 keV ;the range of 10-30 keV electrons in polymeric resists is 5-10 μm and the resist thickness is usually less than 1 μm. This means that electrons easily reach the substrate, where a fraction of them is scattered back in the resist layer, exposing again the resist.

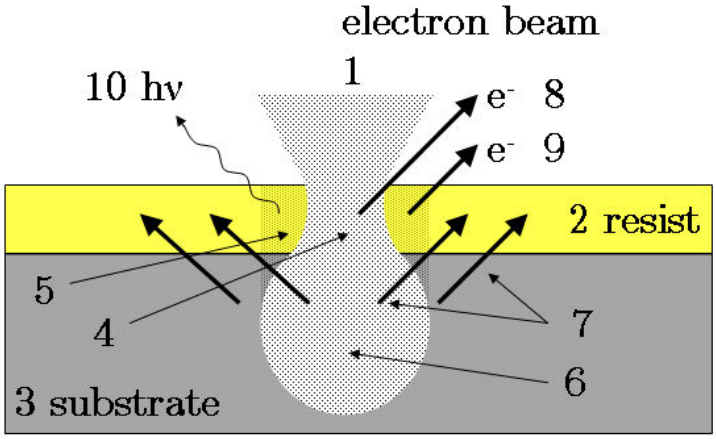


Figure 2.6: Interactions of an electron beam with a resist-substrate sample: 1) electron beam, 2) resist layer, 3) substrate, 4-5) resist regions exposed by electrons passing straight through and by the electrons reflected from the substrate, 6) electron blur in the substrate, 7) reflected electrons, 8-9) electrons leaving the resist, 10) emitted photons.

Few electrons that are scattered outward, i.e. reflected, leave the resist; further secondary electrons, optical and X-ray radiation also occur. All these processes, shown in Figure 2.6, may influence energy distribution that we are interested in. It is worth to deepen the penetration of electrons in the resist. When an electron enters the resist, there is a chance that it will be scattered by the first atom encountered on its path. The scattering process can be elastic or inelastic; in the former case the electron is scattered from the heavy nucleus of the atom and changes direction but preserves the energy. In the inelastic case the interaction is with the electron cloud of the atom, and not only direction but also energy is changed. Because of the energy dissipated by the primary electrons (that is the incident electrons of the beam) in the inelastic process, the atom can be excited (one of its electrons goes to an excited level) or ionized (one of its electrons leaves the atom, creating a secondary electron traveling in the material). As the atom is part of a resist molecule, its excitation or ionization causes the breaking of a molecular chain in the resist, thus exposing the resist itself. At this point it seems rather complex to compute analytically the electrons trajectories and the dissipated energy distribution. Indeed it is more convenient to approach the problem by

means of a Monte Carlo method of statistical trials; for this purpose the mean free path length has to be considered:

$$\lambda = \frac{1}{\sum_j N_j \left(\sigma_j^e + \sigma_j^n e\right)} \qquad \text{(Eq. 2.2)}$$

where σ_j^e and $\sigma_j^n e$ are the total elastic and inelastic scattering cross sections and N_j is the volume density of j atoms. Finally, if the scattering cross sections are known, it is possible to model electron trajectories and dissipated energy by Monte Carlo simulations [18].

In the following, we state the scattering problem approach to explicit calculation of scattering cross sections. A full explanation of all the calculation problems and their solutions is far from the tasks of the present dissertation (the interested reader can refer to [17]), however we report some important results which allow to implement the Monte Carlo method. Finally the typical dissipated energy distribution is commented (proximity effect). In Figure 2.7 simulations for two different beam energies are reported; it is clear how initial energy of electrons affects the spreading inside the target. Figure 2.8 instead shows the absorbed energy distribution as obtained by a Monte Carlo simulation; a double-gaussian curve is used for the fitting procedure (the physical meaning of the double gaussian will be cleared in the next section).

2.3.3. Proximity effect

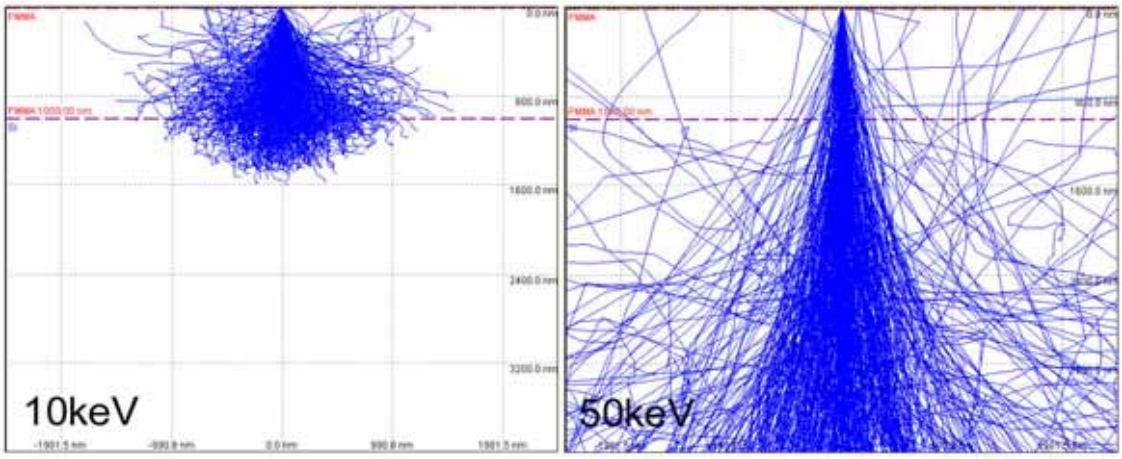


Figure 2.7: Monte-Carlo simulations for a 10keV electron beam (on the left) and a 50keV beam (on the right);the sample is a 1µm PMMA resist layer on a Si substrate. As evident, for higher energies the beam spreading occurs in the substrate, thus not causing the exposure of the resist.

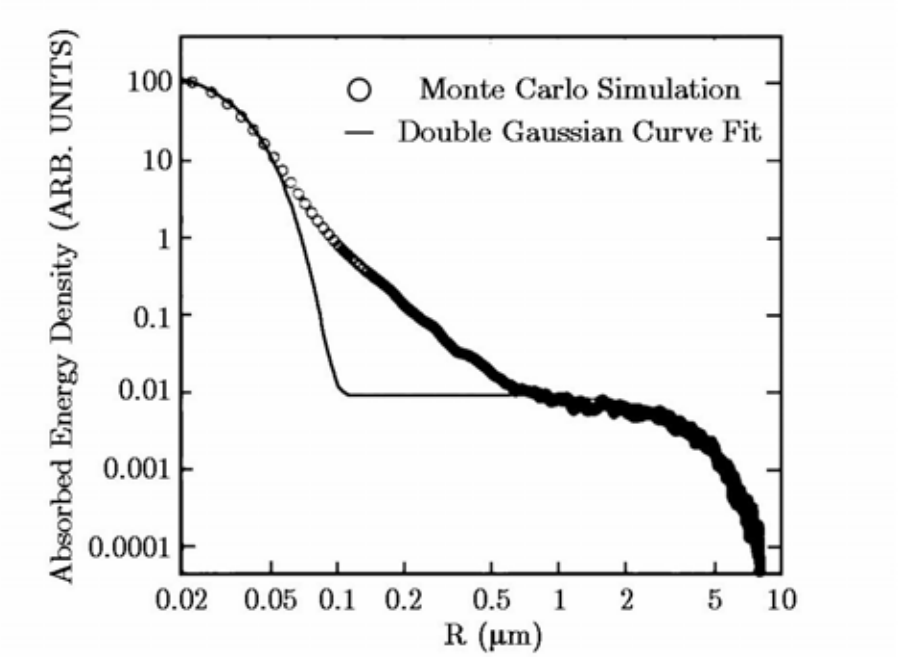


Figure 2.8: Absorbed energy as a function of the distance from the electron beam incident point: the open circles are the result of a Monte-Carlo simulation, the solid line is the double-gaussian fit of the simulation's result.

We shortly summarize the results shown above. When an electron beam penetrates the resist, scattering phenomena occur. We can distinguish between elastic and inelastic collisions: in the former case electrons change their directions without any dissipation of energy, i.e. no resist exposure is achieved; in the latter, energy is transferred to the resist, causing the breaking of molecular chains and the exposure of the resist. Furthermore, secondary electrons with low energies (compared to the energy of incident electrons) are produced by inelastic collisions, due to Auger effect and ionization of the atoms. As the range of 10-30 keV electrons in the resist is in the range 5-10 μm and the usual resist thickness is less than 1μm, a large part of electrons arrive to the substrate. The scattering with the substrate causes many electrons to be reflected in resist layer, thus exposing again the resist. The spreading of the beam due to the scattering of the electrons when they pass the first time through the resist is referred to as forward scattering, whereas the spreading due to the electrons reflected by the substrate is called back-scattering. The proximity effect is the loss of resolution due to both the forward and the back-scattering. We consider now an electron beam incident on a resist layer and try to evaluate the shape of exposed resist. As starting point we have a Gaussian distributed beam, which is the beam produced by a typical EBL apparatus; it is possible to show [17] that the density of energy dissipated when the

beam passes through the resist is Gaussian distributed in every plane perpendicular to the beam axis, with increasing width of the exposed region as the beam penetrates more deeply in the layer. Calculating the contributions due to the reflected electrons is much more complicated and some attempts [19] produced only qualitative results. In this case Monte Carlo simulations are very helpful and in practice analytical functions, fitting the Monte Carlo results (Figure 2.8), are used to take into account the back-scattered electrons contributions. At this point we can introduce a proximity function *f(r)* as the energy density distribution produced by a δ shaped electron beam; the results of Monte Carlo simulations can be approximated by the sum of two Gaussian functions:

$$f(r) = C\left[\frac{1}{\beta_f^2} e^{-r^2/\beta_f^2} + \frac{\eta}{\beta_b^2} e^{-r^2/\beta_b^2}\right] \qquad \text{(Eq. 2.3)}$$

The widths of the two Gaussian functions, β_f and β_b, are quite different: $\beta_f << \beta_b$. The first Gaussian function takes into account the contributions of the forward scattering, while the second function is associated with the back scattering contributions. The η parameter describes the ratio of the back-scattering dissipated energy to the forward scattering energy, while is a normalizing factor.

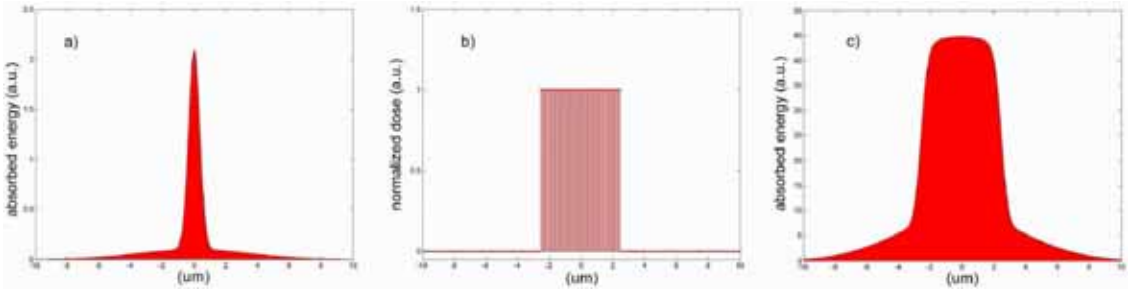


Figure 2.9: Exposure of a rectangular pattern (one dimensional scheme): a) Proximity function, b) pattern scanned by the electron beam, c) the result of the exposure. Exposed resist (c) is the convolution of the proximity function (a) with the pattern profile (b).

The proximity function for a real (not shaped) beam is the convolution of Eq. 2.3 with the intensity distribution function over a transverse section of the beam. Convolution of a Gaussian beam of radius r_b with the Eq. 2.3 gives as result again a function of the Eq. 2.3 type, but with the following parameters:

$$\beta_f' = \left(\beta_f^2 + r_b^2\right)^{1/2} \qquad \beta_r' = \left(\beta_r^2 + r_b^2\right)^{1/2} \qquad \text{(Eq. 2.4)}$$

When writing a pattern, the energy delivered to the resist, i.e. the exposure dose, will be the convolution of the proximity function with the pattern structures; as the pattern is divided in rectangular pixels before exposure, we can limit ourselves to the case of a single rectangle pattern. In Figure 2.9 qualitative results are reported for the exposure of a rectangle; as shown there are exposed areas outside the desired pattern, due to the proximity effect. The delivered dose decreases as the distance from the edges of the pattern increases, following the profile of an error function; in fact if we consider the simpler case of a mono-dimensional rectangle pattern, that is a segment *p(x)*, and calculate the convolution of the mono-dimensional proximity function, *f(x)*, with the segment, it is easy to find the delivered dose *D(x)*:

$$p(x) = \begin{cases} 1 & x_1 < x < x_2 \\ 0 & elsewhere \end{cases} \qquad \text{(Eq. 2.5)}$$

$$f(r) = C\left[\frac{1}{\beta_f} e^{-x^2/2\beta_f^2} + \frac{\eta}{\beta_b} e^{-x^2/2\beta_b^2}\right] \qquad \text{(Eq. 2.6)}$$

$$D(x) = C\left\{\sqrt{\frac{\pi}{2}}\left[erf\left(\frac{x_2 - x}{\sqrt{2}\beta_f}\right) - erf\left(\frac{x_1 - x}{\sqrt{2}\beta_f}\right)\right] + \eta\sqrt{\frac{\pi}{2}}\left[erf\left(\frac{x_2 - x}{\sqrt{2}\beta_b}\right) - erf\left(\frac{x_1 - x}{\sqrt{2}\beta_b}\right)\right]\right\} \qquad \text{(Eq. 2.7)}$$

In the two-dimensional case similar results are obtained.

When high resolution, high density structures are desired, the proximity effect must be taken into account and corrections for the addressed dose must be calculated in order to have well defined and well separated shapes. Purpose of the corrections is to obtain a roughly uniform delivered dose for the pattern, with fast decay of absorbed energy at the edges of desired structures. Consequently the pattern really exposed will be different from the desired one; in general, the corrected pattern will require a multi-level exposure (in which each level has a different dose), also if the desired pattern is a single level one. In the past several methods have been proposed to calculate proximity corrections, like the Matrix method [20], the Fourier analysis [21] and the Optimization method [22].

2.4. X-Ray Lithography

The X-ray lithography (XRL) is a parallel writing technique: a mask is used to transfer the pattern to the resist layer, as shown in Figure 2.10. The main elements are the X-ray source, the mask and the sample with the resist layer. The mask is made up of transparent and opaque areas for the X-radiation, according to the desired pattern. As the X-ray source irradiates all the mask, only the resist areas corresponding to transparent mask regions are exposed; usually a gap is present between the mask and the resist layer. This technique is referred to as proximity lithography; of course the best feasible resolution is decided by the diffraction phenomena and hence by the wavelength of the used radiation.

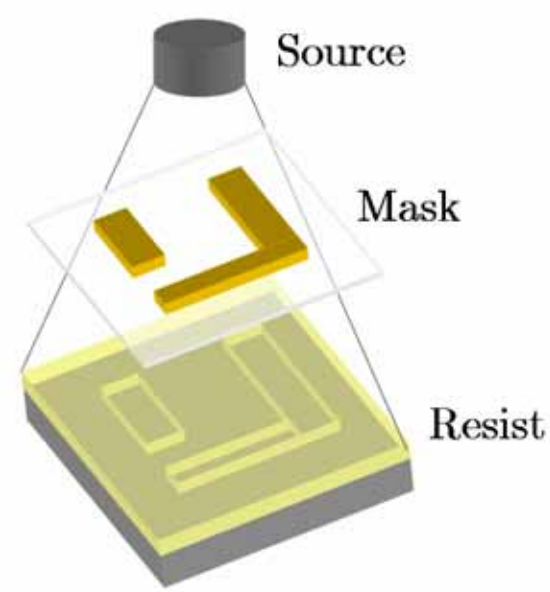


Figure 2.10: Schematic representation of the X-ray lithography.

For this reason XRL achieves better resolutions than optical lithography (which employs UV radiation); furthermore, high penetration power of X-rays allows to pattern thick resist layers. On the other hand a X-ray lithography system is much more complex (and expensive)than optical lithography. The lithographic step reported in Figure 2.10 can be modeled as an input-output system [23]: for a given input signal, i.e. the incident radiation, it returns a resist image as output. We can decompose the system in a series of cascaded subsystems: the mask, the gap and the resist. The output signal of each element constitutes the input signal of the next one. If we consider transversal planes normal to system's optical axis,

the propagating signal is the radiation's electric field ˜E(x,y,ω), which describes the image on the planes. As output of the three subsystems we have respectively:

- mask image: the intensity of the electric field, I(x, y), just after the mask;
- aerial image: the intensity I(x, y) after propagation in the gap and just before the resist;
- latent image: the absorbed energy distribution in the resist layer;

In the following we give more details regarding the three subsystems.

2.4.1. Mask

As mentioned above, the mask is composed of transparent and opaque areas for the X-radiation, according to the pattern to be transferred to the resist. The mask is usually composed by a thin Si_3N_4 membrane with gold structures (Figure 2.11Figure 2.11: Schematic drawing of a X-ray mask.); the membrane is the transparent base of the mask, whereas the gold structures constitute the opaque areas. Several thickness of the Si_3N_4 are used (usually 100 nm, 500 nm, 1 μm and 2 μm), according to the purposes and the lateral dimensions of the membrane (100 nm membranes with lateral size of several mm are too much fragile). The membrane is covered with a thin Cr-Au bilayer (10 nm of Cr and 20 nm of Au) by means of evaporation: the Cr-Au layer provides electric contact in order to grow subsequently the gold absorbent elements by electrolysis. For this purpose a different lithographic technique, e.g. the electron lithography, is exploited to produce a resist pattern on the membrane; then, the resist is used as a mask in a gold electrolytic growth and finally the resist is stripped away. As a result gold elements are present on the membrane.

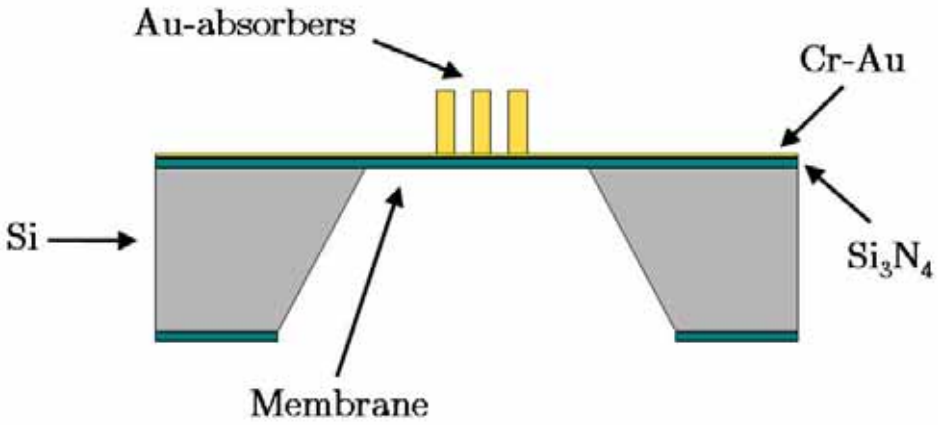


Figure 2.11: Schematic drawing of a X-ray mask.

Of course the shadow effect of the mask is never perfect because different mask regions do not provide neither total absorption nor total transparency. Then, neglecting the diffraction effects at this moment, the pattern profile is well defined if a good ratio of absorption to transparency is achieved. We can introduce a contrast parameter defined as:

$$C = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \qquad \text{(Eq. 2.8)}$$

I_{max} and I_{min} are respectively the maximum and minimum values of radiation intensity just after the mask, corresponding to the transparent and opaque areas.

The highest value is 1, achieved when I_{min} is null, that is for total absorption; instead the lowest value is zero, for $I_{max} = I_{min}$. Obviously higher is the contrast, better is the delineation of the image in the mask section; if the contrast is null, the image is completely wasted.

For a monochromatic radiation of intensity I_0 it follows that:

$$I_{\max} = I_0 e^{-\mu_m d_m} \qquad I_{\min} = I_0 e^{-\mu_m d_m - \mu d} \qquad \text{(Eq. 2.9)}$$

where μ_m and μ are respectively the absorption coefficient of the membrane and that of the gold structures; dm and d the corresponding thickness.

By substitution of the last formulas in Eq. 2.8 the contrast is:

$$C = \frac{1 - e^{-\mu d}}{1 + e^{-\mu d}} \qquad \text{(Eq. 2.8)}$$

which only depends upon the opaque structures parameters.

It is worth to remember that the absorption coefficient is wavelength dependent; therefore in the case of non-monochromatic radiation it is necessary to average the result over the frequency spectrum. In literature [23] a contrast value larger than 0.6 is reported to be enough for a satisfactory image formation; from Eq. 2.10 it follows that a thickness $d > (\ln 4)/\mu$ is needed for the absorbent areas.

In XRL soft X-rays are employed, that is 1-3 keV energy beams, and in this energy range the Au absorbent coefficient is some μm^{-1}: if $\mu = 5\mu m^{-1}$ then a thickness of 280 nm is required for the gold elements. Typically a 350 nm Au thickness is deposited in order to achieve a well shaped image and to compensate other possible losses of definition, e.g. in the resist processing or during the development step.

2.4.2. Gap and aerial image

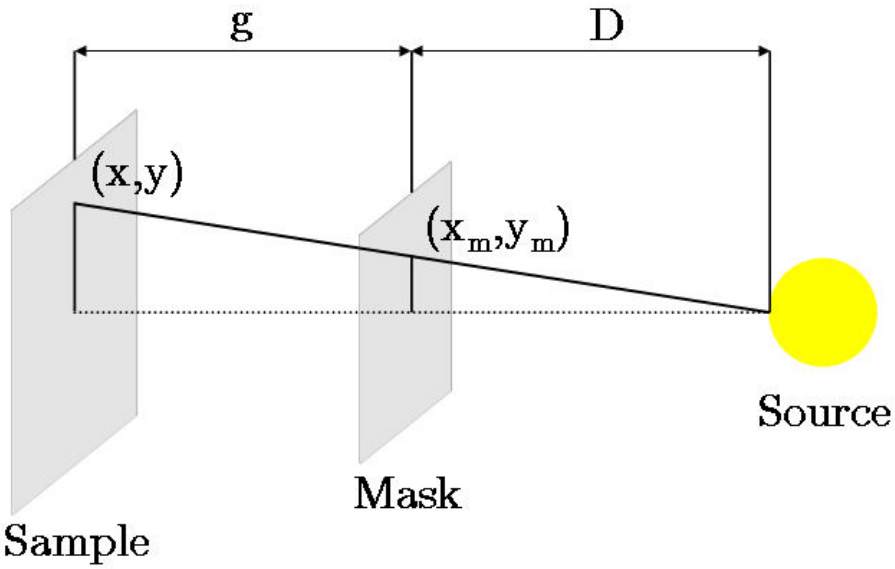


Figure 2.12: Magnification of the X-ray mask's pattern due to the source's divergence.

In turn to describe the aerial image formation starting from the mask image, both geometrical and diffractive arguments are needed. As already pointed out, the proximity lithography is just a shadow projection and the effects of the mask upon the aerial image are contained in its transmission function, $T(x_m, y_m)$, where (x_m, y_m) is a point on the mask. From a geometrical point of view the aerial image will be:

$$I(x,y) = T(x_m, y_m) \cdot I_0 e^{-x_0 g}$$
$$x = x_m(1 + g/D);\ y = y_m(1 + g/D) \qquad \text{(Eq. 2.11)}$$

where D is the distance between the radiation source and the mask, μ_0 is the absorbent coefficient of the material in the gap and g is the width of the gap itself. Referring to the Figure 2.12, the last two formulas for x and y are obtained by a simple geometrical similitude and they describe an exact replication of the mask pattern, except for the magnification due to the angular divergence of the source. If we have g = 20 μm and D = 20 cm, the magnification g/D is 100 ppm, that surely is acceptable. Anyway it is a uniform magnification that can be taken into account during the fabrication of the mask. More than the geometrical contributions, the diffraction effects play an important role in the aerial image formation.

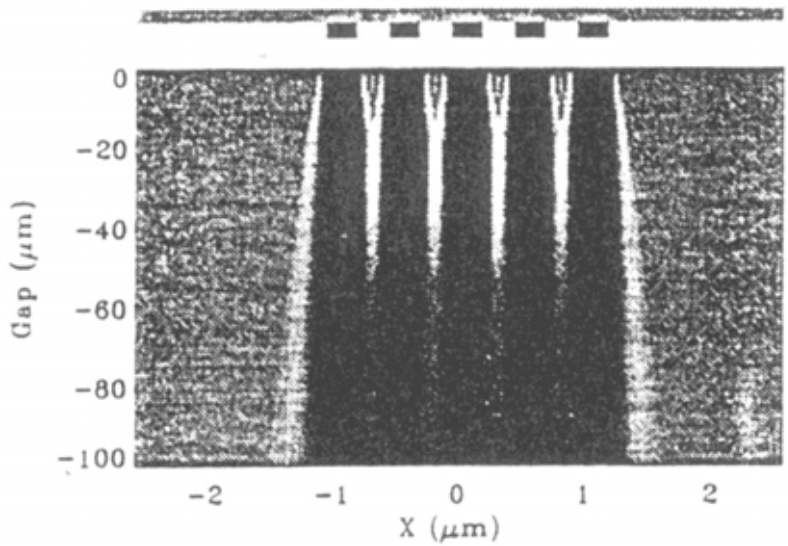


Figure 2.13: Diffractive effects in the aerial image after the X-ray mask.

Let us consider the following example: a slit of width a irradiated by a wavelength λ produces a spreading of the beam $\delta \approx g(\lambda/a)$, with g distance from the mask. Substituting some typical values of X-ray lithography, λ = 1 nm, g = 20 μm and a = 250 nm, we obtain $\delta \approx$ 80 nm, which is significant if compared to a. In Figure 2.13 the intensity of the radiation after the mask is shown: they are evident the diffraction's effects with increasing gap values. For the short calculation we just performed, we used the results of the Fraunhofer diffraction theory, but due to the short wavelength and to the small gap the real situation is better described by the Fresnel diffraction. For the single slit problem, the condition to distinguish between Fresnel diffraction (near-field) and Fraunhofer diffraction (far-field) is given by [24]:

$$\frac{ka^2}{g} > 1 \qquad \text{(Eq. 2.12)}$$

with $k = 2\pi/\lambda$; when the Eq. 2.12 is true, the Fresnel diffraction has to be considered. Substituting the values of the previous example we find that the Eq. 2.12 is true; furthermore in X-ray lithography the parameters (wavelength, gap and structure's resolution) always fulfill the Eq. 2.12. Unfortunately the theory of the near-field diffraction is much more complex than one of the Fraunhofer diffraction: the case of a wave incident on a semi-infinite opaque plane is reported in [24] and the radiation spreading in the shadow region is found to be $\delta \approx \sqrt{\lambda g}$. We can use this result to find the resolution limit of X-ray lithography. A standard

procedure consists in applying the usual Rayleigh criterion to the Fresnel diffraction:

$$\delta = k_1\sqrt{\lambda g} \qquad \text{(Eq. 2.13)}$$

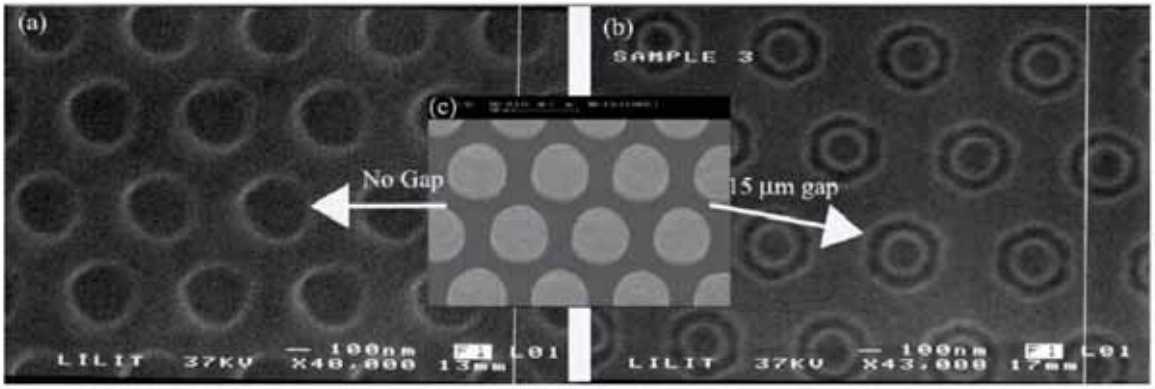


Figure 2.14: Results of XRL process exploiting the mask reported in the inset (c) with two different gaps: (a)gap g =0µm (null gap), (b)gap g =15µm.

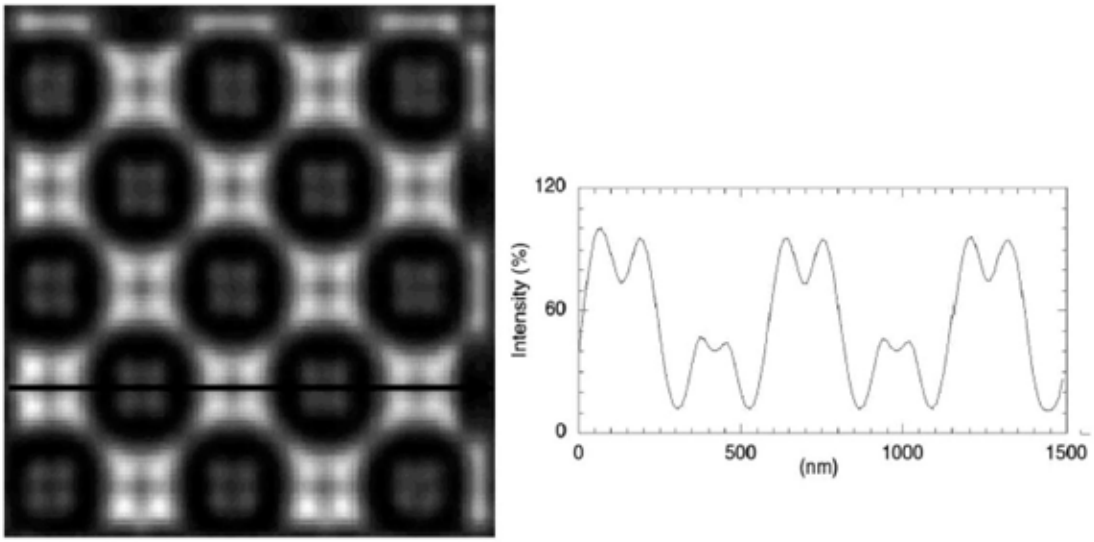


Figure 2.15: Aerial image computed by means of free-code TOOLSET for the mask reported in fig. 2.15; on the right, the intensity corresponding to the marked line is shown.

As starting point. pre-factor k_1 is imposed to be 1.22, as in the traditional Rayleigh criterion, but it is possible to achieve better resolutions than those expected with this value: in [25, 26] the obtained resolutions indicate that experimentally $k_1 \approx 0.6$ and also the good resolutions of [27, 28] achieved with

larger g values are in agreement with a k_1 pre-factor smaller than the traditional Rayleigh one. The point is that the final resolution of X-ray lithography can be improved with the process optimization and the diffraction can also be exploited to achieve higher resolutions than those of mask structures. As an example we report two patterns fabricated by means of XRL with a negative tone resist (Figure 2.14) [29]: the starting mask is the same for both patterns and it is an array of circular dots (lattice period 660 nm, dot diameter 330 nm). The first pattern is replicated with a null gap, $g = 0$ μm, whereas for the second pattern a large gap, $g = 15$μm, is used; in Figure 2.14 the diffraction effects are evident and the optimization of the process allows us to achieve structures with a higher resolution than the mask dots. The diffraction effects and the aerial image formation can be simulated (e.g. the TOOLSET code) and predicted; consequently it is possible to exploit the diffraction for the fabrication of patterns, taking into account the diffraction effects while fabricating the x-ray masks. Figure 2.15 shows the aerial image computed by TOOLSET for the mask of Figure 2.4 with a 15 μm gap.

2.4.3. Resist and latent image

When the X-photons penetrate the resist, as input signal we have the aerial image and as output the latent image is formed in the resist. Regarding the present dissertation, ionization of atoms is the main phenomenon occurring as X-rays travel through the resist layer. When the photon energy is transferred to an atom's electron, the electron may undergo transitions from ground state to an excited state of the discrete spectrum (excitation of the atom) or to the continuous spectrum (ionization of the atom). In the latter case photoelectrons are produced and the holes are soon recombined by fluorescence or Auger effect. The energy of photoelectrons is of several hundreds eV; of course photoelectrons behave as the electrons that penetrate the resist in electron lithography: they undergo scattering processes causing the breaking of polymeric chains of the resist, thus exposing the resist itself. While the absorption of the photons is well localized on an atomic site and the non-absorbed photons go straight without trajectory variations, the photoelectrons and the Auger electrons provide energy in the nearby regions, thus creating a blur like the proximity effects of the electron lithography. It is possible to calculate the shape of the exposed resist distribution taking into account this blur effect: the energy distribution of the aerial image is weighted with the absorption coefficient of the resist, in order to compute the absorbed energy distribution (dose image), and the convolution of the dose image with the blur function provide the dissipated energy distribution in the resist. Considering the

threshold energy density, or dose D_0, under which the resist is not exposed, the dissipated energy distribution allows to know which resist regions will be removed during the development. It is worth to point out that the blur due to the photoelectrons is not so relevant for the final resolution, in contrast to what happens for the proximity effect in electron lithography. In fact in X-ray lithography the radiation energy is not higher than 2-3 keV and consequently also the photoelectrons can not have energies exceeding 2-3 keV; on the contrary the primary electrons in EBL technique have usually some ten keV energies.

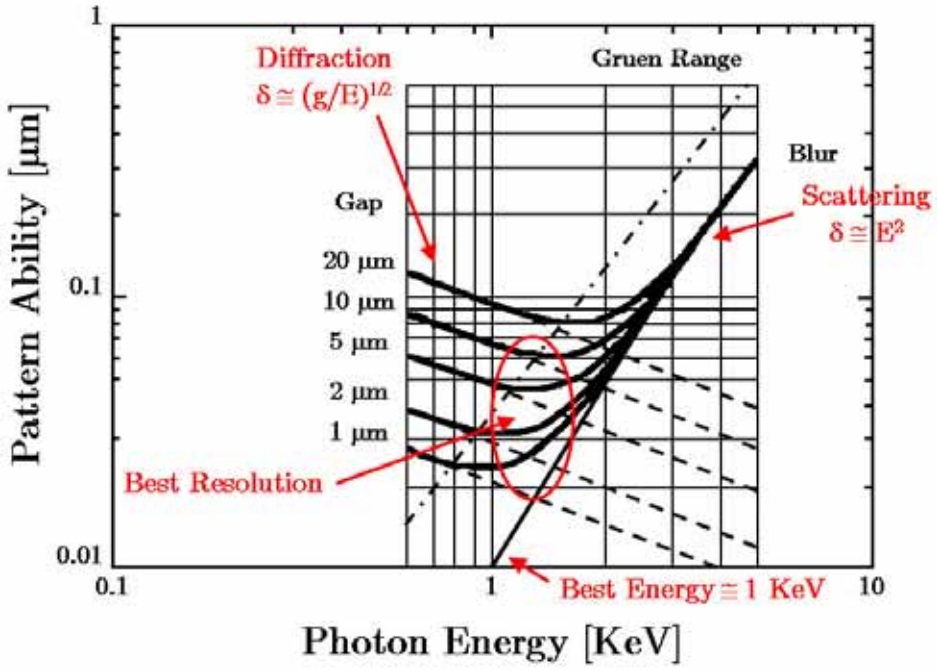


Figure 2.16: Achievable resolution for different gap values: at low energies the patterning ability is dominated by diffraction's effects; at high energies the electron blur (scattering contributions) is the dominant effect. The best resolution is achieved for 1-2 keV energy.

As a consequence, the ranges of the electrons in the two cases are much different, the XRL range being much smaller than the EBL one. Actually 1-2 keV energy is chosen in XRL because it is the best the combination of diffractive and blur effects. In the aerial image the achievable resolution is limited by the diffraction:

$$\delta = k_1\sqrt{\lambda g} = k_1(hc)^{1/2}\sqrt{\frac{9}{E}} \qquad \text{(Eq. 2.14)}$$

while the photoelectron blur in the latent image has the following spreading:

$$\delta = A\left(E_0^2 - I_0\right) \qquad \text{(Eq. 2.15)}$$

The combination of the last two formulas provides the best resolution, that is for photon energy in the range 1-2 keV (Figure 2.16).

2.4.4. LILIT beamline

LILIT beamline (Laboratory for Interdisciplinary LIThography) is located at the Elettra Synchrotron (Trieste, Italy) and it is mainly devoted to nanofabrication processes. Among the available techniques X-ray lithography is present. Up to now XRL requires a synchrotron because it is the only radiation source with brilliance and stability suitable for this purpose.

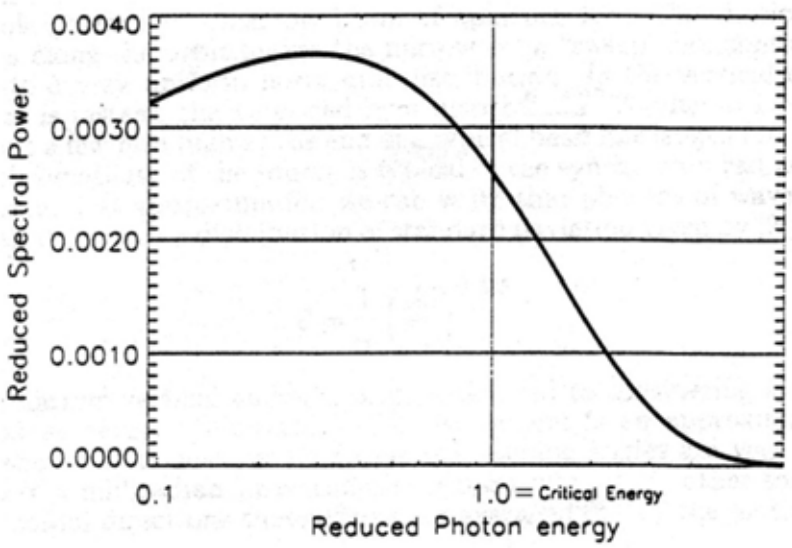


Figure 2.17: Universal curve for the spectral density energy of a synchrotron.

Synchrotron radiation is described by several universal formulas [23]:

$$\begin{aligned} \lambda_c\left(A^\circ\right) &= 5.59R/E^3 \text{ critical wavelength} \\ \varepsilon_c\left(keV\right) &= 2.218E^3/R = 12.398/\lambda_c \quad \text{critical energy} \\ P\left(kV\right) &= 88.47E^4 I/R = 39.89\varepsilon_c EI \quad \text{total power} \end{aligned} \qquad \text{(Eq. 2.16)}$$

where E is the energy, in GeV, of the electrons in the storage ring, I is the electron current in Ampere and R is the curvature radius, in meters, of the electron orbit. The critical energy ε_c is the value above which the irradiated intensity decreases,

as shown in Figure 2.17 which reports a universal curve for the spectral energy density of a synchrotron. Substituting in Eq. 2.16 Elettra parameters, a critical energy ε_c = 3.2 keV is found. As evident from Figure 2.17, the synchrotron radiation has a wide energy spectrum. The beamline selects the desired part of the spectrum and delivers the radiation to the X-ray stepper, the apparatus that allows to perform the lithographic exposures.

In Figure 2.18 a scheme of LILIT beamline is presented; we want to highlight the effects on the incoming radiation of the Beryllium windows and the X-ray mirrors: the former act as a high-pass filter, whereas the mirrors act as a low-pass filter. Combination of both these effects selects the working spectral window. It is worth to point out that a particular feature of LILIT beamline is the wide range of available spectra: adjusting the incident angles on the two mirrors, different spectra may be selected, with peak energy in the range 1-12 keV (Figure 2.19). Lowest energies are suitable for high resolution patterning (as mentioned above), whereas highest energies allow to perform deep lithography on thicker resist layers.

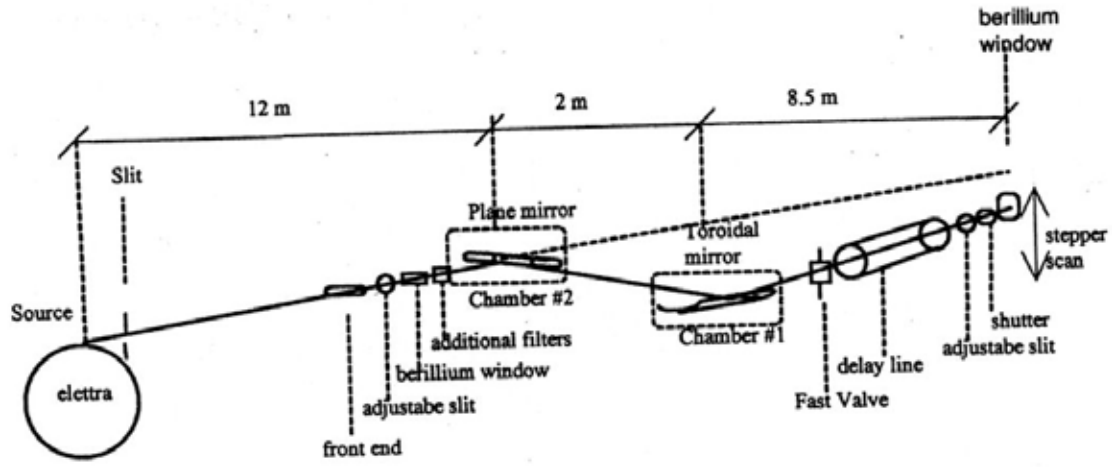


Figure 2.18: Layout of LILIT beamline.

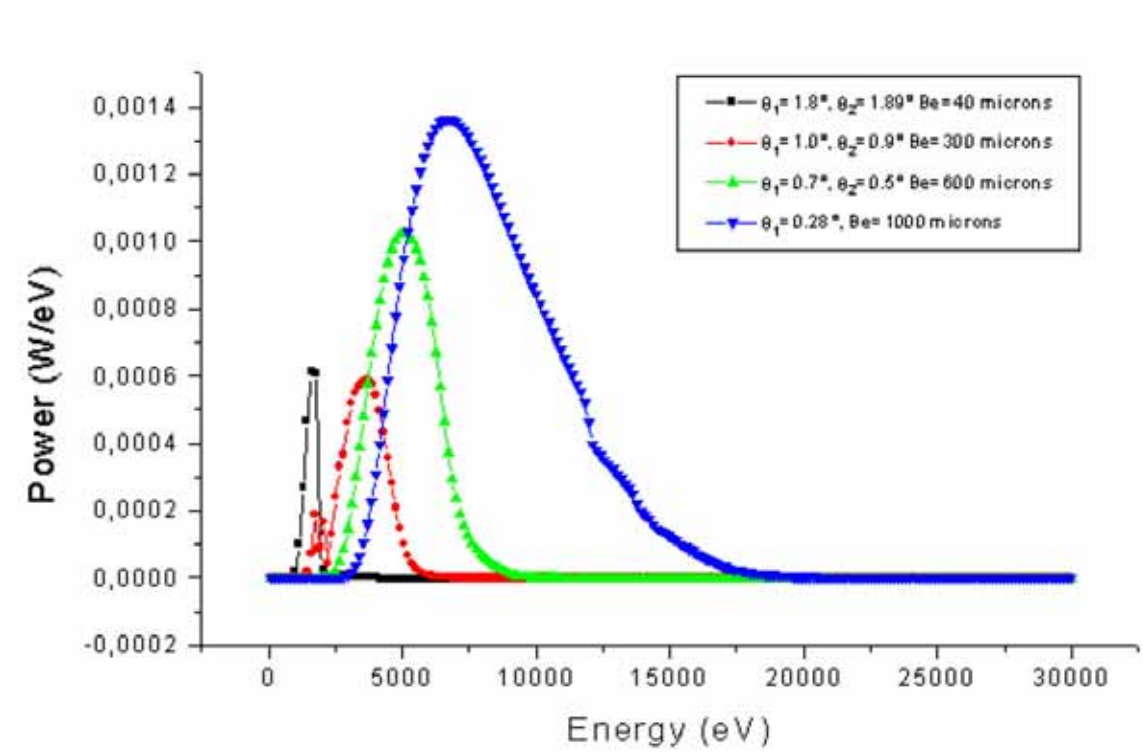


Figure 2.19: Energy spectra of LILIT beamline for different orientations of the mirrors and thickness of the Be windows.

2.5. Optical lithography

The operating principle of optical lithography is the same as in X-ray lithography. In Figure 2.20 a schematic system for optical lithography is shown. An ultra-violet (UV) radiation source (typically an UV lamp) lights up a sample through a mask and by shadow projection the mask-pattern is replicated on the photoresist. This technique (as in XRL) is called proximity lithography and obviously it is a parallel lithography (time-saving, suitable for large production, but a mask is required). Generally an optical lithography mask is made of glass with a Cr layer on one side; electron beam lithography (or another serial writing technique), combined with an etching process (usually wet-etching), is exploited to remove the Cr layer in some regions, according to the pattern; at the end the mask will present some opaque areas (with Cr) and other transparent ones (without Cr). As in XRL, the resolution available to optical lithography is limited by diffraction phenomena, and consequently by the used wavelength. In the biggest micro-electronic companies very short wavelengths (extreme UV) are exploited to reduce diffraction limit; combining 193 nm radiation wavelength with appropriate optical setups, 0.13 μm features are produced in microprocessor technology on a large scale production.

More often standard optical lithography can achieve lower resolutions, with feature's minimum dimensions in the μm scale. This is also the limit for the optical lithography system existing at LILIT laboratory, where it is available a Karl-Suss mask aligner with UV lamp working on the so called i-line of Hg (365 nm).

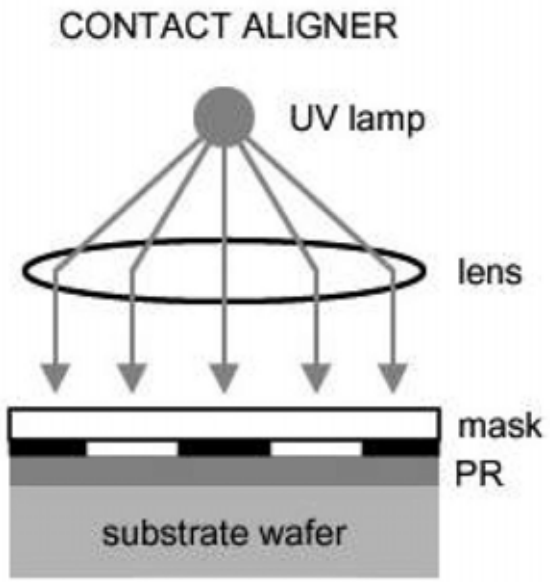


Figure 2.20: Scheme for the contact optical lithography.

2.5.1. Optical resolution

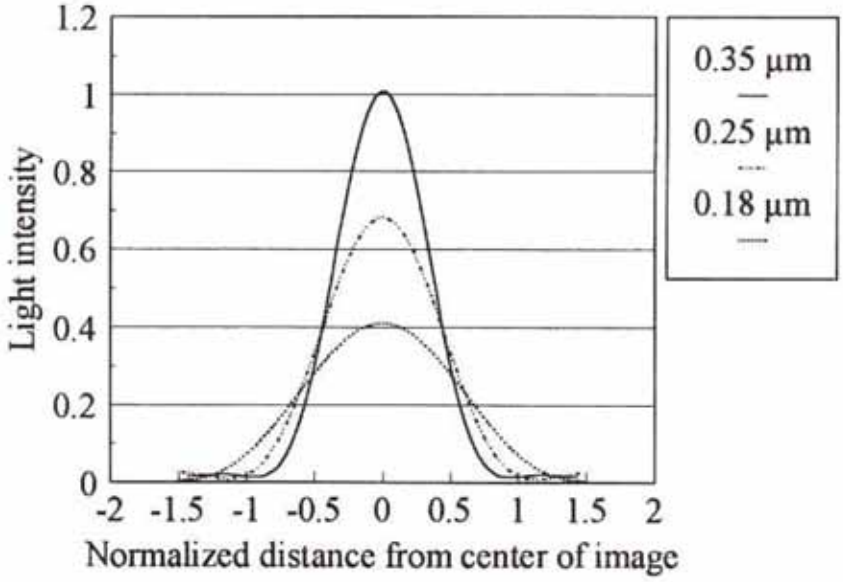


Figure 2.21: Optical resolution problem: optical profile after the mask is less clear when the mask's features decrease. Different curves are shown for different single slit's sizes.

For a lithographic system it is not trivial to give a strict definition of resolution, meaning the ability to resolve two distinct objects. Starting from the Rayleigh criterion, a convergent circular lens has the following resolving limit, *d*:

$$d = 0.61\frac{\lambda}{n \cdot \sin \vartheta} \qquad \text{(Eq. 2.16)}$$

where λ is the wavelength, n is the refraction index and 2θ is the angle seen by the lens; the term $n \sin \theta$ is called numerical aperture (*NA*). It is worth to remember that the Rayleigh criterion is derived considering point sources of radiation and surely this is not the case of optical lithography. The resolution problem for optical lithography is rather the one proposed in Figure 2.21: as mask structures' dimensions decrease, the optical profile after the mask is flatter and it is more difficult to identify structures' edges. Although the statement of the problem is different from that of the point sources, scale properties of the Rayleigh criterion are still applicable. If the numerical aperture is increased of 20%, a similar resolution improvement is expected for the lithographic performances. Generally the Rayleigh criterion is re-written as follows:

$$\delta = k_1 \frac{\lambda}{NA} \qquad \text{(Eq. 2.17)}$$

where δ is the resolution, NA the numerical aperture and the factor 0.61 is replaced by a k_1 pre-factor. This pre-factor is called k-factor of the process and it is found experimentally for a given setup (lamp, lenses, resist).

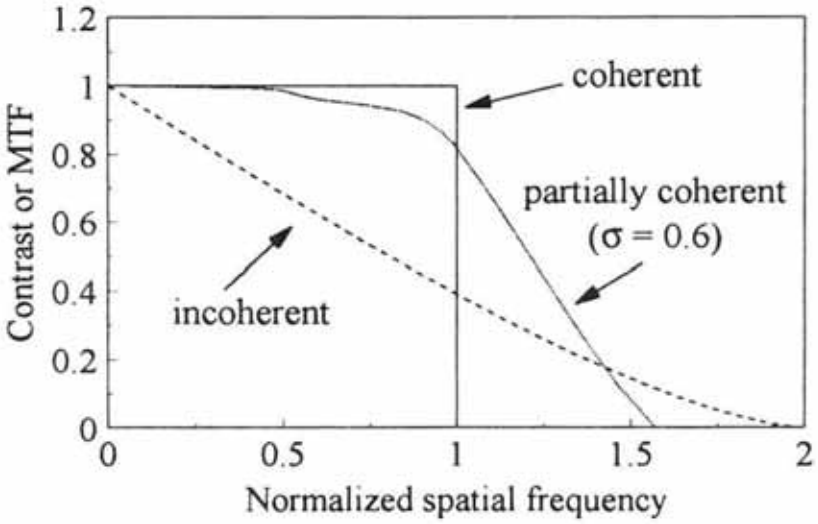


Figure 2.22: Contrast (or Modulation Transfer Function, MTF) as a function of the spatial frequency for radiations with different coherence.

Some considerations about radiation coherence are necessary. In the Rayleigh criterion we suppose to deal with incoherent light; but if we consider totally coherent radiation the minimum resolution is [30]:

$$d = 0.77 \frac{\lambda}{n \cdot \sin \vartheta} \qquad \text{(Eq. 2.18)}$$

The same result as for the incoherent case, except for the pre-factor value: for coherent light the resolution decreases. Incoherent light is advantageous also because it allows to perform lithography near resolution limit. For a given set of optical elements, periodical lattices in the mask pattern are well replicated by coherent light until the lattice dimensions are decreased down to a specific value, but beyond these dimensions they are not replicated at all. This dramatic cut-off in resolving power is due to the coherence of the radiation. In Figure 2.22 it is presented the modulation transfer function (MTF) or contrast of a given optical system as a function of the spatial frequency of the pattern (for example of a

lattice). It is clear that the coherent light causes an impressive contrast loss beyond resolution limit, while totally incoherent light has very poor results also before this limit; the best situation is obtained with partially coherent light, which provides an almost complete contrast before the resolution limit and slowly decreasing performances beyond it.

2.6. Etching process

The resist pattern produced by lithographic step is not the final device, but only a temporary structure. It is used as a mask in both an additive and a subtractive process; in the former it covers the areas where no deposited material is required, in the latter it saves from damage the sample's areas not to be removed. As mentioned above, the removal step is called etching; in this section we shortly present the widely employed etching processes. Etching processes can be divided in two main families: wet-etching and dry-etching. As suggested by the name, wet-etching techniques exploit chemical solutions to remove the material from the sample; dry-etching is usually a plasma-based process in which ionized atoms bombardment removes the material. For both wet and dry etching two important features are: selectivity towards the materials and the fidelity to the pattern. The selectivity is defined as the ratio of etch rates between different materials. We consider the situation of Figure 2.23, in which a substrate (e.g. silicon wafer) covered with a different material layer (e.g. metallic film) constitutes the sample and the resist pattern has to be transferred to the underlying film. Of course for the resist mask it is preferred an etch rate much smaller than the rate of the metallic film; furthermore also selectivity towards the silicon substrate would be helpful as endpoint of the process. Chemical reactions usually provide satisfactory selectivity, while pure ion bombardment etching has a really poor selectivity (that is similar etch rates for the mask and the film) and the success of the process relies on a resist layer thicker than the metallic film. Fidelity is obviously related to the differences between mask pattern and the transferred one. With reference to Figure 2.24 we can introduce the bias as a parameter to quantify fidelity. A zero-bias process produces a vertical edge profile coincident with the edge of the mask (as in Figure 2.23); in this case there is no etching in the lateral direction and the process is called totally anisotropic etching.

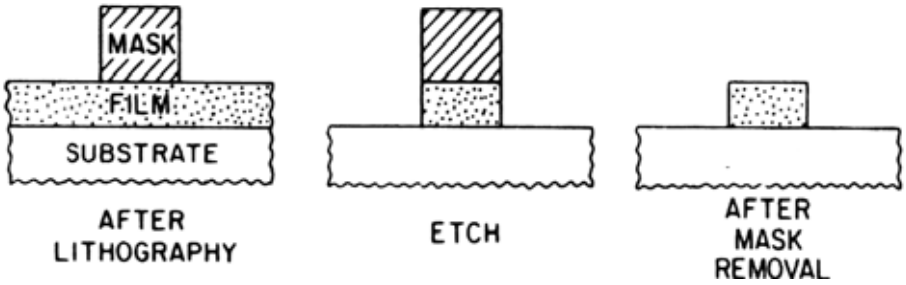


Figure 2.23: Selective etching process of a masked film (subtractive process).

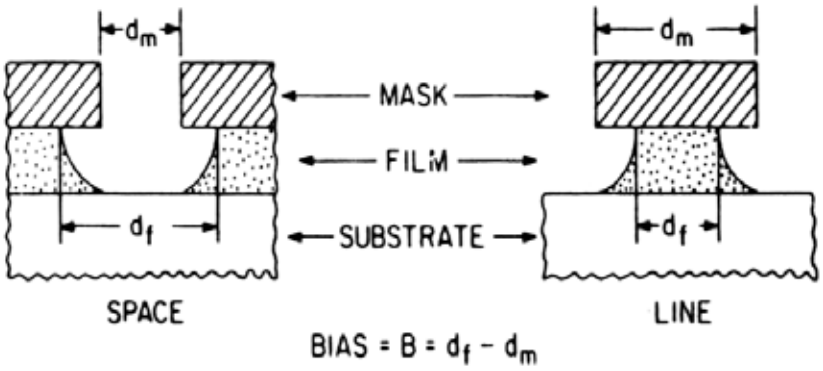


Figure 2.24: Definition of the bias parameter for the etching process.

When the vertical and the lateral etch rates are equal or, more precisely, when the etch rate is direction independent, the edge profile has a circular shape (as in Figure 2.24); in this case we have a pure isotropic etching and a large bias value. Intermediate situations can also occur and they are evidently the result of an anisotropic etching, but only the total anisotropic process (vertical sidewalls) is usually referred to as anisotropic.

2.6.1. Wet etching

Wet-etching processes are strictly based on chemical reactions and frequently have a high selectivity; on the other hand in most cases they are isotropic, as the liquid solution wets the lateral walls of the structures as well as the bottom of them. Only crystalline materials have a chance to be etched in an anisotropic way, different crystallographic planes being removed in some cases with different rates. Sometime wet-etching requires strongly basic or acid solutions or temperatures much higher than the room temperature; in these cases the resist may provide an inadequate mask action, due to the dissolution of the resist itself or to adhesions problems. In fact only Van Der Waals forces bind the resist and the sample and it

is not surprising that the chemical solution can damage the interface region, causing the separation of the resist layer. If the resist is not the right candidate for a suitable mask action, a different material can be deposited as etching mask by means of an additive process. The most part of dielectric materials, due to their amorphous or polycrystalline nature, are removed isotropically in wet-etching solutions. As example, the silicon dioxide (SiO_2) in all his forms (thermal oxide, fused silica or simply glass) presents an etch rate independent of direction in aqueous solution of HF. The etch rate will depend upon the density of the material, density of defects and the residual stresses (thermal and mechanical);for this reason SiO_2 layers deposited by different techniques will have different etch rates. Wet-etching solutions for many materials have been adjusted [31].

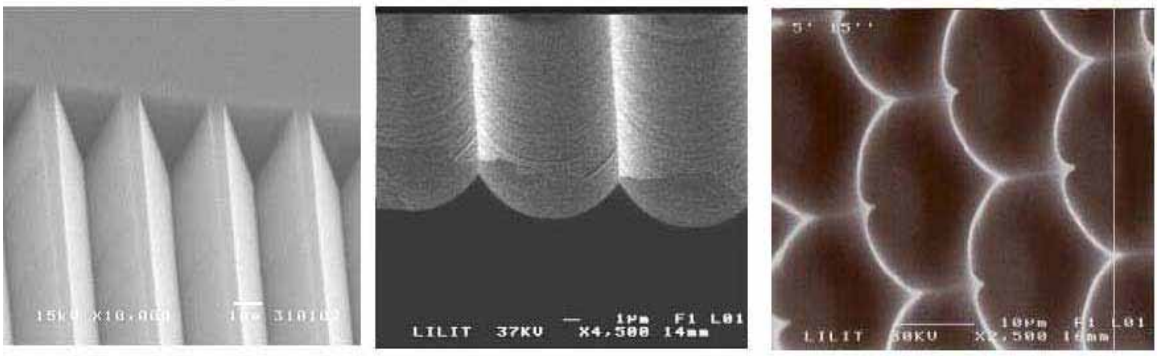


Figure 2.25: Wet-etching results: on the left, anisotropic etching of Si {100}; in the middle and on the right, isotropic etching of SiO_2 (the structures on the right after [15]).

Also if isotropic etching does not cause an exact transfer of the mask pattern, it proved to be very useful for the fabrication of curved profile structures [15], as those shown in Figure 2.25. If the material to be etched presents a crystalline structure, wet etching may provide anisotropic results. Generally the etched structures exhibit as surfaces the crystallographic planes with the smallest etch rate. In the case of crystalline silicon, the {111} plane is the most packed (highest surface density of atoms) and consequently it has the smallest etch rate. If a {100} surface silicon wafer is employed as substrate, the etched structures will present tilted walls, where the tilting angle is 54°45', that is the angle between the {100} and the {111} planes (Figure 2.25). The most used solution for the Si wet-etching is a *KOH* aqueous solution; the reaction mechanism is the following:

$$Si + 4KOH \rightarrow SiO_4^{4-} + 4K^+ + 2H^2 \qquad \text{(Eq. 2.19)}$$

The etch rate for the {111} planes is two order of magnitude smaller than that for the {100} and {110} planes. The etch rate for all the planes is strongly dependent upon solution concentration and temperature. In order to fabricate well shaped structures with flat surfaces by means of this technique it is very important to have the structure edges parallel to the {010} and {001} planes.

2.6.2. Dry etching

Dry etching is synonymous with plasma-assisted etching. Their earliest applications in silicon IC technology dates back to the late 1960s, but only later they were largely investigated. In the middle of 1970s silicon nitride turned to have many interesting features to be employed as inorganic passivation layer, but no wet etching was available; the problem was solved by means of dry etching. This event marks the beginning of large scale efforts to develop plasma etching techniques. It was soon clear that plasma etching allowed to achieve high fidelity in pattern transfer and that the loss of resolution encountered in wet etching was dramatically reduced. Plasma etching was no longer seen as a substitute of wet etching, but rather as a new technique having suitable capabilities to improve the pattern transfer process. Plasma assisted techniques rely on partially ionized gases composed of ions, electrons and neutral species produced by low pressure (~10^{-4} to 10^{+1} torr) electric discharges. The generic term "plasma assisted etching" includes ion milling, sputter etching, reactive ion etching and reactive ion beam etching. These techniques differ for discharge conditions, type of gas and apparatus, but all of them are based on the formation of plasma.

2.6.3. Plasma

A useful definition for the plasma state is given by Chen [32]: "a plasma is a quasi-neutral gas of charged and neutral particles which exhibits collective behavior". It contains the most important aspects of a plasma, namely the collective behavior and the quasi-neutral state. Indeed it is not correct to call just any ionized gas a plasma; ordinary air under normal conditions is also ionized (ion density $\approx 10^{-12}$, that is very low) but surely it is nota plasma. In common air the motion of the particles is determined only by mutual collisions, as they are mostly neutral and the gravity is negligible. In a plasma there is a larger density of charged particles and the Coulomb interactions are not negligible; as it is a long-range interaction, the movement of the particles is determined not only by local conditions, but also by the status of remote regions; consequently collective behaviors arise.

Furthermore Coulomb interactions will prevent charge fluctuations and deviations from the neutrality to persist for a long time, hence the term quasi-neutral in Chen's definition. The main charge carriers in plasmas are electrons. Due to their low mass, they can follow external electric fields or charge perturbations quicker than the heavier ions. The distance over which an external potential can perturb a plasma, or, vice versa, the distance that the plasma needs to screen out the perturbing potential, is known as self-shielding or Debye length, λD. Solving the relative Poisson's equation leads to:

$$\lambda_D = \sqrt{\frac{\varepsilon kT}{e^2 n_0}} \qquad \text{(Eq. 2.20)}$$

where ε is the dielectric constant, k and T respectively the Boltzmann's constant and the electrons temperature, e and n_0 the electron charge and the electrons density in the bulk of the plasma (far from the region of the external potential).
Another way to point out ability of plasma to react to external perturbations is a time-domain analysis of the problem. Consider an external force displacing the plasma electrons over some length and disappearing soon after this. The displacement causes an electric field to arise, pulling back the electrons to their original position. But due to their inertia, they will overshoot the position and a displacement in the opposite direction will occur. The result is an oscillation of the electrons density, whose frequency is:

$$\omega_e = \sqrt{\frac{e^2 n_0}{\varepsilon m_e}} \qquad \text{(Eq. 2.21)}$$

where m_e is the electron mass. The frequency ω_e is called electron plasma frequency and it is an indication of the time taken by the plasma to restore the charge equilibrium. Both Debye length and plasma frequency are exhibitions of collective behavior and according to them some conditions can be stated to discriminate which gases can truly be called plasmas. The first criterion regards the finite dimensions of the plasma: external potentials and local charge concentrations should be shielded out on a scale that is small compared to the dimensions of the plasma. It follows that:

$$\lambda_D << l \qquad \text{(Eq. 2.22)}$$

where is the smallest dimension of the system. The second criterion states that the main interactions in a plasma have an electromagnetic nature rather than a fluid-mechanical nature; in other words electrons should interact with each other more frequently than with the neutral species. This means:

$$\omega_e > \nu_e \qquad \text{(Eq. 2.23)}$$

where ν_e is the electron-neutral collision frequency. The third criterion has a statistical origin and states that the number of electrons in a Debye-length sphere should be large; indeed in the derivation of the Debye length λ_D a statistical density distribution for the electrons is exploited and this procedure is correct only for a large number of electrons:

$$N_D = \frac{4\pi}{3} n_0 \lambda_D^3 >> 1 \qquad \text{(Eq. 2.24)}$$

where N_D is the number of electrons inside a Debye sphere.

Another parameter to introduce is the fractional degree of ionization, α, defined as:

$$\alpha = \frac{n_0}{n_g} \qquad \text{(Eq. 2.25)}$$

where n_0 and n_g are respectively the densities of the charged and neutral particles. It is worth to point out that the above conditions for the existence of a plasma take into account only the charge particles density and not the ionization factor; this means that a large range for α is available (actually from 10^{-6} up to 10^{6}).

Finally, we want to highlight that in plasmas the electron and ion populations have quite different characteristics, due to their different masses. As an example, they have different and independent velocity distributions and no thermal equilibrium is achieved by means of elastic collisions. In fact elastic particles collisions result in a significant impulse transfer only if the particle masses are comparable, otherwise the lighter particle retains almost all its energy. The same result is still true for collisions of electrons with neutral gas atoms. On the other hand the ion population reaches thermal equilibrium with the neutral gas, since ions and neutral atoms masses are comparable. If an external field is applied:

$$F = qE = m\frac{dv}{dt} \qquad \text{(Eq. 2.26)}$$

the electrons reach a larger speed than ions, because of the different mass. As the speed is related to the temperature T via the energy equipartition theorem, the two population also have different temperature. Indeed it is usual that in plasmas the electrons have very high temperatures (several ten thousands Kelvin degrees), while ion temperature is the same of the neutral gas and it is much smaller (several hundreds Kelvin degrees). The high electron temperature should not be surprising, as the heat transfer to the external world is related to heat capacity which is very small for plasma electrons (due to both their density and mass). Only for plasmas with ionization factor $\alpha > 1\%$ the Coulomb interactions between electrons and ions outweigh the collisions of ions with neutral atoms; the ionization factor increases the thermal equilibrium is achieved and the overall plasma temperature rises. These systems are called hot-plasmas and examples are provided by the central part of the sun, the nuclear fusion reactors and the arcs used in welding. In many plasmas for technological applications (in the etching as well as in the deposition technology, like in sputtering systems and Plasma Enhanced Chemical Vapors Deposition, PECVD) the ionization factor is much smaller than 1% (in the $10^{-6} - 10^{-4}$ range)and we always deal with cold plasmas.

2.6.4. Reactive Ion Etching

One of the most widely used plasma etching systems is Reactive Ion Etching (RIE). As shown in Figure 2.26 a RIE apparatus is composed by a vacuum chamber, input lines for the gaseous species and a radio-frequency (*rf*) power supply connected to an electrode inside the vacuum chamber; all the other walls of the chamber are electrically grounded. The sample is placed on the powered electrode and reactive gaseous species (like CF_4, H_2, O_2, etc.) as well as inactive species (usually Ar) may be introduced in the chamber, according to the requests of the process. The working pressure is kept constant during the process by means of the vacuum system; it usually ranges from 10^{-3} to 10^{-1} Torr. The *rf* voltage is exploited to produce and to sustain a plasma inside the chamber; the ions species formed in the plasma are accelerated towards the sample, where physical and chemical removing take place. The physical action is due to the ion bombardment of the sample and material is removed by sputtering effect; the chemical removal is realized when chemical reactions of the sample with the reactive gases produce volatile compounds, expelled by the vacuum system.

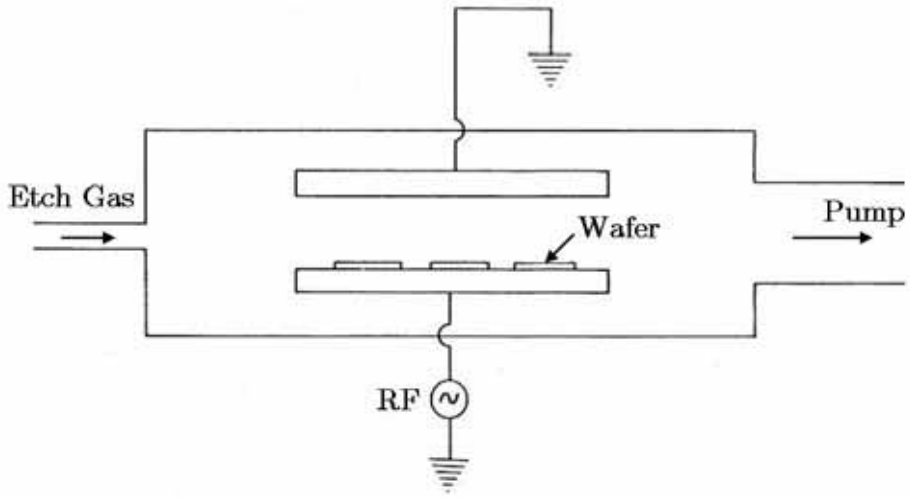


Figure 2.26: Scheme of a RIE (Reactive Ion Etching) apparatus: the samples are placed on the powered electrode.

In common RIE systems, a *rf* of 13.56 MHz is employed. While the plasma frequency (Eq. 2.21) for the electrons is larger than this value, the ion plasma frequency is much smaller, because of the ions mass. As a consequence, electrons are able to follow the rf signal while ions population is not perturbed; the result is an electron depleted region near the electrode called ion sheath. As in plasma the electrons are the main charge carriers, the ion sheath constitutes a non-conducting region and the *rf* coupling is capacitive, where area and thickness of the sheath determines the capacitance. By means of the area theorem, firstly proposed to explain the voltage division in low-pressure sputtering systems [33], we obtain:

$$\frac{V_{rf}}{V_g} = \left(\frac{A_g}{A_{rf}} \right)^4 \qquad \text{(Eq. 2.27)}$$

$$V_{dc} = V_{rf} - V_g$$

where V_{rf} and V_g are respectively the potentials developed at the powered electrode and the grounded one, A_{rf} and Ag are the electrode areas and V_{dc} is a dc self-bias voltage. It is evident that a powered electrode smaller than the grounded one causes large potential differences and the ions are accelerated by the self-bias voltage V_{dc} towards the sample. Typical values for the bias voltage range up to 350V. Besides the bias-voltage, the operating pressure is another important parameter during a RIE process. As known the mean free path of particles in a gas is proportional to the inverse of pressure. Hence smaller is the pressure, larger is

the mean free path. If we consider incident ions on a mask-pattern, accelerated in a direction normal to the sample surface, it is evident that the lateral etching is mainly due to the collisions with other particles, like in Figure 2.27. For this reason the anisotropy of etching process is strongly related to the mean free path: lower pressure corresponds to higher anisotropy.

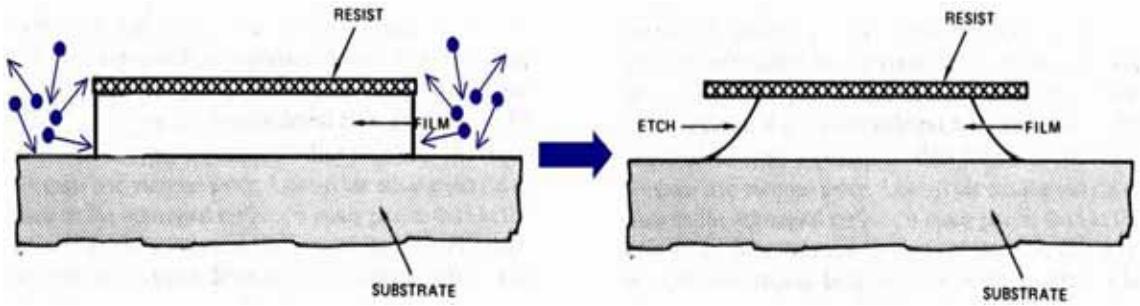


Figure 2.27: Effects of high pressure during a RIE process: the mean free path of ions is reduced and consequently lateral etching (undercutting)is produced.

2.7. Deposition techniques

Solid can be deposited onto a substrate from a liquid, a plasma, a gas, or a solid state. These additive techniques are often accompanied or followed by thermal processing to obtain desired materials properties and substrate adhesion. Although deposition methods, especially in the thin film arena, are generally the same as in integrated circuits (ICs), additive processes in miniaturization science span a much wider range from inorganic to organic materials. Besides the typical microelectronic elements (Si, Al, Au, Ti, W, Cu, Cr, O, N, and Ni-Fe alloys), miniaturizations science involves deposition of several atypical elements such as Zr, Ta, Ir, C, Pt, Pd, Ag, Zn, and Nb. Moreover, a plethora of exotic compounds ranging from enzymes to shape memory alloys and from hydrogels to piezoelectric are used. The number of materials and compound involved in MEMS fabrication, in comparison, is very limited. In miniaturization science, particularly in chemical sensors and biomedical devices, modular, thick film technologies are more prevalent.

In depositions methods two many categories can be distinguished: direct line-of-sight impingement deposition techniques, called physical vapor deposition (PVD), and diffusive-convective mass transfer techniques, i.e., chemical vapor deposition (CVD).

Only underlying physical and chemical principles of the various additive process are explained. A limited number of examples of deposition process in

micromachining are given in the end of the chapter. Electrochemical and electroless metal deposition and micromolding are gaining renewed interest because of the emerging importance of replication methodologies. Those methods are dealt with separately.

2.7.1. Physical vapor deposition

Many different kinds of thin films in ICs and micromachines are deposed by evaporation and sputtering, both of which are examples of PVD. PVD reactors may use solid, liquid, or vapor as raw material in a variety of source configurations. Other PVD techniques, very useful in the deposition of complex compound materials, are molecular beam epitaxy (MBE) and laser ablation deposition. Ion plating and cluster deposition are based on a combination of evaporation and plasma ionization and offer some of the advantages inherent to both techniques. The key distinguishing attribute to a PVD reactor is that the deposition of material onto the substrate is a line-of-site impingement type. At low pressure employed in a PVD reactor, the vaporized material encounters few intermolecular collisions while traveling to the substrate. Modeling of the deposition rate is consequently a relatively straightforward exercise in geometry.

2.7.1.1. Thermal evaporation

Thermal evaporation represents one of the oldest of thin film deposition techniques. Evaporation is based on the boiling off (or sublimating) of heated material onto a substrate in a vacuum. From thermodynamic considerations, the number of molecules leaving a unit area of evaporant per second or flux F is given by:

$$F = N_0 e^{-\frac{\Phi_e}{kT}} \qquad \text{(Eq. 2.28)}$$

where N_0 is a slowly varying function of temperature (T) and Φ_e is the activation energy (in eV) required to evaporate one molecule of the material. The activation energy for evaporation is related to the enthalpy of formation of the evaporant, H, as:

$$H = \Phi_e \times e \times N \text{ (J/mol)} \qquad \text{(Eq. 2.29)}$$

where N is the Avogadro's number.

Table 2.1 suggests the need for a good vacuum during evaporation; even at a pressure of 10^{-5} Torr, 4.4 contaminating monolayer per second will redeposit on the substrate [34]. For reference purposes, the number of atoms per unit area corresponding to a monolayer for a metal is about 10^5 atoms/cm^2. Moreover, to avoid reactions at the source (e.g., oxide impurities being formed), the oxygen partial pressure needs to be less than 10^{-8} Torr.

Pressure (Torr)	Mean free path (cm)	Number impingement rate (S^{-1}, cm^{-2})	Monolayer impingement rate (s^{-1})
10^{1}	0.5	3.8×10^{18}	4400
10^{-4}	51	3.8×10^{16}	44
10^{-5}	510	3.8×10^{15}	4.4
10^{-7}	5.1×10^{4}	3.8×10^{13}	4.4×10^{-2}
10^{-9}	5.1×10^{6}	3.8×10^{11}	4.4×10^{-4}

Table 2.1: Kinetic data for air as a function of pressure

In laboratory settings, a metal is usually evaporated by passing a high current through a highly refractory metal containment structures (e.g., a tungsten boat or filament). This method is called "resistive heating" (Figure 2.28a). Resistive evaporation is simple but easily spreads contaminants that are present in the filament, and the small size of filament limits the thickness of the deposit film. In industrial applications, resistive heating has been surpassed by electron-beam (e-beam) and RF induction evaporation. In the e-beam mode of evaporation, a high-intensity electron beam gun (3 to 20 keV) is focused on the target material that is placed in a recess in a water-cooled copper hearth. As shown in Figure 2.28b, the electron beam is magnetically direct onto the evaporant, which melt locally. In this manner, the metal forms its own crucible, and the contact with the hearth is too cool for chemical reactions, resulting is fewer source-contamination problems than in the case of resistive heating. E-beam evaporation not only results in higher quality films, it also provides a higher deposition rate (5 - 500 nm/min). Two disadvantages of e-beam evaporation are that the process may induce x-ray damage and possibly even some ion damage to the substrate (at voltage > 10kV, the incident electron beam will cause x-ray emission) and that the deposition equipment is more complex. X-ray damage may be avoided by using a focused high-power laser beam instead of an electron beam. However this technique has not yet

penetrated commercial applications. In RF induction heating, a water-cooled RF coupling coil surround a crucible with the material to be evaporated. Since about two-third of RF energy is absorbed within one skin depth of the surface, the frequency of RF supply must decrease as the size of the evaporant charge increases. With a charge of few grams of evaporant, frequencies of several hundred kilohertz are sufficient.

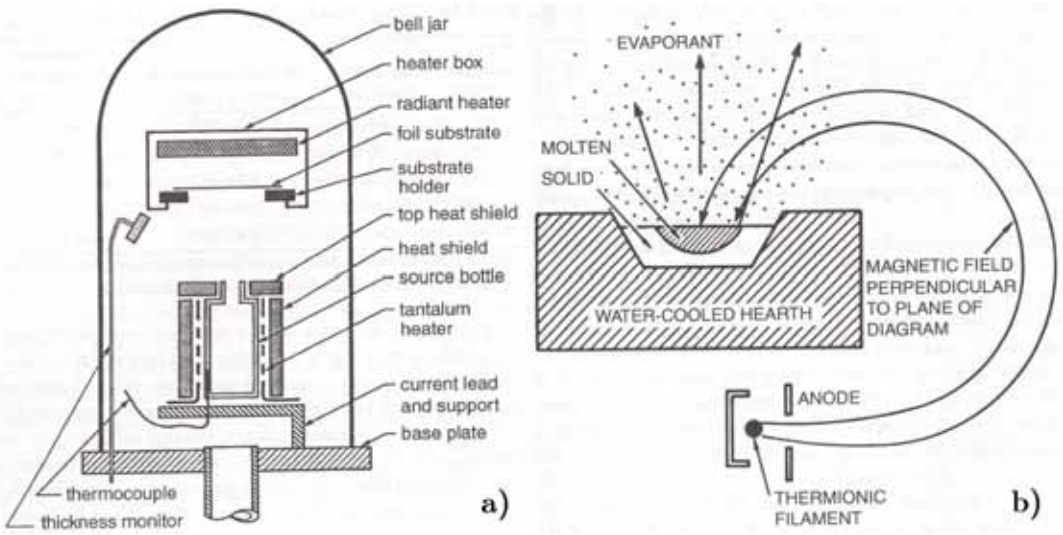


Figure 2.28: Thin film deposition: a) typical evaporation setup; b) Diagram of magnetized deflection electron-beam evaporation system

The substances used most frequently for thin-film formation by evaporation are elements or simple compounds whose vapor pressures range 1 to 10^{-2} Torr in the temperature interval from 600 to 1200°C. Refractory metals such as platinum, molybdenum, tantalum do not easily heat to the temperature required to reach the vapor pressure range. To obtain a deposition rate enough for practical applications, the vapor pressure at the source must be above the background pressure and reach at least 10^{-2} Torr. An evaporation source operating at a vapor pressure of 10^{-1} Torr, for example, can deliver rates about 1000 atomic layers per second.

Evaporation is fast (e.g., 0.5 μm/min for Al) and comparatively simple; it registers a low-energy impact on the substrate (~0.1eV); that is, no surface damage results except when using e-beam evaporation. Under proper experimental conditions, evaporation can provide films of extreme purity and known structure. In cases where the purity of the deposed film is of prime importance, evaporation is the preferred technique. For example, for electrochemical sensor, where the electrocatalytic activity of the top monolayer of the sensing electrode determines proper operation, material purity is a priority consideration. In chemical sensor,

the surface purity of films is more important than the bulk resistivity (a criterion often used to evaluate IC films). As we just learned, evaporators emit material from a point source (e.g., a small tungsten filament), resulting is shadowing and sometimes causing problems with the deposition, especially on high-aspect structures. Difficulties also arise for large areas where highly homogeneous film are required, unless special setups are chosen. The different problem to overcome arises from the source materials decomposing at high evaporation temperature. While this risk does not exist when evaporating pure elements, it becomes a problem in the evaporation of compounds and substance mixtures. The e-beam heating system have an advantage here, since only a small part of the metal source evaporates. The initial evaporant stream is richer in the high-vapor-pressure component; however, the melt depletes that constituent locally, and eventually an equilibrium rate is established. The best way to deposit complex metal alloy is then to e-beam evaporate metallic elements from different sources as modern quartz-crystal deposition rate monitors enable excellent control in alloy composition. To form oxides of the deposed metals, evaporation is performed in a low-pressure oxygen atmosphere. This process is known as reactive evaporation. The oxygen supply comes from a jet directed at the substrate during deposition. To obtain the correct stoichiometry, the deposition needs to take place on a heated substrate.

Evaporates thin film are usually under tensile stress, and the higher material's melting point the higher stress. Tungsten and nickel, for example, can have stress in excess of 500 MPa, which may lead to curling or peeling [35]. Raising the temperature of the substrate tends to lower the internal stresses in the thin films.

2.7.2. Chemical vapour deposition

During CVD, the constituents of a vapor phase, often diluted with an inert carried gas, react at a hot surface (typically higher than 300°C) to deposit a solid film. In CVD, the diffusive-convective transport to the substrate involves intermolecular collisions. Mass and heat transfer modeling of deposition rates is consequently much more complex than in PVD. In the reaction chamber, the reactants are adsorbed on the heated substrate surface, and the adatoms undergo migration and film-forming reactions. Gaseous by-products are desorbed and removed from the reaction chamber. The reactions forming a solid material do not always occur on or close to the heated substrate (heterogeneous reactions) but can also occur in the gas phase (homogeneous reactions). As homogeneous reaction led to gas phase cluster deposition and result in poor adhesion, low density, and high-defect films,

heterogeneous reactions are preferred. The slowest of any of the CVD step mentioned, the gas phase or surface process determines the rate deposition. The sample surface chemistry, its temperature, and thermodynamics determine the compounds deposited. The most favorable end product of physical and chemical interactions on the substrate surface is a stoichiometric-correct film. Several activation barriers need to be surmounted to arrive at this end product. Some energy source, such as thermal, photon, or ion bombardment, is required to achieve this.

The CVD method is very versatile and works at low or atmospheric pressure and at relatively low temperatures. Amorphous, polycrystalline, epitaxial, and uniaxially oriented polycrystalline layer can be deposited with high degree of purity, control, and economy. CVD is used extensively in the semiconductor industry and has played an important role in transistor miniaturization by introducing a very thin film deposition silicon. Most recently, CVD copper and low-dielectric insulator ($\varepsilon < 3$) also have become important CVD applications. CVD embodies the principal building technique in surface micromachining. In the case of a CVD reactor, the diffusive-convective transport to the substrate involves many intermolecular collisions. Accordingly, mass and heat transfer modeling of the deposition rates becomes more complex. When a molecule has reached the surface, the required reaction analysis the same, regardless of the deposition method. The molecular phenomena at the surface to be considered include sticking coefficient, surface adsorption, surface diffusion, surface reaction, desorption, and film or crystal growth

2.7.3. Electrodeposition

MEMS microfabrication depends critically on metal deposition to replicate master photoresist mold. This additive process might involve electrodeposition.
In a typical electrodeposition through a polymer mask in IC applications, metal conductor must be deposited on a dielectric substrate. The dielectric is usually made conductive first by sputtering a thin adhesion metal layer (Ti, Cr, etc.) and then a conductive seed layer (Au, Pt, Cu, Ni, NiFe, etc.). The thickness of the thin refractory metal adhesion layer may be as small as 50 to 100 Å, while the thickness of the conducting seed layer can range from 150 to 300 Å [36]. The key requirement for seed layer is that it is electrically continuous and offers low sheet resistance. After forming a pattern in a spin-coated polymer by UV exposure, e-beam, or X-ray radiation and developing away the exposed resist, contact is made to the seed layer and electrodeposition is carried out.

In IC industry, one trend to avoid wet chemistry, but both IC and micromachining needs are forcing reconsideration of electrochemical techniques as a viable solution [37]. In micromachining, for example, electrochemical deposition enables metal replication of high-aspect-ratio resist mold while maintaining highest fidelity. Specifically in the electroforming of micro devices with LIGA, a conductive substrate, carrying the resist structure, serve as cathode. The metal layer growing on the substrate fills the gaps in the resist configuration, thus forming a complementary metal structure.

The use of solvent-containing development agent ensures a substrate surface completely free of grease and ready for plating. The fabrication of metallic relief structures is a well known art in the electroforming industry. The technology is used, for example, to make fabrication tools for compact discs where structural details in the submicron range is transferred. Because of the extreme aspect ratio (several orders of magnitude larger than in the crafting of CD master), electroplating in LIGA poses new challenges.

2.7.3.1. Electroplating

Electroplating takes place in an electrolytic cell. The reactions involve current flow under an imposed bias. As an example, we consider the deposition of Ni from $NiCl_2$ in KCl solution with a graphite anode (not readily attacked by Cl_2) and an Au cathode (inert surface for Ni deposition). With the cathode sufficiently negative and the anode sufficiently positive with respect to the solution, Ni deposits on the cathode and Cl_2 evolves at the anode. The process differs from electroless Ni deposition in that the anodic and cathodic processes occur on separate electrodes and that the reduction is affected by the imposed bias rather than a chemical reductant. Important process parameters are pH, current density, temperature, agitation, and solution composition. The amount of hydrogen evolving and competing with Ni deposition depends on the pH, the temperature, and the current density. Since one of the most important causes of defects in metallic LIGA microstructures is the appearance of hydrogen bubbles, these three parameters need very precise control. Pollutants cause hydrogen bubbles to cling to the PMMA structures, resulting in pores in nickel deposit, so the bath must be kept clean, for example, by circulating through a membrane filter with 0.3μm pore openings [38]. Besides typical impurities such as airborne dust or dissolved anode material, main impurities are nickel hydroxide formed at increased pH value in the cathode vicinity and organic decomposition product from the wetting agents. The letter two can be avoided to some degree by monitoring and controlling the pH and by adsorption of the organic decomposition products on activated carbon.

2.7.3.2. Diffusion-limited reactions

Species in the electrolyte must be transported to and from the electrode before electrode reaction can occur. As species are being consumed or generated at the electrode surface via electrochemical reactions, the concentration of these species at the electrode surface will become smaller or larger, respectively, than in the bulk of the electrolyte. Suppose we are dealing with a species of concentration C^0 in the bulk of the electrolyte ($x = \infty$) being consumed at the electrode, the concentration gradient at the electrode is then given as:

$$\frac{dc}{dx} = \frac{C^0_{x=\infty} - C_{x=0}}{\delta} \qquad \text{(Eq. 2.30)}$$

where $C_{x=0}$ is the concentration of the species at the electrode surface, and δ represents the boundary layer thickness. From the thermodynamic relationship between the potential and concentration difference leads to an overpotential:

$$\eta_c = \frac{RT}{nF} \ln \frac{C_{x=0}}{C^0_\infty} \qquad \text{(Eq. 2.31)}$$

i.e., the expression for concentration polarization, ηc, where n is the number of electrons involved in the reaction. On the basis of Faraday's law, we can rewrite Eq 2.30 in terms of the current density as:

$$i = nFD_0 \frac{C^0_\infty - C_{x=0}}{\delta} \qquad \text{(Eq. 2.32)}$$

with D_0 (cm^2s^{-1}) representing the diffusion coefficient of the electro-active species and n the number of electrons transferred.

At a certain potential ηc, all of species arriving at the electrode are immediately consumed, and from that potential on, the concentration of the electro-active species at the surface falls to *0*. that is, $C_{x=0} = 0$, and we reach the limiting current density i_l:

$$i_I = nFD_0 \frac{C_\infty^0}{\delta} \qquad \text{(Eq. 2.33)}$$

This equation shows that the limiting current is proportional to the bulk concentration of the reacting species. On the basis, classical amperometric (current-based) sensors are used as analytical devices. We can equate $C_{x=0}/C_\infty^0$ in Eq. 2.31 with 1- i/i_I or:

$$i = i_I\left(1 - e^{\frac{nF\eta_c}{RT}}\right) \qquad \text{(Eq. 2.34)}$$

This expression is illustrated for a cathodically diffusion-limited reaction in Figure 2.29. The cathodic current quickly becomes diffusion limited, whereas the activation-controlled current keeps on rising.

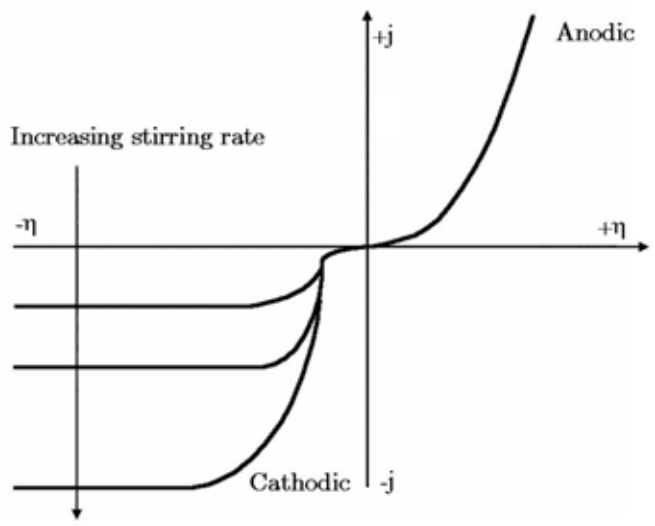


Figure 2.29: The cathodic limiting current is indicated for different stirring rates. The cathodic limiting current appears as a horizontal straight line limiting the current that can be achieved at any large negative value of the overpotential.

2.7.4. Spin coating

Spin coating technology has been optimized for deposition of thin layers of photoresist, about 1 to 2 µm thick, on round and nearly ideally flat Si wafers. Resists are applied by dropping resist solution, a polymer, a sensitizer (for two-component resist), and a solvent on the wafer. The wafer is then rotated on a

spinning wheel at high speed so that centrifugal forces push the excess solution over the edge of the wafer, and a residue on the wafer remains due to surface tension. In this way, films down to 0.1 μm can be made. However, biosensor substrate rarely are round or flat, and many chemical sensor membranes require a thickness considerably greater than 1μm for proper functioning. For example, a typical ion selective electrode (ISE) membrane is 50 μm thick. Consequently, spin coating technology does not necessary fit in with thick chemical membranes on a variety of substrates, and for monolayer deposition the method is not adequate, either.

3 MEMS micromirror scanning probe

3.1. Introduction

MEMS are small integrated devices or systems that combine electrical and mechanical components. They range in size from the sub micrometer (or sub micron) level to the millimeter level, and there can be any number, from a few to millions, in a particular system. MEMS extend the fabrication techniques developed for the integrated circuit industry to add mechanical elements such as beams, gears, diaphragms, and springs to devices.

Examples of MEMS device applications include pressure sensors [39], accelerometers [40], inertial sensors [41], microengines [42, 43]. New applications are emerging as the existing technology is applied to the miniaturization and integration of conventional devices.

These systems can sense, control, and activate mechanical processes on the micro scale, and function individually or in arrays to generate effects on the macro scale. The microfabrication technology enables fabrication of large arrays of devices, which individually perform simple tasks, but in combination can accomplish complicated functions.

MEMS are not about any one application or device, nor are they defined by a single fabrication process or limited to a few materials. They are a fabrication approach that conveys the advantages of miniaturization, multiple components, and microelectronics to the design and construction of integrated electromechanical

systems. MEMS are not only about miniaturization of mechanical systems; they are also a new paradigm to design mechanical devices and systems.

Around 1982, the term micromachining came into use to designate the fabrication of micromechanical parts (such as pressure-sensor diaphragms or accelerometer suspension beams) for Si microsensors. The micromechanical parts were fabricated by selectively etching areas of the Si substrate away in order to leave behind the desired geometries. Isotropic etching of Si was developed in the early 1960s for transistor fabrication. Anisotropic etching of Si then came about in 1967. Various etch-stop techniques were subsequently developed to provide further process flexibility.

These techniques also form the basis of the bulk micromachining processing techniques. Bulk micromachining designates the point at which the bulk of the Si substrate is etched away to leave behind the desired micromechanical elements [39]. Bulk micromachining has remained a powerful technique for the fabrication of micromechanical elements. However, the need for flexibility in device design and performance improvement has motivated the development of new concepts and techniques for micromachining.

Among these is the sacrificial layer technique, first demonstrated in 1965 by Nathanson and Wickstrom [44], in which a layer of material is deposited between structural layers for mechanical separation and isolation. This layer is removed during the release etch to free the structural layers and to allow mechanical devices to move relative to the substrate. A layer is releasable when a sacrificial layer separates it from the substrate. The application of the sacrificial layer technique to micromachining in 1985 gave rise to surface micromachining, in which the Si substrate is primarily used as a mechanical support upon which the micromechanical elements are fabricated. Surface micromachining enables the fabrication of complex multicomponent integrated micromechanical structures that would not be possible with traditional bulk micromachining. This technique encases specific structural parts of a device in layers of a sacrificial material during the fabrication process.

Prior to 1987, these micromechanical structures were limited in motion. During 1987-1988, a turning point was reached in micromachining when, for the first time, techniques for integrated fabrication of mechanisms (i.e. rigid bodies connected by joints for transmitting, controlling, or constraining relative movement) on Si were demonstrated. During a series of three separate workshops on microdynamics held in 1987, the term MEMS was coined. Equivalent terms for MEMS are microsystems (preferred in Europe) and micromachines (preferred in Japan).

Another fabrication technology for the creation of MEMS devices is mold micromachining. In this technology, mechanical features are created by first creating a mold. This mold is in the shape of the desired mechanical structure. The mold is then filled with a structural material to create the desired mechanical devices. The mold can then be etched away, or the mechanical device can be removed from the mold. The most famous technology for mold micromachining is the LIGA technology [45]. LIGA creates very small and very precise mechanical structures using the synchrotron radiation source.

Each of the three basic microsystems technology processes we have seen, bulk micromachining, sacrificial surface micromachining, and micromolding/LIGA, employs a different set of capital and intellectual resources. MEMS manufacturing firms must choose which specific microsystems manufacturing techniques to invest in [46].

MEMS technology has the potential to change our daily lives as much as the computer has. However, the material needs of the MEMS field are at a preliminary stage. A thorough understanding of the properties of existing MEMS materials is just as important as the development of new MEMS materials.

Future MEMS applications will be driven by processes enabling greater functionality through higher levels of electronic-mechanical integration and greater numbers of mechanical components working alone or together to enable a complex action. Future MEMS products will demand higher levels of electrical-mechanical integration and more intimate interaction with the physical world. The high up-front investment costs for large-volume commercialization of MEMS will likely limit the initial involvement to larger companies in the IC industry. Advancing from their success as sensors, MEMS products will be embedded in larger non-MEMS systems, such as printers, automobiles, and biomedical diagnostic equipment, and will enable new and improved systems.

3.2. Actuation techniques

There are many different types of MEMS actuator and sensors. The kingdom of MEMS actuators mainly consists of four families: electrostatic, piezoelectric, thermal and magnetic.

An actuator is a device that responds to the electrical signals within the transduction system. Specifically, a mechanical actuator is one that translates a signal from the electrical domain into the mechanical domain. In the ideal case, we

would like the conversion to be 100% efficient. Of course, any real system cannot achieve a figure anywhere near this, owing to internal and external losses. Typical micromechanical actuators offer an efficiency between 5% and 35%. Other factors such as ease of fabrication, robustness, resistance to external effects (i.e., temperature, humidity), and range of motion, result in a series of trade-offs for selecting the appropriate mechanism.

3.2.1. Electrostatic

Electrostatic actuators are based on the fundamental principle that two plates of opposite charge will attract each other. They are quite extensive as they are relatively straightforward to fabricate. They do, however, have a nonlinear force-to-voltage relationship.

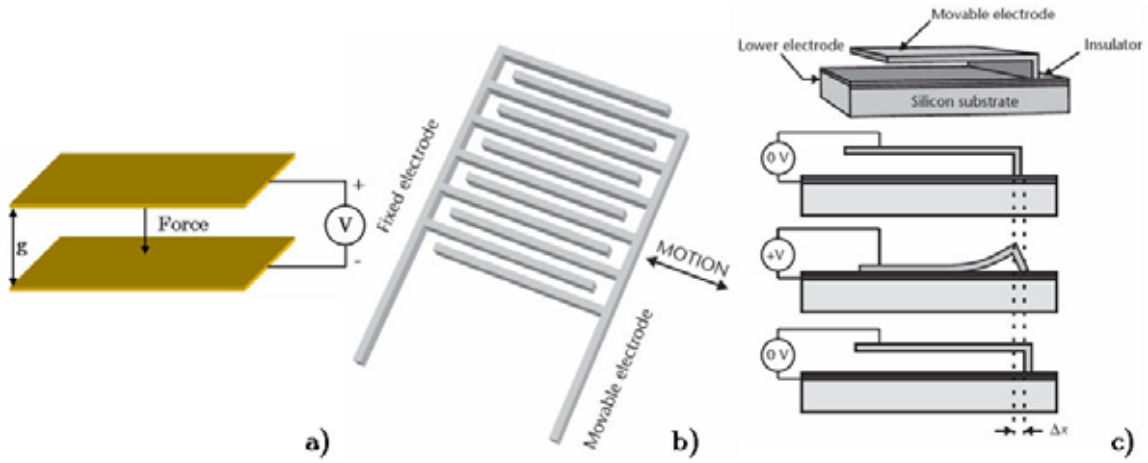


Figure 3.1: a) a simple planar capacitor electrostatic actuator; b) illustration of the electrostatic comb-drive actuator; c) illustration of the principle of operation of the electrostatic scratch drive actuator as described by Akiyama and Katsufusa.

The force is a nonlinear function of both the applied voltage (V) and the gap separation (g) (Figure 3.1a). Use of closed loop control techniques can linearize the response.

An alternative type of electrostatic actuator is the so-called comb-drive, which is comprised of many interdigitated electrodes (fingers) that are actuated by applying a voltage between them. The geometry is such that the thickness of the fingers is small in comparison to their lengths and widths.

The attractive forces are therefore mainly due to the fringing fields rather than the parallel plate fields, as seen in the simple structure above. The movement

generated is in the lateral direction, as shown in Figure 3.1b, and because the capacitance is varied by changing the area of overlap and the gap remains fixed, the displacement varies as the square of the voltage.

The fixed electrode is rigidly supported to the substrate, and the movable electrode must be held in place by anchoring at a suitable point away from the active fingers. Additional parasitic capacitances such as those between the fingers and the substrate and the asymmetry of the fringing fields can lead to out-of-plane forces, which can be minimized with more sophisticated designs.

Electrostatic actuation techniques have also been used to developed rotary motor structures. With these devices, a central rotor having surrounding capacitive plates is made to rotate by the application of voltages of the correct phase to induce rotation. Such devices have been shown to have a limited lifetime and require lubrication to prevent the rotor from seizing. The practical use has therefore been limited, but they are, nevertheless, the subject of intensive research.

Another interesting type of electrostatic actuator is the so-called scratch drive actuator (SDA) as described by Akiyama and Katsufusa [47]. The device comprises a flexible, electrode plate and a small bushing at one end. It is depicted in Figure 3.1c, which also illustrates the principle of operation. The free end of the electrode in the actual device is usually supported by a thin beam, but this is not shown in the figure. When a voltage is applied between the electrode plate and the buried electrode layer on the substrate, the plate buckles down and so causes the bushing to “scratch” along the insulator, thereby resulting in a small forward movement. When the voltage is removed, the plate returns to its original shape, thereby resulting in a net movement of the plate. The cycle can be repeated for stepwise linear motion.

3.2.2. Piezoelectric

As we have already seen, piezoelectric devices can be used for both sensor and actuator applications. An applied voltage across the electrodes of a piezoelectric material will result in a deformation that is proportional to the magnitude of the voltage (strictly electric field). The displacement across a bulk sample of PZT with an actuation voltage of several hundred volts, for example, is only a small fraction of a micron. When such a system is scaled down to that of a typical MEMS actuator, a displacement of several orders of magnitude less is obtained. For this reason, some form of mechanical amplification is needed in order to generate useful displacements. Such a device can be fabricated by depositing a piezoelectric film onto a substrate in the form of a cantilever beam as shown in Figure 3.2.

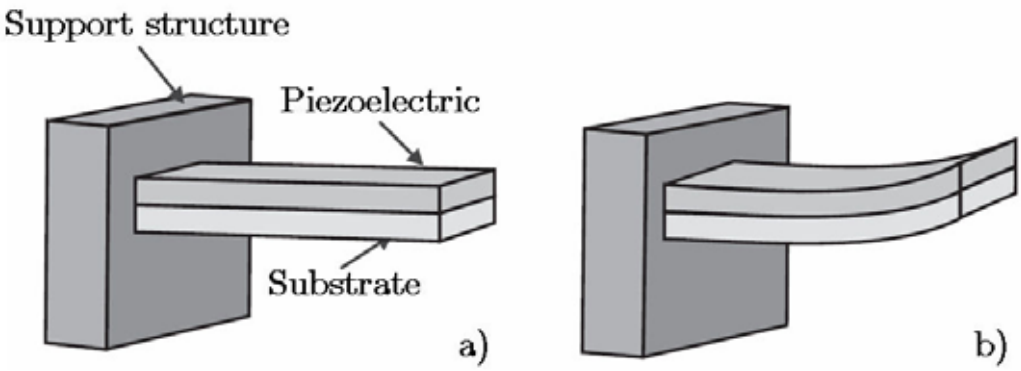


Figure 3.2: an example of a simple cantilever beam with a deposited piezoelectric layer: a) the structure with no applied voltage; and b) how the tip of the beam moves upon the application of an applied voltage.

This type of structure is referred to as a piezoelectric unimorph. The deflection at the free end of the beam is greater than that produced in the film itself. Piezoelectric actuators are often used in micropumps as a way of deflecting a thin membrane, which in turn alters the volume within a chamber below. Such a structure is depicted in Figure 3.3. The device comprises two silicon wafers bonded together. The lower wafer comprises an inlet and outlet port, which have been fabricated using bulk micromachining techniques. The upper wafer has been etched to form the pump chamber. The shape of the ports gives rise to a preferential direction for the fluid flow, although there is a degree of flow in the reverse direction during pumping. So the ports behave in a similar manner to valves. An alternative structure comprises cantilever-type flaps across the ports, but these often suffer from stiction during pumping. When a voltage is applied to the piezoelectric material, this results in a deformation of the thin membrane and hence changes the volume within the chamber. This is depicted in Figure 3.3b. Typical flow rates are in the range of nanoliters to microliters per minute, depending on the dimension of the micropump.

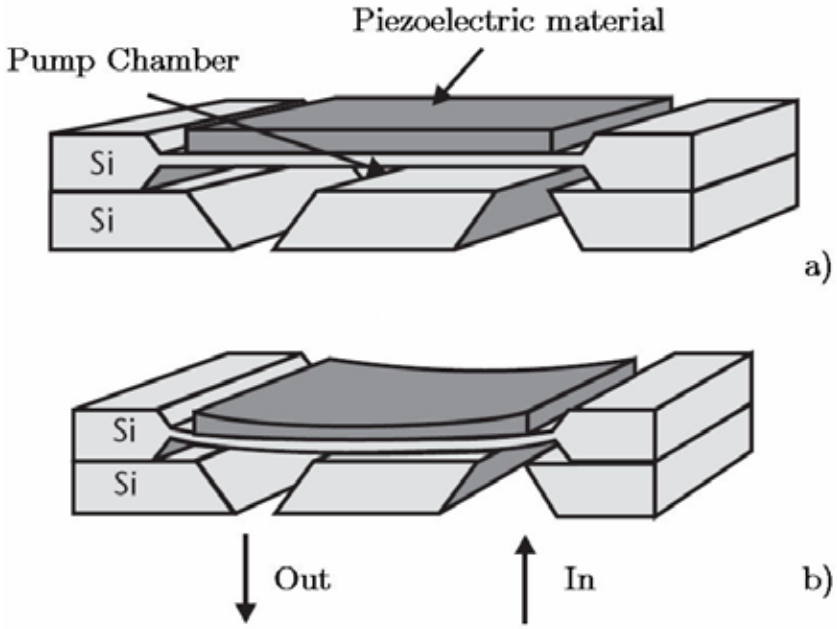


Figure 3.3: A simple micropump with a piezoelectric actuator: a) the situation with no applied voltage; and b) the effect of applying a voltage to the piezoelectric layer.

3.2.3. Thermal

Thermal actuation techniques tend to consume more power than electrostatic or piezoelectric methods, but the forces generated are also greater. One of the basic approaches is to exploit the difference in linear expansion coefficients of two materials bonded together. Such structures are often referred to as thermal bimorphs and are analogous to the familiar bimetallic strips often used in thermostats. One layer expands by a different amount to the other, resulting in thermal stresses at the interface leading to bending of the structure. The amount of bending depends on the difference
in thermal coefficients of expansion and also on the temperature. An illustration of a thermal bimorph is shown in Figure 3.4.

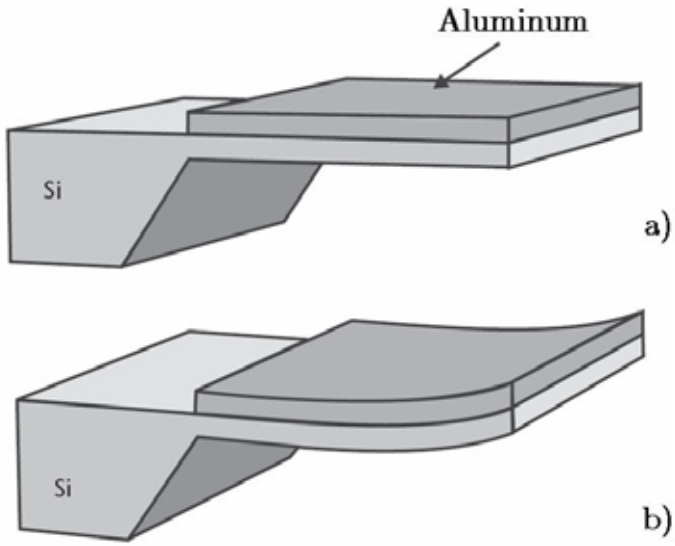


Figure 3.4: A simple thermal bimorph actuator a) before and b) after the application of electric current.

If an electric current is passed through the aluminum layer, it heats up (Joule heating), thereby causing the free end of the beam to move. These devices are relatively straightforward to fabricate and in addition to consuming relatively large amounts of power, they also have a low bandwidth because of the thermal time constant of the overall structure (i.e., beam and support).

An example of a commercial device based on thermal actuation is the so-called fluistor from Redwood Microsystems in California. This device is comprised of a cavity with a sealed fluid that can be heated and thus expanded. The heat is applied to the fluid via a thin-film resistive element. If one section of the cavity, such as a wall, is made more compliant than the other sections, then it will deform under pressure, thereby generating a mechanical force. The cavity is formed by bulk micromachining in silicon and is sealed using a Pyrex wafer, containing the heating element, anodically bonded to the silicon. Strictly, this is a thermopneumatic actuator and the commercial device is often used as a microvalve in applications such as medical instrumentation, gas mixers, and process control equipment. Such actuators may require up to 2W of power to operate.

Another thermal effect that can be exploited in thermal actuators is the shapememory effect, which is a property of a special class of metal alloys known as shape-memory alloys. When these materials are heated beyond a critical transition temperature, they return to a predetermined shape. The SMA material has a temperature-dependent crystal structure such that, at temperatures below the transition point, it possesses a low yield strength crystallography referred to as a

Martensite. In this state, the alloy is relatively soft and easy to deform into different shapes.

It will retain this shape until the temperature exceeds the phase transition temperature, at which point the material reverts to its parent structure known as Austenite. One of the most widely used SMA materials is an alloy of nickel and titanium called Nitinol. This has excellent electrical and mechanical properties and a long fatigue life. In its bulk form, it is capable of producing up to 5% strain. The transition temperature of Nitinol can be tailored between –100°C and +100°C by controlling the impurity concentration. The material has been used in MEMS by sputter depositing TiNi thin-film layers [48].

3.2.4. Magnetic

If a current-carrying element is placed within a magnetic field, an electromagnetic force (Lorentz force) will occur in a direction perpendicular to the current and magnetic field. The magnitude of the force is proportional to the current, length of the element, and the magnetic field. The availability of permanent magnetic materials, which are compatible with MEMS processing, is very limited, and thus it is common for the magnetic field to be generated externally. Discrete magnetic actuators often comprise coils, but such structures are not currently achievable with conventional MEMS processing and planar coils must be used.

Another approach that can be used as the basis of a magnetic actuator is the magnetostrictive effect. Magnetostriction is defined as the dimensional change of a magnetic material caused by a change in its magnetic state. Like the piezoelectric effect, it is reversible, and an applied stress results in a change of magnetic state. All magnetic materials exhibit varying degrees of magnetostriction. J. P. Joule discovered the effect in 1847 by observing the change in length of an iron bar when it was magnetized. A popular modern-day magnetostrictive material is Terfenol-D, an alloy of terbium, dysprosium, and iron. The magnetostriction of Terfenol-D is several orders of magnitude greater than that of iron, nickel, or cobalt and gives rise to strains in the region of 2×10^{-3}. Bulk Terfenol-D produces much larger strains than those achievable with piezoelectric materials. Research has been undertaken to investigate the feasibility of depositing thin and thick-films of magnetostrictive material onto substrates such as silicon, glass, and alumina; the magnetostriction achievable, however, is inferior to that of the bulk material. fig. 3.5 shows an example of a magnetic actuator as described by Judy et al. [49].

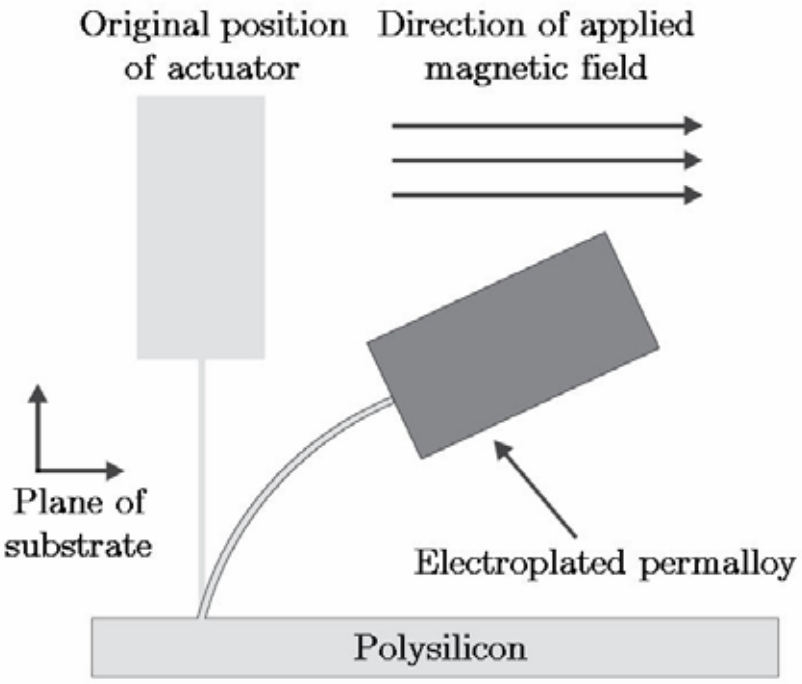


Figure 3.5: An example of an in-plane magnetic actuator[49]

The device comprises a 7 μm-thick layer of Permalloy, which was electroplated onto a polysilicon cantilever. The root of the beam is thin and narrow and acts as a spring, thereby allowing the tip to deflect over a wide angular range.

The magnetic field is applied externally to the device, and this causes a deflection of the actuator in the direction of the plane of the substrate. The device is made using polysilicon surface micromachining techniques. Deflections exceeding 90° were achieved with this configuration.

3.3. Micromirror design and modeling

Electromagnetic actuation is an attractive way to actuate micromirrors for applications that require large deflection angles and robust structures. With electrostatic, thermal, and piezoelectric actuation, large deflection angles require a low spring constant, obtained with very thin mechanical structures. Torsional micromirror rotates about one axis, but is unable to translate; for this chain, it has a good dynamic response and is well suited for applications requiring high speed.

In this work, robust free-standing torsional micromirror structure is proposed. The special design of the structural frame permits one to easily handle

the micromirror, self-align the mirror on the top of the scanning probe, and replace it.

3.3.1. Micromirror layout

Simulation of micromachined systems and sensors is becoming increasingly important. The motivation is that before fabricating a prototype, it is necessary to virtually build the device and predict its behavior. This allows for the optimization of the various design parameters according to the specifications. As it is a virtual device, parameters can be changed much more quickly than actually fabricating a prototype, then redesigning and fabricating it again. This considerably reduces the time to market and also the cost to develop a commercial device.

Simulation software tools for electronic circuits are very mature nowadays, and the level of realism is striking. Often the first fabricated prototype of a novel circuit works in a very similar way as predicted by the simulation. In MEMS, however, this degree of realism cannot be achieved in many cases for two reasons. First, the simulation tools have not reached a similar maturity as their electronic equivalents; second, and more important, simulation of MEMS devices is much more complex. A MEMS device typically comprises many physical domains such as mechanical, electrical, thermal, and optical. All these domains interact and influence each other, making the problem orders of magnitude more difficult.

Originally developed to study stresses in aeronautical structures, the Finite Elements Method (FEM) has known in this decade a widespread diffusion in every field of engineering, such as fluid mechanics, electromagnetism, heat conduction and many more. This flexibility is due to the particular way the FEM uses to simplify complex problems: the geometry of the system under analysis (main domain, generally complex in shape) is divided in smaller parts with a simple shape, the sub-domains, also called elements.
Every element shares some points with neighboring ones called nodes. The finite element procedure allows us to reduce the number of variables involved in the physics of the process and to simplify calculations. The simplification is introduced by considering each element as a sum of differential equations that describe the physics involved in the model. These functions, called shape functions, are spotted by the values of the dependent variables on the nodes.

For the design and modeling of the MEMS micromirror we used finite elements software to study a wide range of microdevices. The mechanical behavior of a micromirror has been studied in order to evaluate the bending effect of an applied load. A bended mirror reflects a light beam that diverges from the optical

theoretical path obtained from a perfect flat mirror. This effect has been minimized in order to maximize the reflection and to obtain the best optical signal.

The micromirror is built completely in nickel. The whole system is made of two parts: a support frame which keep the mirror suspended and take it back to the initial position when the load is removed, and the mirror body, which divert the geometric path of the optical signal.

To simulate the presence of the surrounding nickel structure, we applied the boundary conditions to the opposite faces of the support (Figure 3.6). In this way we take into account the structure to which the mirror is linked without depicting it completely. This allows us to have a model with less nodes, which reqire less computational time, without loosing in accuracy.

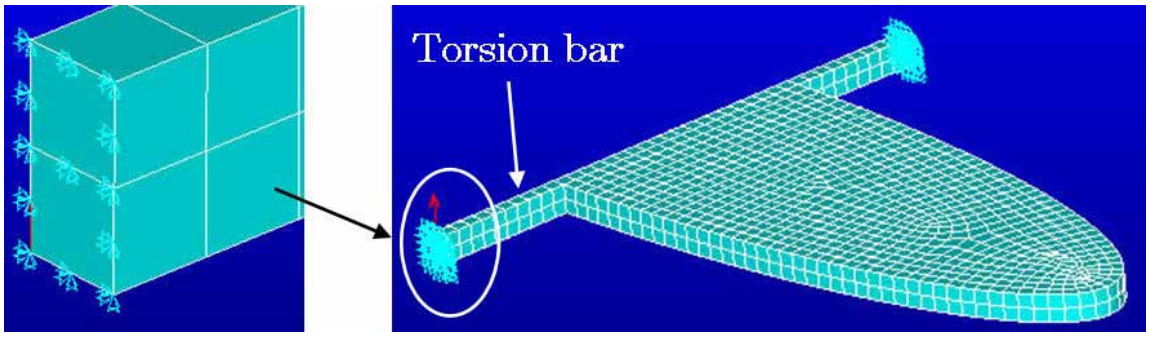
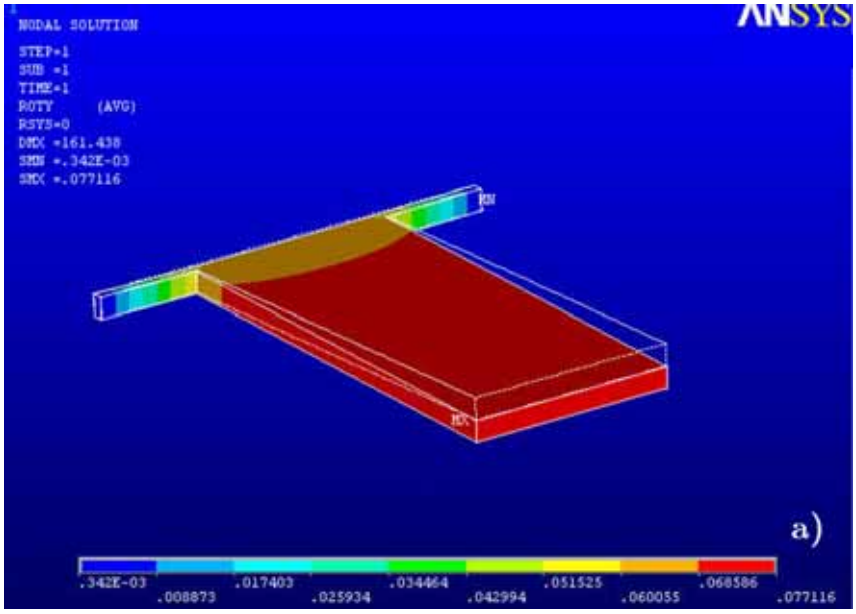
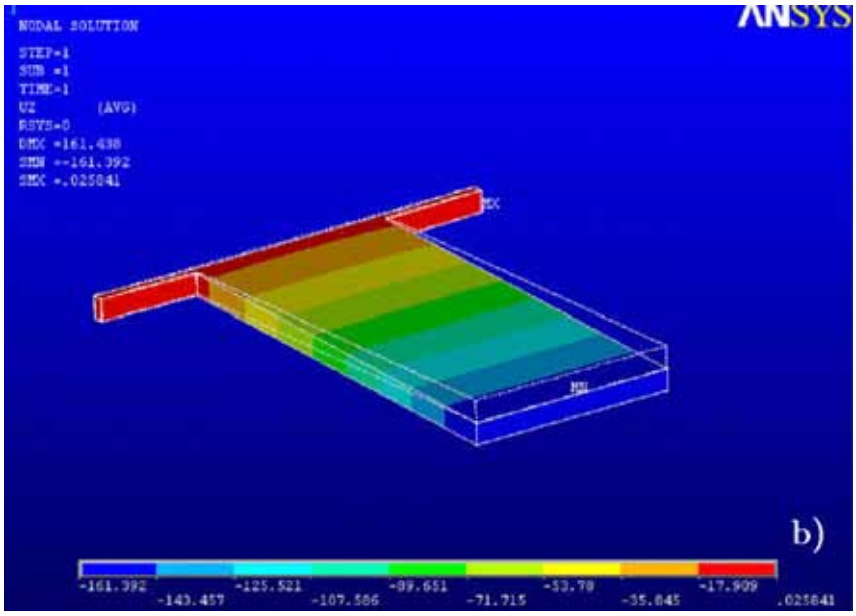


Figure 3.6: Micromirror mesh and boundary condition

We applied a pressure load over the surface of the mirror in order to evaluate the bending angle along the length due to a uniform load distributed over the surface. The pressure load applied is 1×10^{-6} Pa. Furthermore the weight of the system is simulated by a body load which uses an acceleration vector of 9.8 m/s^2 oriented in the same direction as the pressure load. In Figure 3.7, the results gained from the simulations are presented. All the analysis performed were in the elastic range of stress. The data collected investigated the bending angles (Figure 3.7a), the displacement of the mirror's free edge (Figure 3.7b) and the Mises equivalent stresses. This test has been evidenced that the weak points are the edges between the torsion bar and the mirror (red zones in the Figure 3.7c)


NODAL SOLUTION
STEP=1
SUB =1
TIME=1
ROTY (AVG)
RSYS=0
DMX =161.438
SMN =.342E-03
SMX =.077116
a)
.342E-03
.008873
.017403
.025934
.034464
.042994
.051525
.060055
.068586
.077116


NODAL SOLUTION
STEP=1
SUB =1
TIME=1
UZ (AVG)
RSYS=0
DMX =161.438
SMN =-161.392
SMX =.025841
b)
-161.392
-143.457
-125.521
-107.586
-89.651
-71.715
-53.78
-35.845
-17.909
.025841

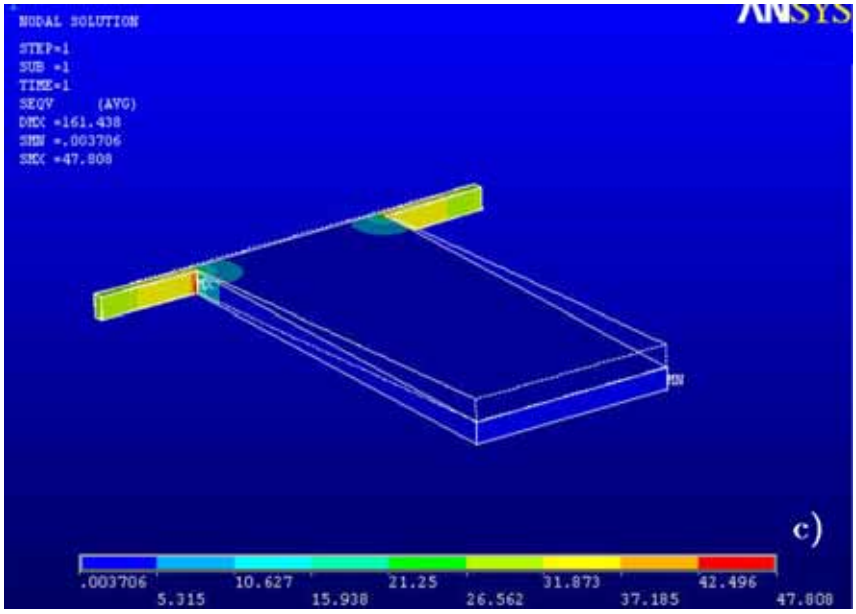


Figure 3.7: Simulation results: a) the bending angle in the rectangular geometry; b) the displacement of the system's nodes along the vertical direction (loads' direction); c) Mises equivalent stress.

To simulate both electromagnetic and mechanic behaviour, we use a coupled field analysis in which two simulations in sequence are performed: the output results given by the electromagnetic analysis solutor are taken as input loads in the mechanical analysis.

3.3.2. Theory of operation

The projected micromirror structure is shown in Figure 3.8. It consists of a structural support plate made of electroplated copper on which a free-standing micromirror body is resting.

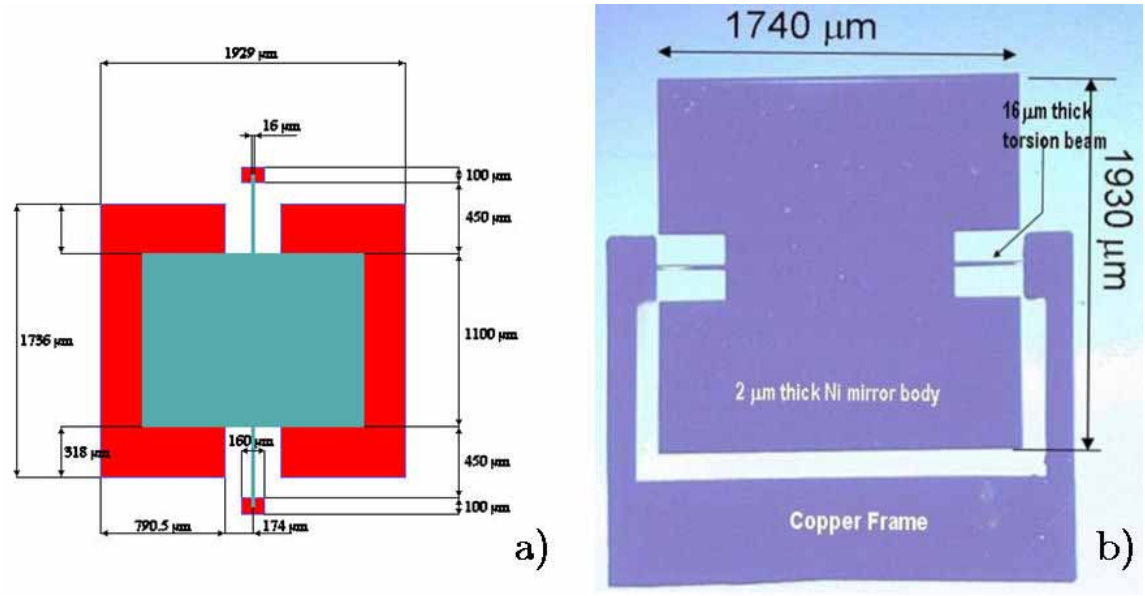


Figure 3.8: a) final layout of micromirror body; b) optical micrograph of fabricated MEMS micromirror with the physical dimensions.

The commonly used materials in MEMS, such as silicon and silicon dioxide, exhibit exceptionally low magnetic susceptibility and thus cannot develop an appreciable internal magnetization [50]. The proposed micromirror is composed of a reflective body and a torsion spring which provide restoring torque and support the mirror structure made in low stress electroplated nickel. Nickel is a ferromagnetic material, therefore strongly attracted by a magnetic force. Copper (diamagnetic material) used for support frame is weakly attracted.

All materials show a diamagnetic response in an applied magnetic field; however for materials which show some other form of magnetism (such as ferromagnetism or paramagnetism), the diamagnetism is completely swamped. Diamagnetic materials have a relative magnetic permeability that is less than 1, thus a magnetic susceptibility which is less than 0, and are therefore repelled by magnetic fields. However, since diamagnetism is such a weak property its effects are not observable in every-day life. Nevertheless these values are orders of magnitudes smaller than the magnetism exhibited by paramagnets and ferromagnets.

When an external electromagnetic field is applied, the micromirror rotates about the torsion bar axis until the restoring torque and the magnetic moment generated are equal. At the rotational angle θ, the restoring torque by the torsion spring can be written as [51]:

$$T_r = k\theta = \frac{2Gw_s^3 t_s}{3l_s}\theta\left[1 - \frac{192}{\pi^5}\frac{w_s}{t_s}\tanh\left(\frac{\pi t_s}{2w_s}\right)\right] \quad \text{(Eq. 3.1)}$$

where k is the spring constant and G, t_s, w_s, and l_s are the shear modulus, the thickness, the width and the length of the torsion spring. The natural frequency equation for the device is derived from damped spring model for torsional displacement and is given by:

$$f_0 = \frac{1}{2\pi}\sqrt{\frac{k}{I}} \quad \text{(Eq. 3.2)}$$

where f_0 is the resonance frequency of the device and I is the mass moment of inertia of the micromirror body.

As the mirror rotates, shear stress is induced by the torsion spring deformation. The special layout used, permits to maximize the length of the torsion spring avoiding the plastic deformation region for big deflection angle without sacrificing the total reflective area.

3.4. Micromirror fabrication process overview

For the fabrication of electromagnetically actuated torsion beam micromirror an extremely simple 2-mask exposure fabrication process using photoresist as sacrificial layer was developed [52, 53]. The important feature of this process include that the torsion beam and micromirror body are made of low stress electroplated nickel, and the torsion beam micromirror is resting on a free standing structural frame also made of electroplated copper. This device structure has resulted in a simplified fabrication process. The major fabrication process steps for free standing surface micromachined electromagnetically actuated structure studied is summarized in Figure 3.9. The individual processing steps are summarized below.

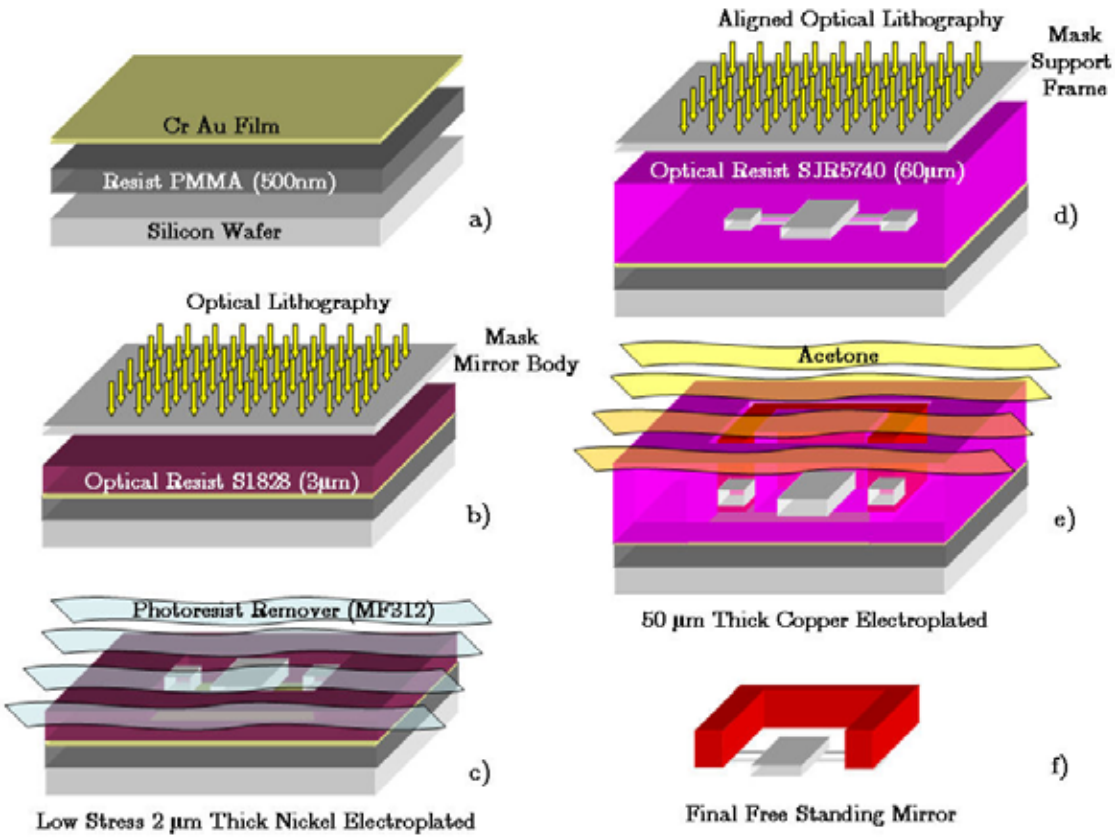


Figure 3.9: Micromirror fabrication process: a) based on a PMMA sacrificial layer; b) and c) optical lithography techniques employed for fabrication of micromirror nickel body and; d) and e) copper support frame; and f) releasing of the final free-standing structure.

3.4.1. Substrate preparation

Although the most popular MEMS fabrication technology is bulk micromachining using deep wet or dry etching below the surface of the silicon, surface micromachining provides a complementary technique in which materials are added above the surface. These materials act as spacers or sacrificial layers to be removed at a later stage to produce freestanding structures and moveable parts. The MEMS freestanding micromirror fabrication process is based on poly-methylmethacrylate (PMMA) resist as sacrificial layer. On a Si substrate is deposed a 500 nm thick PMMA resist by spinning process and backed at 170°C on a hotplate for 10 min. Over PMMA sacrificial layer, a seed layer of Cr/Au (10 nm/20 nm) is deposited by thermal evaporation to facilitate the electroplating of the integrated actuator body comprising of mirror and torsion bar in nickel and also the plating of copper for structural plate on which the mirror will rest finally (Figure 3.10). Once all the processing steps are completed the device is released by removing PMMA

sacrificial layer. The metals can be also deposited by CVD, e-beam evaporation, plasma spray deposition, or sputtering processing. Chromium thin layer is deposed between the PMMA layer and the gold thin film to improve adhesion. The stress in evaporated films is, in general, tensile, with the actual value depending on the pressure in the chamber and the temperature of the substrate. Both parameters were controlled to avoid cracks on the substrate. On this substrate, it is possible to perform the standard optical lithography process. The exposure field required for a single micromirror structure is 4000 µm, which corresponds to the minimum dimension of the starting silicon substrate.

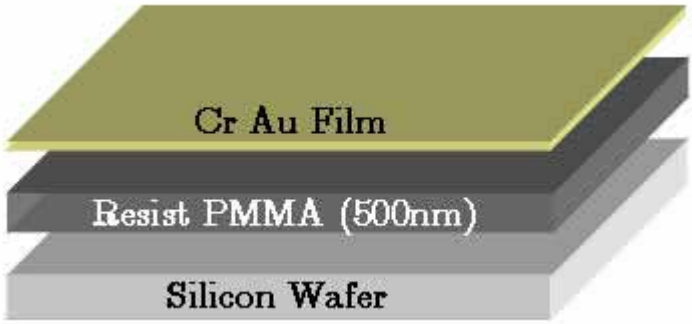


Figure 3.10: Substrate preparation: between the Si substrate and evaporated Cr/Au seed layer was deposed a 500 nm thick PMMA resist

3.4.2. Lithographic process micromirror body

The first step involves the optical lithography process by which the micromirror body and the torsion bar patterns are formed in a chemically resistant polymer, applied by spinning it on to the sacrificial layer substrate (Figure 3.11a). 3µm thick positive photoresist S1828 is spun coated and baked for 2 min at 90 °C on a hotplate. The polymer is exposed to UV light through a quartz mask with an micromirror body and the torsion bar patterned chrome layer on it to break the polymer chains A photomask, with patterns of the integrated actuator body, is exposed (Figure 3.11b). The optical lithography mask is made of glass with a Cr layer on one side; electron beam lithography of micromirror body, combined with a wet-etching process, is exploited to remove the Cr layer in some regions, according to the pattern; at the end the mask will present some opaque areas (with Cr) and other transparent ones (without Cr).

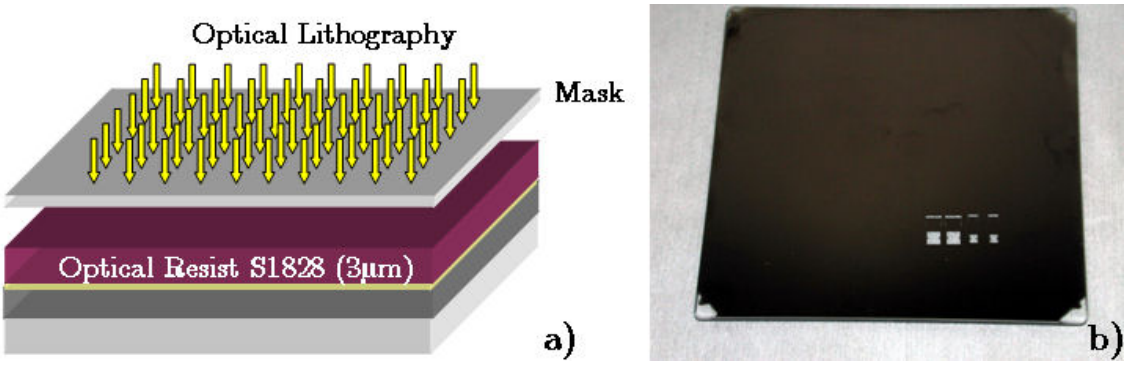


Figure 3.11: first process step: a) optical lithography for the fabrication of the torsion beam and micromirror body; b) photomask with patterns of the integrated actuator body

After exposure the soluble resist (the broken polymer chains) is removed in MF322 developer and the remaining resist is baked in order to harden it against chemical attack. The sacrificial layer substrate is unaffected throughout optical lithography process as it is resistant to etchants for optical resists. After resist development a 30s O_2 plasma cleaning treatment is given to obtain clean surface for efficient electroplating.

3.4.3. Nickel electrodeposition

1.7µm thick nickel is electroplated (Figure 3.12). Nickel, along with Copper, is one of the most widely used materials in electroplating structures. A Nickel sulfamate bath composition, optimized for Ni electroplating, is used. In addition to Nickel sulfamate and boric acid as buffer, a small quantity af an anion-active wetting agent is added. Nickel sulfamate bath produce low internal stress deposits without the need for additional agents, thus avoiding a cause of more defects. The bath is operated at 56°C and at a pH value between 3.5 to 4.0. Metal deposition is carried out at current densities up to 10 A/dm^2. Growth rate vary from 12 (at 1 A/dm^2) to 120 µm/hr (at 10 A/dm^2).

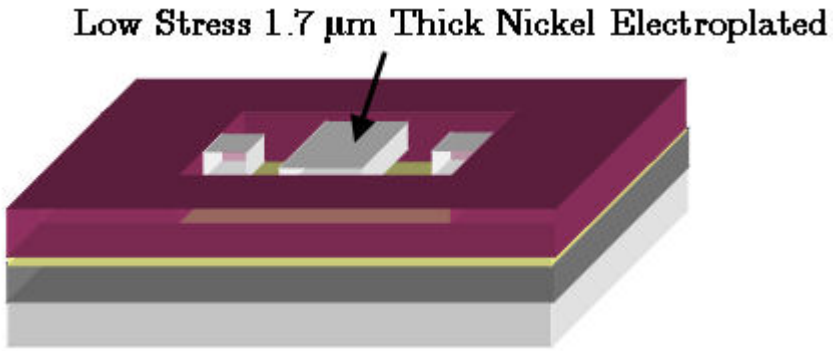


Figure 3.12: electrolitic nickel low stress deposition

The hardness of the Ni deposits can be adjusted from 200 to 350 Vickers by varying the operating conditions. Hardness decreases with increasing current density. To reach a high hardness of 350 Vickers, the electroforming must proceed at reduced current of 1 A/dm^2. Also, for low compressive stress of 20 N/mm^2 or less, a reduced current density must be used. Internal stress in the Ni deposits is not only influenced by current density but also by the layer thickness, pH, temperature, and solution agitation. In the case of pulse plating, pulse frequency has a distinct influence as well [54]. For thin Ni deposits, the stress is high and decrease very fast as function of thickness. For thick Ni deposits (> 30 µm), the stress as a function of thickness reaches a plateau. At a current density of 10 A/dm^2, these thick Ni films are under compressive stress; and at 5 A/dm^2, the internal stress reduces to practical zero. Stirring of the plating solution reduces stress dramatically, indicating that mass transport to the cathode is an important factor in determining the ultimate internal stress. Consequently, since high-aspect-ratio features do not experience the same agitation of a stirred solution as bigger features do, this result in higher stress concentration in the smallest features of the electroplated structure.

For the fabrication of micro device with high deposition rates exceeding 120 µm/hr, a reduction of the internal stress can occur only by raising the temperature of the bath or by using alternative electrodeposition methods. Raising of the bath temperature is not an attractive option, but using alternative electrochemical deposition techniques deserves further exploration. For example, using pulsed (= 500 Hz) galvanic deposition instead of a DC method can influence several important properties of the Ni deposit. Properties such as the grain size, purity, and porosity can be manipulated this way without the addition of organic additives. In pulse plating, the current pulse is characterized by three parameters: pulse current, i_p; pulse duration, t_d; and pulse pause, t_p. These three independent

variables determine the average current density, i_a, which is the important parameter influencing the deposit quality and is given by:

$$i_a = \frac{t_d}{t_d + t_p} i_p \qquad \text{(Eq. 3.3)}$$

Pulse plating leads to smaller metal grain size and smaller porosity due to a higher deposition potential. Because each pulse pause allows some time for Ni_2^+ replenishment at the cathode (Ni_2^+ enrichment) and for diffusion away of undesirable reaction products that might otherwise get entrapped in the Ni deposit, a cleaner Ni deposit results. The higher frequency of pulse leads to the smaller internal stress in the resulting metal deposit [54]. Pulse plating represents only one of many emerging electrochemical plating techniques considered for microfabrication.

3.4.4. Selective resist removal and silver evaporation

After electroplating, the photoresist has been given a flood UV light exposure. To increase the reflectivity of the micromirror body, we evaporated 50 nm thick silver; the roughness of the surface is under 15 nm. The reflectivity of Ni in the visible range is 72% against 97% Silver in the same range. The oxidation does not influence the optical measurement of the scanner and the reflectivity of the mirror is still acceptable.

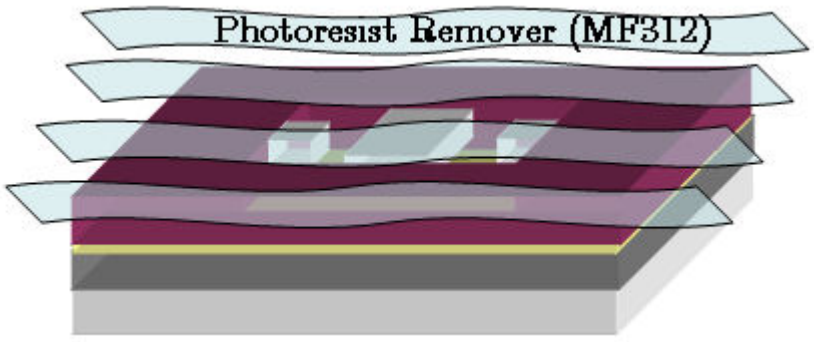


Figure 3.13: Selective remotion of the S1828 optical resist (lift-off process)

The silver deposition is performed over all the sample, covering both the resist structures and the substrate; at the end a resist stripping step (MF322 developer) removes the S1828 optical resist (Figure 3.13).

3.4.5. Lithographic process support frame

The second lithographic step involves the structural support frame fabrication; a 60 µm thick positive tone SJR5740 photoresist is spun coated with a disuniformity less than 0.1% and baked at 96 °C on a hotplate for 15 min (Figure 3.14a). After final soft bake, wafer is kept in ambient for about 90 min to reabsorb water moisture prior to photolithographic exposure. An optical lithography aligned with the micromirror body pattern using a photomask with a pattern of a structural plate is performed (Figure 3.14b).

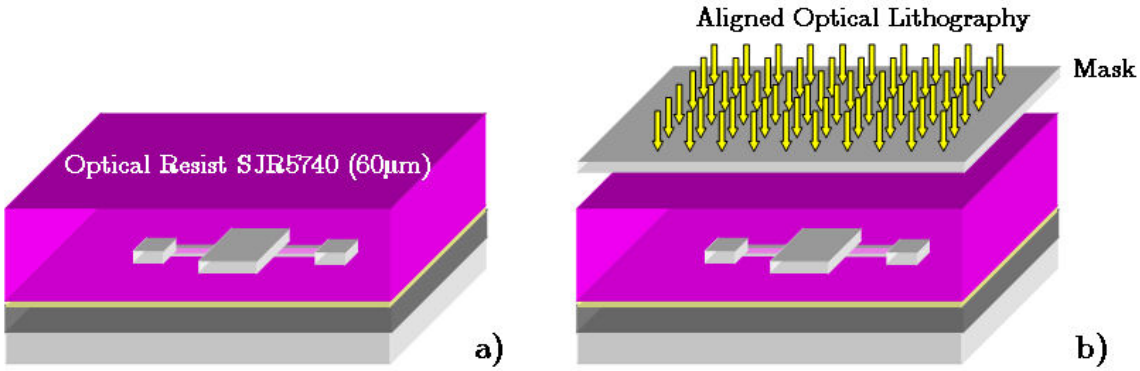


Figure 3.14: a) spin coating process 60 µm thick positive tone SJR5740 photoresist; b) the second optical lithography is aligned with the first one.

Exposed resist is developed in 1:5 mixture of CD-2401 developer and water mixture and rinsed with DI water. After resist development a 90s O_2 plasma cleaning treatment is given to obtain clean surface for efficient electroplating.

3.4.6. Copper electrodeposition

On cleaned surface, about 50 µm thick copper is electroplated. Electrochemically deposited copper is important material for integrated circuit interconnect and packaging applications. Copper's low resistivity value and high electromigration resistance make this material especially attractive for 90 nm and beyond technology nodes of integrated circuit manufacturing, where integration of electroplated Cu and low-k and ultra low-k dielectrics will be essential to further reduction of RC time constants.

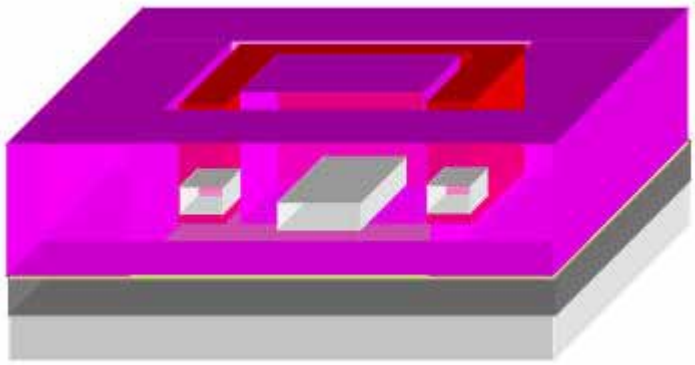


Figure 3.15: 50 μm thick copper electroplated

For our application, we are interested mainly in mechanical and magnetic properties.

Copper is deposited using a commercial available acid-Cu plating solution. At the target current density (3-5 A/dm^2), the deposition rate is 20 μm/hr. An annulus around the periphery of the sample, left bare of photoresist, served as an on-wafer "thief electrode". Although this on-wafer thief is not powered indipendently, good thickness is achieved; across-wafer thickness variation was typically about 3%. The thickness uniformity depended primarily on the layout of the photoresist pattern. Since the current efficiency of the Cu deposition is 100% the deposit thickness is easily controlled once the active area of the photoresist pattern is accurately known.

3.4.7. Selective etching of seed layer coating and removal of sacrificial layer

After, electroplating, resist is given flood exposure and developed in pure CD-2401 developer to remove the photoresist. Last step involves selective etching of Cr/Au seed layer and removal of PMMA photoresist sacrificial layer. Final structure is then released by removing sacrificial PMMA resist layer in hot acetone and rinsed in acetone (Figure 3.16a) and then allowed to dry in air at room temperature.

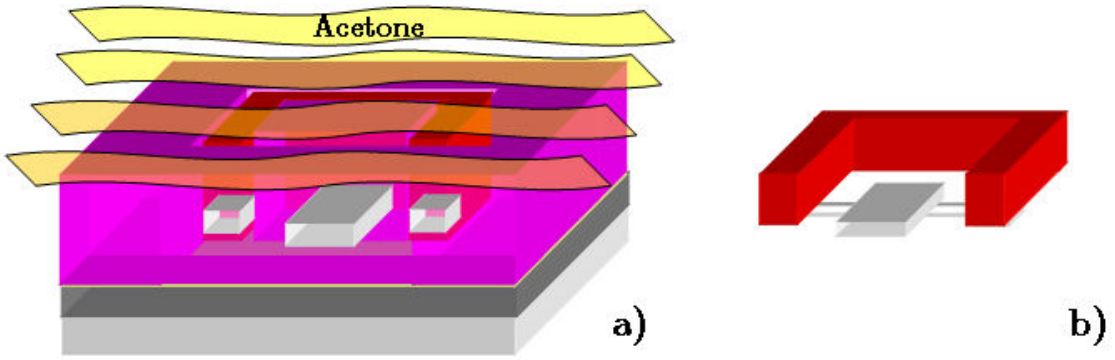


Figure 3.16: a) the final structure is released by removing the sacrificial PMMA resist layer in hot acetone; b) free standing structure, torsion beam and micromirror body in nickel are rested on the copper structural support frame

The challenges with surface micromachining are to control the mechanical properties of the structural layer to prevent the formation of internal residual stresses and to ensure that the released structures do not stick to the surface of the wafer after they are dried. In surface micromachining, structures are generally released by wet etching the sacrificial layer followed by rinsing in water. This gives rise to capillary forces as the wafers are dried causing the structures to stick to the underlying substrate. Many methods for preventing this stiction have been developed. One approach is to process the wafers through a series of rinses such that the final rinse is in a hydrophobic liquid such as hexane or toluene [55]. Another approach relies on changing the phase of the liquid in which the wafers are finally rinsed, either by freezing or heating the liquid into a supercritical state. T-butyl can be frozen solid and sublimed at low vacuum pressures [56]. In the supercritical drying method the final rinse is done in a pressure vessel in liquid CO2, which is then raised into a supercritical state. In this state, the interface between the liquid and gas phases is indistinguishable and there are no surface tension forces [57]. Thus, the CO2 gas can be vented without affecting the structures. Other methods involve dry release of the structures. One such method is etching the oxide in an HF vapor [58], and another entails rinsing the wafers in acetone then adding photoresist, which fills the gaps beneath the structures after the acetone has evaporated. The resist can then be removed in an oxygen plasma.

For the MEMS micromirror sacrificial layer removal, in the final liquid-drying process, a magnetic field is applied to the sample so that the mirror flaps remain perpendicular to the substrate during the liquid evaporation; the magnetic field is maintained until the acetone is completely removed. The torque due to surface tensions is overcome, and sticking is avoided.

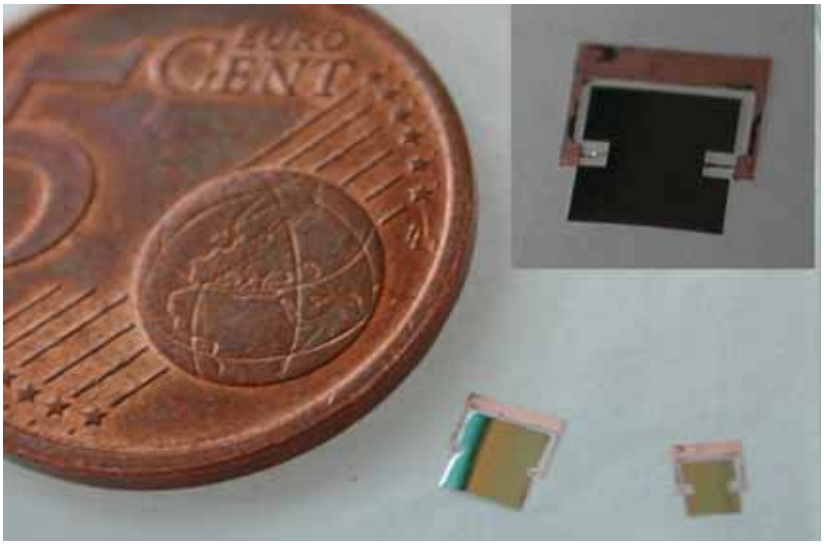


Figure 3.17: Optical micrograph of fabricated MEMS micromirrors

The SEM micrograph of the fabricated micromirror with torsional bar is shown in Figure 3.17. The top of the micromirror is the electroplated copper structural plate beneath of which lie the two torsion beam support pad, which acts as hinges during rotation.

3.5. Micromirror characterization

3.5.1. Natural frequency

The study of the natural frequency of the micromirror structure proposed is important for the evaluation of the static and dynamic performance of the micromirror. The natural frequency is specified by the micromirror layout. Operation close to this frequency must be avoided.

The micromirrors, shown schematically in Fig. 3.8a, can be described as follows [59].

The spring constant k[N·m], is a function of the physical dimensions of the torsion spring:

$$k = \frac{2Gw_s^3 t_s}{3l_s}\left[1 - \frac{192}{\pi^5}\frac{w_s}{t_s}\tanh\left(\frac{\pi t_s}{2w_s}\right)\right] \qquad \text{(Eq. 3.4)}$$

Where, t_s, w_s, l_s, and G are the thickness, width, length of spring (torsion bar) and the sheer modulus of spring material. The natural frequency equation for the device is derived from a damped spring model for torsional displacement is given by:

$$\omega_0 = 2\pi f_0 = \sqrt{\frac{k}{I}} \qquad \text{(Eq. 3.5)}$$

Where, I is the mass moment of inertia and k is the spring constant of the torsion spring (Eq. 3.1).

The mass moment of inertia for the torsion bar micromirror can be associated to a rectangular mass rotates around its symmetry axis and given by:

$$I = \frac{\rho ABZ}{12}\left(A^2 + B^2\right) = \frac{M}{12}\left(A^2 + B^2\right) \qquad \text{(Eq. 3.6)}$$

Where, A, B, Z and M are the thickness, the length, the width and the mass of the micromirror body as shown in Figure 3.18a

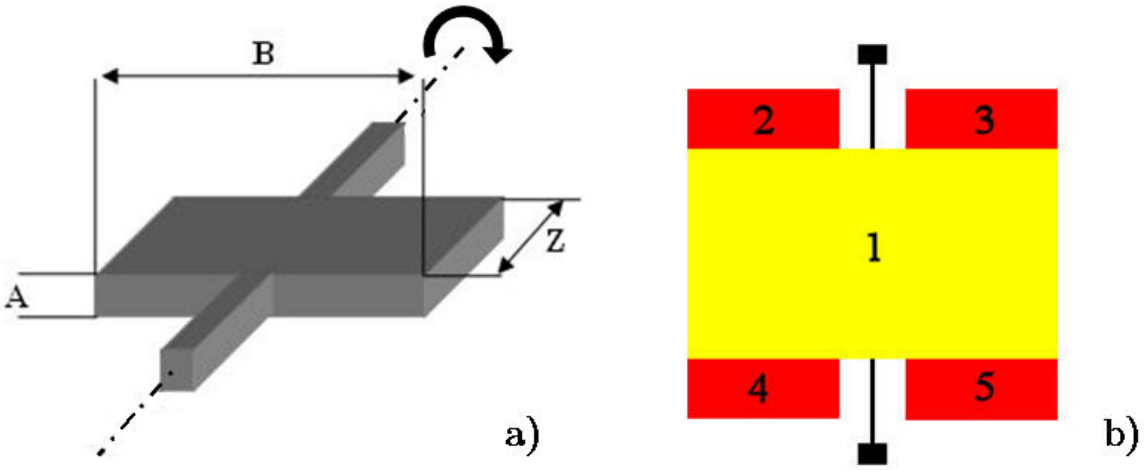


Figure 3.18: micromirror geometrical symmetry a) rectangular mass rotates around its symmetry axis; b) micromirror mass division for the mass moment of inertia calculation

The mass moment of inertia for a rectangular mass rotates around a generic axis is given by:

$$I = \frac{M}{3}\left(\left(A_{\max}^2 + A_{\max}A_{\min} + A_{\min}^2\right) + \left(B_{\max}^2 + B_{\max}B_{\min} + B_{\min}^2\right)\right) \qquad \text{(Eq. 3.7)}$$

Where, (A_{max} - A_{min}), (B_{max} - B_{min}), M are the length, the width and the mass of rectangular mass.

For the proposed micromirror mass moment of inertia calculation, the micromirror body can be considered as composition of 5 fundamental parts, as shown in Figure 3.18b

The physical dimension of the part 1 of the central micromirror body (shown in Figure 3.8a) are:

A = 1.7 µm;
B = 1929 µm;
Z = 1100 µm;
ρ = 8908 Kg·m^{-3} (Nickel volume density).

Solving the Eq. 3.6 is obtained:

$$I_1 = 9.9641 \cdot 10^{-15} \text{ kg} \cdot \text{m}^2$$

The physical dimension of the part 2 of the central micromirror body (shown in Figure 3.8a) are:

A_{min} = -0.85µm;
A_{max} = 0.85µm;
B_{min} = 174µm;
B_{max} = 964.5µm;
Z = 318µm;
ρ = 8908 Kg·m^{-3}(Nickel volume density).

Solving the Eq. 3.7 is obtained:

$$I_2 = 1.4318 \cdot 10^{-15} \text{ Kg} \cdot \text{m}^2$$

The total mass moment of inertia of the micromirror is given by:

$$I_{tot} = I_1 + I_2 + I_3 + I_4 + I_5 = 9.9641 \cdot 10^{-15} + 4 \cdot (1.4318 \cdot 10^{-15}) = 1.5691 \cdot 10^{-14} \text{ Kg} \cdot \text{m}^2$$

From the Figure 3.8a:

t_s = 1.7 µm;

w_s = 16 µm;
l_s = 450 µm;
G = 200 GPa (Nickel shear modulus)

Solving the Eq. 3.4 is obtained:

$$k = 4.8516 \cdot 10^{-8} \text{ N} \cdot \text{m}$$

Finally, the natural frequency is given by:

$$f_0 = \frac{1}{2\pi}\sqrt{\frac{k}{I}} \qquad \text{(Eq. 3.8)}$$

and solving the Eq. 3.8 is obtained:

$$f_0 = 279.856 \text{ Hz}$$

The resulting natural frequency is chosen considering the mechanical stability of the structure, the reflective area dimension of the micromirror and the acquisition frequency of the conoscopic system. The natural frequency of the micromirror is strongly correlated to the torsion bar dimension as is shown in Figure 3.19.

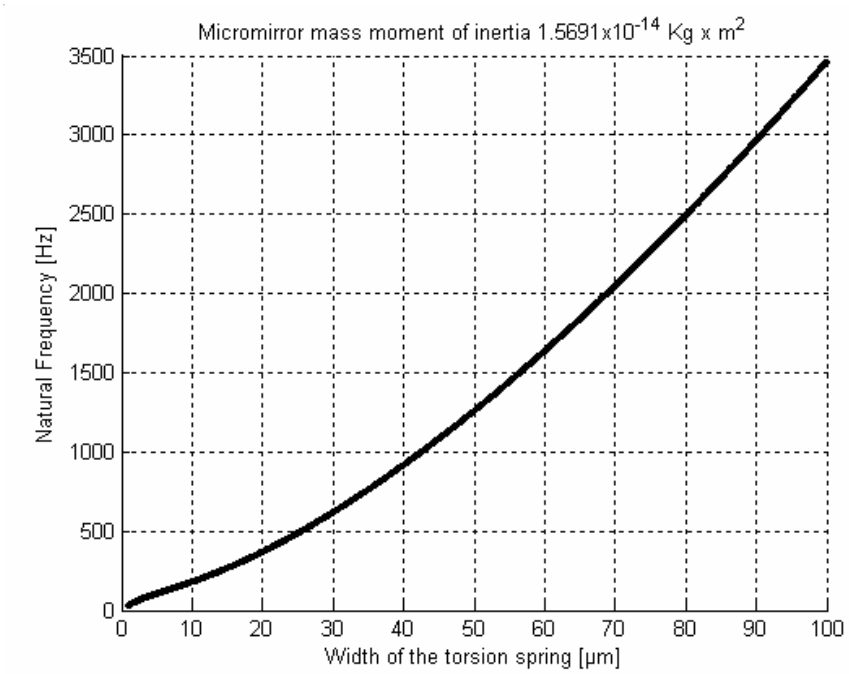


Figure 3.19: Increasing the width of the torsion spring, the natural frequency increases, the maximum displacement decreases

3.5.2. Mechanical test

Micromechanical elements are normally built-up of a movable and a fixed part which are joined by beams. A deflection of the movable element leads to a stress in the torsion beam which is at a maximum at the fixing point of the beam. In most cases the deformation of the torsion beam is periodical. In metallic materials this stress leads to fatigue depending on the amplitude of the stress and the number of stress cycles.

The mechanisms of fatigue in macroscopic structures are well known [60]. Usually the stress leads to a crack initiation which often begins at the surface of the structure. The stress at the end of the crack increases with increasing crack length and accelerates the crack propagation, which first leads to a change of the elastic behavior and finally to a break of the microstructure. The number of stress cycles necessary to destroy the structure, depend on the stress amplitude [61]. There is a strong dependence on the fabrication technique of the material.

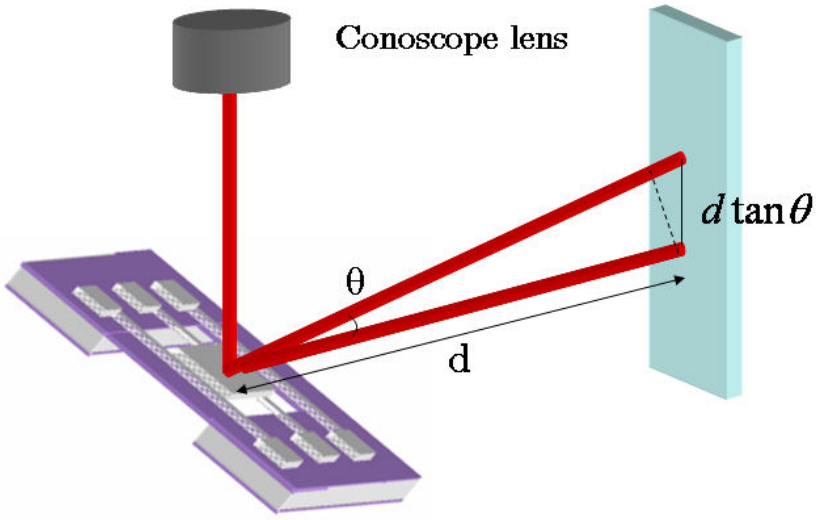


Figure 3.20: Experimental setup for mechanical test

In Figure 3.20 is shown the experimental setup to determine the stress behavior for MEMS micromirror. As the micromirror body is made of electroplating nickel, which shows ferromagnetic behavior, an electromagnetic activation of microstructures, in contrast to silicon microstructures, is possible. Therefore, the MEMS micromirror is positioned over an iron yoke, magnetized by two electrical

coils twisted around it. By connecting alternating voltages to the coils an alternating magnetic field is generated and the micromirror body starts oscillation. To detect the dynamic behavior of the micromirror, a conoscopic system (laser system for distance measurement) is used. The distance between the conoscopic system and the measure screen is a function of the angular position of the micromirror. The displacement $d \tan \theta$ is correlated to the stress values induced by alternating electromagnetic field generated.

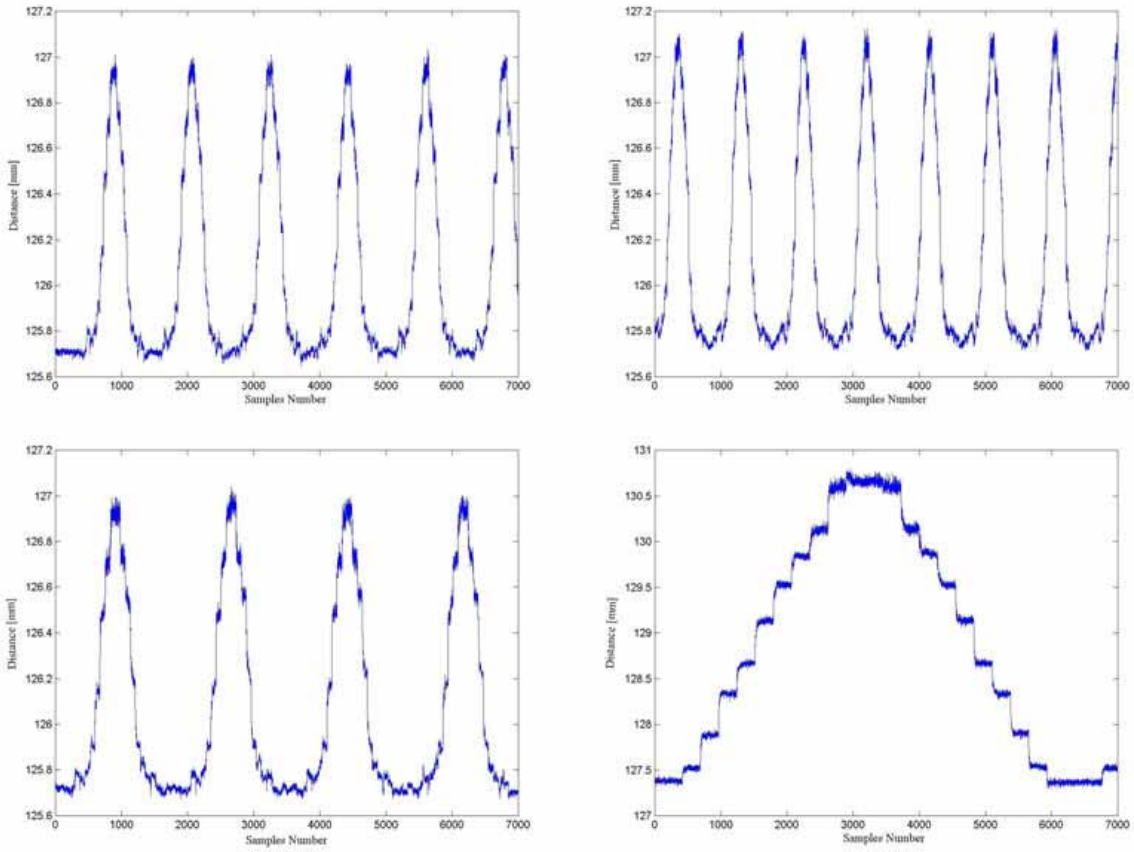


Figure 3.21: Displacement-Number of cycle curves for different actuation frequency. Displacement can be directly associate to the stress value

The MEMS micromirror is exposed to a stress close to the natural frequency for 10^8 stress cycles. From the data test shown in Figure 3.21 it can be concluded that the microstructures could withstand more than 10^8 stress cycles. Moreover, the working frequency for the otoscan3D application is far from the natural frequency. These measurements show in particular that the fatigue effects are negligible in the case of our micromirror where the stress is only more than ten times less. Nevertheless it has to be taken into account that the ratio of surface to volume is much higher in microstructures than in macroscopic structures. As the

crack initiation often starts at the surface of the structures it may be that S-N curves for microstructures will differ from the curves for macroscopic structures.

Experimental results show that, for Ni electroplated structures, the long-term stability reaches the range of comparable literature data for bulk annealed and hardened nickel specimens. Usually, stress leads to crack initiation, which often starts at the surface of the structure. Since microstructures have a higher surface-to-volume ratio, one might have expected the stress curves to differ from macroscopic structures, but so far this has not been observed. To the contrary, it seems that the smaller structures are more stable.

3.6. MEMS scanning probe

3.6.1. Design and characterization of electromagnet and mirror integration

The actuation of the MEMS micromirror is achieved by using a magnetic field generated by an electromagnetic coil stick. Changing the current through copper coils wound on an iron pole, the micromirror deflection angle is controlled. The mirror flipping takes place due to the interaction between the permanent magnetic moment of the mirror and the gradient magnetic field generated by the electromagnet coil. The MEMS micromirror and electromagnet coil assembly compose the optomechanical scanning probe entering the EAC [62]. The layout for the integration of the MEMS micromirror with an electromagnet coil and the concept of electromagnetic actuation under an induced magnetic field is shown in Figure 3.22.

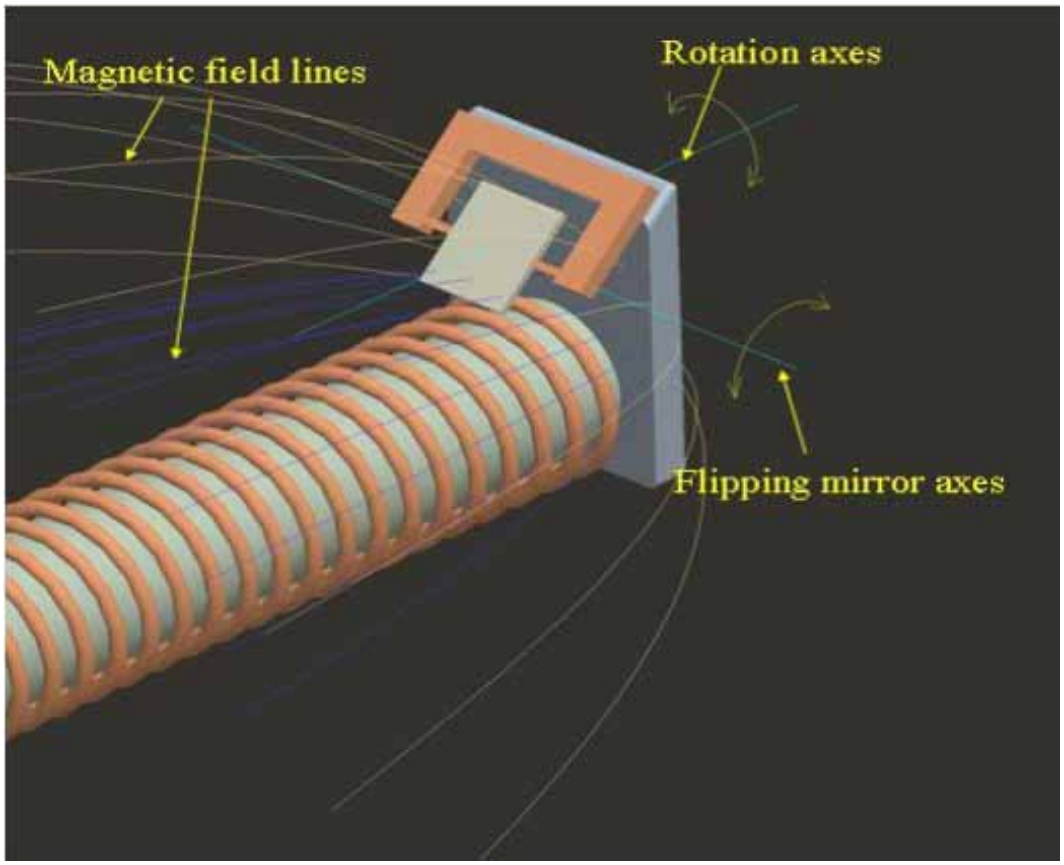


Figure 3.22: Layout for integration of MEMS micromirror with electromagnet coil and concept of electromagnetic actuation under induced magnetic field.

The electromagnet is designed by winding 500 turns of copper wire (150 µm diameter) on a screw core (1 mm diameter) 70mm long (a 13.5 mH inductance) (Figure 3.23). The static and dynamic actuation of the micromirror is achieved by using magnetic fields generated by flowing ac or dc current flows around these copper loops. A micropositioning holder with a 25° slanted top edge is also designed and mounted on the top of the electromagnetic stick.

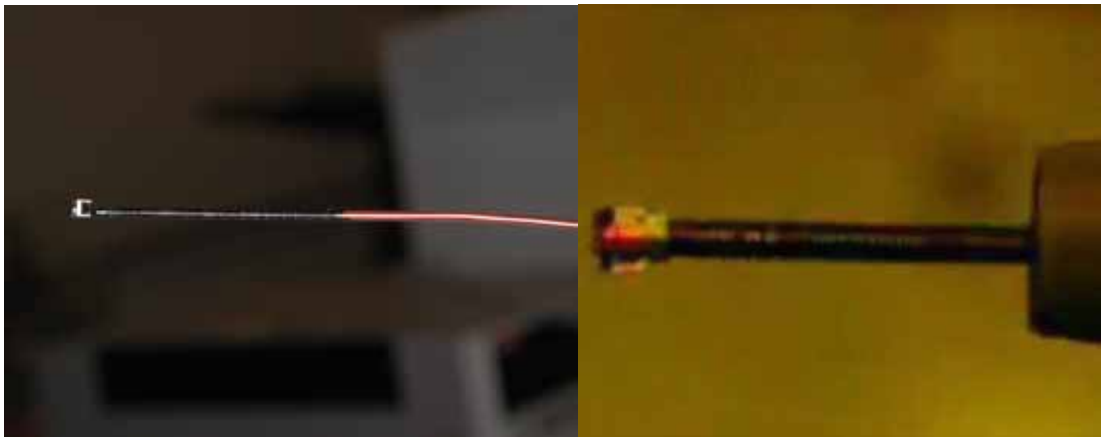


Figure 3.23: Optical image of Layout for integration of MEOMS micromirror with electromagnet coil and concept of electromagnetic actuation under induced magnetic field.

3.6.2. Experimental determination of magnetic field distribution

The dimensions of the positioning holder are optimized by measuring the magnetic field intensity distribution on the top of electromagnet surface by Hall probe. This are done to ascertain that the maximum possible magnetic flux is on the mirror surface. The Hall probe measurement results are summarized in Figure 3.24.

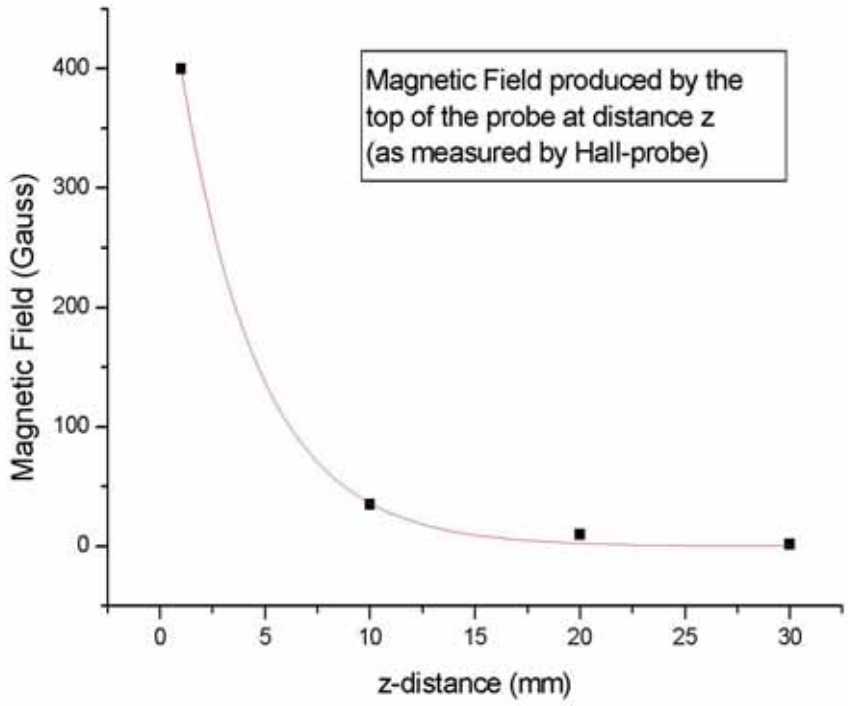


Figure 3.24: Magnetic field intensity distribution on the top of electromagnet surface by Hall probe, measured Magnetic field intensity on top of the electromagnet for fixing actuator in maximum flux position

From measurement, it is found that the magnetic field intensity is maximum at 2.7 mm above the surface of electromagnet coil. Accordingly, the positioning holder is designed so that the mirror axis coincide with this flux of line.

3.6.3. Static and dynamic characterization

The static and dynamic performance of the mirror is characterized by using a laser displacement method. A schematic diagram of our measuring system is illustrated in Figure 3.25.

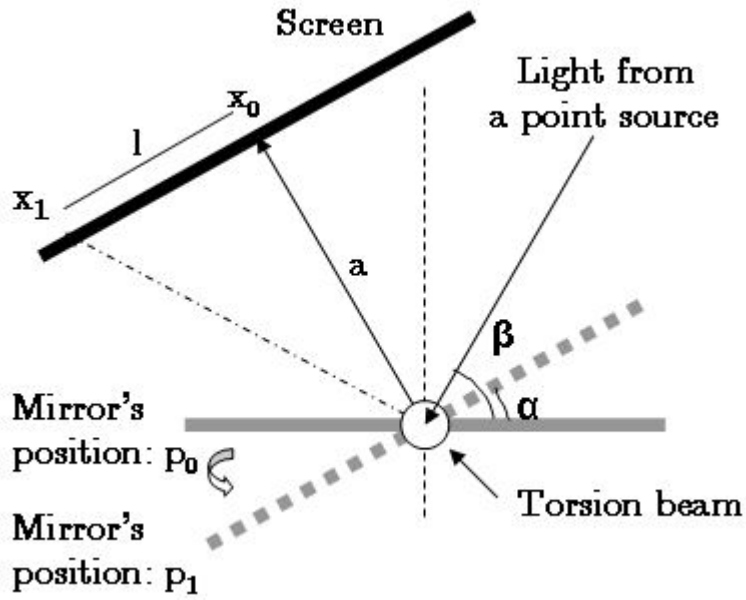


Figure 3.25: Schematic representation of the experimental setup to measure motion characteristics of micromirror

A He-Ne laser is used as a light source because of emitting a good parallel beam. A beam of the He-Ne laser is focused through an optical fiber, with lenses at both ends of it, and a convex lens onto the surface of our micromirror. A graph paper ruled into $1mm^2$ is used as a screen to measure the distance. A light beam from a point source is specularly reflected from the center of a micromirror at a reflection angle ß. The light beam makes a light spot on a screen. The light beam reflected from the mirror is a normal line to the screen. When a micromirror rotates about the torsion beam at an angle α, a mirror's position moves from p_0 to p_1. And the light spot moves from x_0 to x_1 (Figure 3.25) on the screen because of changes of the reflection angle (from β to $\beta - \alpha$ or $\beta + \alpha$). For the angular deflection the following equation can be geometrically derived:

$$\alpha = \frac{1}{2}\arctan\left(\frac{l}{a}\right) \qquad \text{(Eq. 3.9)}$$

where l is the distance between x_0 and x_1 on the screen and a is the distance between the centre of the micromirror and x_0 on the screen.

We measured the angular deflection α by supplying a *dc* magnetic field generated by the electromagnet with an iron core for measurements of static characteristics and a *ac* magnetic field for measurements of dynamic

characteristics. We decided to use an upper limit of angular deflection of 25°, it is obtained for the investigated current of 200 mA.

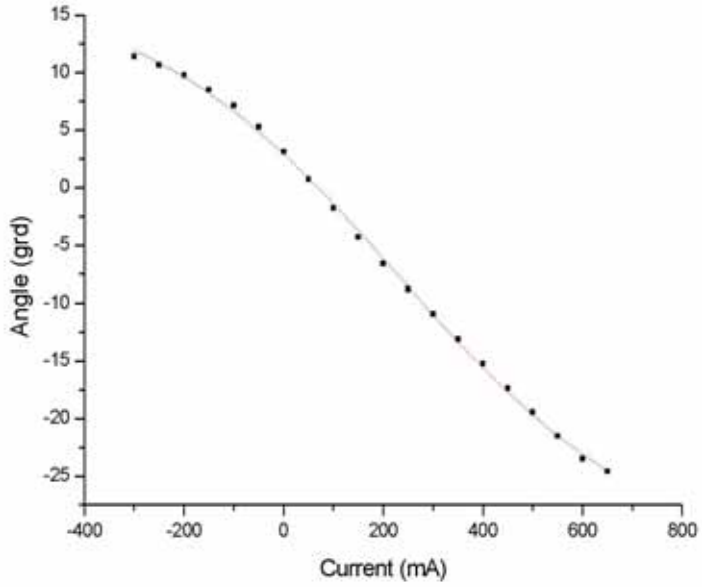


Figure 3.26: Measured and calculated bidirectional deflection as function of coil current.

The micromirror is first activated with the magnetic field increasing in one polarity. The polarity is then reversed and the magnitude of the magnetic bias is increased again. The angular-deflection curves for both cases highlighting a the low field level and the results is plotted in Figure 3.26. Angular deflection α exhibit linear dependence to the coil current for positive bias and fits well with the simulated results.

A sinusoidal current of amplitude of 200 mA is applied to investigate the dynamic response of the micromirror. For a given torsion micromirror, the maximum displacement (Δx_{max}) available for maximum force (F_{max}) can be accounted for by:

$$\Delta x_{\max} = \frac{F_{\max}}{k} \qquad \text{(Eq. 3.10)}$$

and combing Eq. 3.10 with the natural frequency of the micromirror in Eq. 3.8 we obtain:

$$f_0^2 \cdot \Delta x_{\max} = \frac{1}{4\pi^2}\frac{F_{\max}}{I} = k \qquad \text{(Eq. 3.11)}$$

Eq. 3.11 clearly states that the displacement will decrease upon increasing the frequency.

Another factor which is examined to know the current carrying capacity of copper wire of the coil winding of electromagnet. It is critical to know the maximum current flow achievable in order to avoid the burning of the electromagnet coil. The maximum current is determined by using the relation

$$\frac{\partial^2 T}{\partial x^2} = -\left(\frac{I^2 R}{AK_t}\right) \qquad \text{(Eq. 3.12)}$$

and integrating it twice. Temperature (T), the distance from the anchors (x), the thermal conductivity of air (K_t), the cross-sectional area of the coil (A), and resistance per unit length (R) are all taken into account. The above differential equation is solved with boundary conditions $T = T_0$ at $x = 0$ and $T = T_0$ at $x = L$, where the length of the entire wire loop is (L). Using these values, the maximum current that could safely be passed through the coil is estimated to be 800 mA.

Accordingly, all the measurement are carried below this current value to avoid the damage of coil.

4 Conoscopic holography

Conoscopic holography was invented in 1985 by Gabriel Sirat and Demetri Psaltis at the California Institute of Technology. Conoscopic holography is an incoherent holographic technique based on the properties of crystal optics. In the basic interference set-up a point of light is projected on a diffuse object. This point creates a light point, which diffuses light in every direction. In a conoscopic system a complete solid angle of the diffused light is analyzed by the system. The measurement process corresponds to the retrieval of the distance of the light point from a fixed reference plane.

3D measurement systems based on conoscopic holography are steadily gaining ground against older techniques. Conoscopic systems have evolved from an exotic solution to the first choice for all measurement problems not covered by triangulation techniques. Conoscopic systems provide good performance while measuring on a wide variety of materials, which could hardly be measured with other non-contact sensor. The ear has a translucent surface which offer a good response to the conoscopic technique.

In the first part of this chapter, following Gabriel Sirat's considerations, a simplified version of the theory of conoscopic holography is presented. The point-spread function and the transfer function of the conoscopic system are presented and the conoscopic hologram is defined. The basic schemes for reconstruction are presented and finally, the resolution of the system is quantified.

In the second part of the chapter the conoscopic system is presented. The distinction of this system from the conventional triangulation system and the performance characteristics are also discussed.

4.1. Introduction

The theory and experiments on light propagation in uniaxial crystals can be traced back to the early days of optics; the analogy between interference of coherent light and interference in crystalline media is also well established. However, the paper of Sirat and Psaltis [63] was the first to propose the use of crystal optics phenomena directly as the basis of an incoherent holographic system. The term conoscopic is used to describe this phenomenon.

First the conoscopic holography is described in simple word using ray optics. Figure 4.1 presents the basis configuration of the conoscopic holography setup. It is slightly inaccurate because it shows the ordinary and the extraordinary travelling along the same path. Indeed, a more accurate representation will spatially separate the ordinary and the extraordinary rays to take into account the differences of indices and walk-off effect.

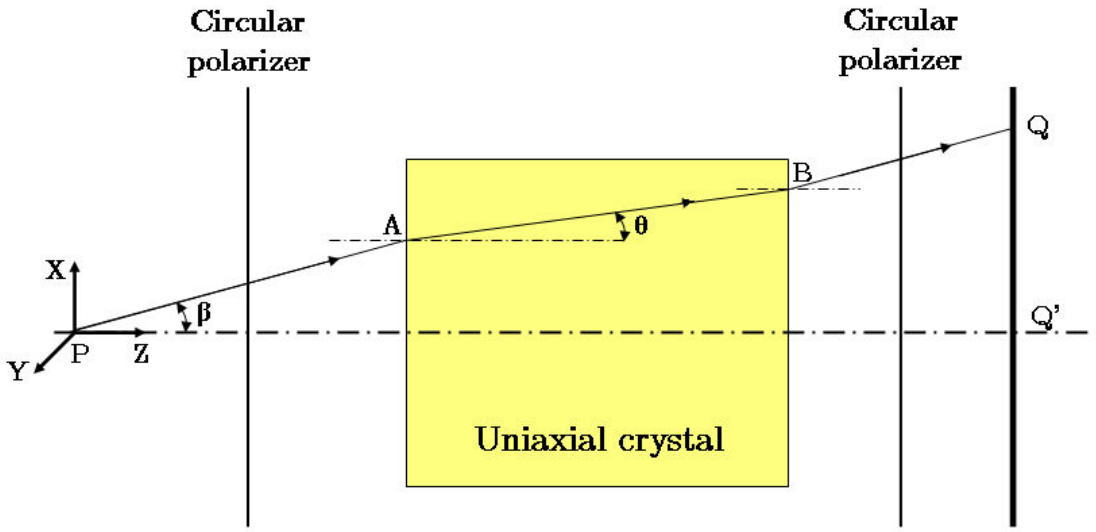


Figure 4.1: Simplified representation of on-axis conoscope

However for accuracy a mean value for the angle of the ray joining point P and Q, both in vacuo (β) and in the crystal (θ), is used. The differences between the phase and the energy-propagation directions are also neglected.

Consider, separately, each point of an object (point P of Figure 4.1) emitting (or diffusing or reflecting) incoherent, unpolarized, quasi-monochromatic light.

Consider also a single ray emerging from this point at an angle β as in Figure 4.1. The light ray passes through a circular polarizer and generated two orthogonal polarizations, one having a quarter-wavelength relative phase delay. The two rays propagated through a uniaxial crystal, along the same geometrical path (in first-order approximation), one in the ordinary mode, one in the extraordinary mode, but both with different light velocities owing to the different refractive indices. The ordinary refractive index is constant, whereas the extraordinary refractive index (and so the phase delay between the two rays) is a function of the direction of the ray relative to the crystal optical axis. These two rays are put back into the same polarization mode by a circular analyzer places after the crystal therefore they interfere. The interference is constructive or destructive depending on the phase delay between the ordinary and the extraordinary rays. The circular analyzer also compensates the initial quarter-wavelength delay.

In an on-axis conoscope the optical axis of the crystal is parallel to the geometrical axis of the system, and, in this case, the conoscopic point-spread function is the Gabor zone lens. The on-axis conoscope is said to be positive if the two polarizer are of the same handedness and negative otherwise.

The conoscopic hologram is the superposition of the elementary conoscopic figures for each point of the object and contains complete information about the object's spatial distribution. Reconstruction is the process of retrieving (optically or numerically) three-dimensional information about the object from the recorded hologram.

The strength of the effect, in a given geometrical configuration, is proportional to the rate of variation of the extraordinary refractive index as a function of the angle. It will be shown to be proportional to the maximum difference between the ordinary and the extraordinary refractive indices: the birefringence scaled by the square of a mean refractive index. One of the properties of conoscopic holography is the strength of static birefringence. It is the major strong effect in optics: The ratio of the birefringence to the square of the mean refractive index can be as much as 0.1 - 0.2, compared with an effective index change typically of 10^{-3} - 10^{-4} in dynamic optical affects (with the exception of liquid crystal).

4.2. Conoscopic point-spread function

Conoscopic holography can be analyzed as a special case of amplitude-splitting systems. The beam is polarized by the first polarizer into a polarization state that

contains an equal amount of the two modes of the crystal. The crystal splits the beam into its two mode: the ordinary and the extraordinary modes. The two modes are slightly attenuated and separated both in phase and in space; the analyzer returns them to the same polarization mode. In the recording plane we obtain the sum of two superimposed fields that interfere, one that propagates according to the ordinary mode of the crystal and one that propagates according to the extraordinary mode of the crystal. They are called the ordinary and the extraordinary beams.

The basic on-axis conoscopic system is represented in Figure 4.1. It is build from a uniaxial crystal placed between two circular polarizers. The ordinary and the extraordinary refractive indices are n_0 and n_e, respectively. The optical axis of the conoscopic system is, in an on-axis system, parallel to the axis of the birefringent crystal. The difference between crystal axis defined as a direction and optical axis defined as geometrical axis is highlighted. The on-axis conoscope is positive if the same handedness and negative otherwise.

The fact that the crystal axis is defined as a direction makes the system space invariant. Note that, in conoscopic holography, the space invariance ensures freedom from all aberration [64]. Indeed, because numerical processing is being considered, the so-called spherical aberration is viewed not as an aberration but as higher-order terms in the point-spread function, which can be compensated for, quite simply.

The space invariance of the phenomenon means also that we can use Fourier analysis in our derivations and that the point-spread functions of the system are functions of the difference in the lateral coordinates of the emitting and detecting point [65].

Let a point P, positioned at (x,y,z), emit quasi-monochromatic light at mean wavelength λ, passing through the on-axis conoscope. The polarizer and the analyzer are taken to be the same handedness, so that the on-axis conoscope is positive; i.e., the interference is constructive for light parallel to the optical axis. The analytical signal describing the complex amplitude field at P is $u(P,t)$, where the spatial dependence of the self-coherence function is written explicitly.

The basic physical description of an amplitude-splitting system is the self-coherence function $\Gamma(P,\tau)$, given by:

$$\Gamma(P,\tau) = \left\langle u(P,t+\tau)u^*(P,t)\right\rangle \qquad \text{(Eq. 4.1)}$$

and the normalized self-coherence function $\gamma(P,\tau)$, given by:

$$\gamma(P,\tau)=\frac{\langle u(P,t+\tau)u^{*}(P,t)\rangle}{\langle u(P,t)u^{*}(P,t)\rangle}=\frac{\Gamma(P,\tau)}{\Gamma(P,0)} \qquad \text{(Eq. 4.2)}$$

where the zero-frequency self-coherence function is related to the intensity emitted at point *P*, *I(P)*, through:

$$I(P)=(\varepsilon_o c/2)\langle|u(P,t)|^2\rangle=(\varepsilon_o c/2)\Gamma(P,0) \qquad \text{(Eq. 4.3)}$$

The self-coherence function is the analytical signal related to the conoscopic hologram. In this chapter a simpler description is chosen, based on ray optics. Simplifications implied by this approach are described and summarized in Appendix B, as some required derivations.

The light from point *P* can be in any polarization state. To simplify the system notation, intensities and amplitudes are the intensities and the amplitudes in the polarization mode of the input polarizer, for the reason that the other polarization is absorbed by the polarizer and cannot be detected. This point has to be taken into account when one is calculating the overall efficiency of a real system.

Let the recording point *Q* be positioning at *(x',y',z')*. At point *Q* the ordinary and the extraordinary fields ($A_o(Q,P)$ and $A_e(Q,P)$) resulting from light emerging from *P* have: the same modulus, because all the simplifying assumptions to make them so (see Appendix B) have been made; and the phase difference between them is $\Delta\phi$, which is calculated below.

Accordingly, the amplitude *A* is defined to be the modulus value of the amplitude of the field:

$$A=|A_o(Q,P)|=|A_e(Q,P)| \qquad \text{(Eq. 4.4)}$$

If the crystal had been replaced with an isotropic medium whose refractive index was equal to the ordinary refractive index, a field equal to twice the ordinary field $A_o(P,Q)$ would be obtained, giving:

$$I_o(Q,P)=(\varepsilon_o c/2)(2A)^2 \qquad \text{(Eq. 4.5)}$$

$I_o(Q,P)$ is the isotropic intensity in mks units. In most cases the angle of the light ray is small (in other words, the detector is positioned quite far from the emitter) so

that the isotropic intensity is close to uniform. Point *P*, emitting quasi-monochromatic light with an intensity *I(P)*, the isotropic intensity, is given by the inverse square law:

$$I_o(Q,P) = \left(T' / z_c^{\,2}\right) I(P) \qquad \text{(Eq. 4.6)}$$

where *T'* is an intensity transmittance coefficient and the approximation of the optical path (Eq. A.11) is used in its simplified version (Eq. A.12). Then we drop the literal recording-point dependence simply by utilizing:

$$A^2 = \frac{2I_o}{4\varepsilon_o c} = \frac{T}{\varepsilon_o c z_c^{\,2}} I(P) \qquad \text{(Eq. 4.7)}$$

where z_c is the conoscopic corrected distance defined in Appendix A (Eq. A.8) and $T = T'/2$. This particular normalization of the intensity transmittance coefficient is chosen because the basic reference measure is the bias that is half the isotropic intensity.

The phase delay $\Delta\Phi$ is given by the difference in the optical path length between the ordinary and the extraordinary rays, in radiant, scaled by the wavelength:

$$\Delta\phi = 2\pi\left(l_e - l_o\right)/\lambda \qquad \text{(Eq. 4.8)}$$

Using the simplifying statement in the Appendix B, assume that the two rays travel along the same geometrical path but a different velocities in the crystal.

With reference to Figure 4.1:

$$l_o = \overrightarrow{PA} + n_o \overrightarrow{AB} + \overrightarrow{BQ} \qquad \text{(Eq. 4.9)}$$

$$l_e = \overrightarrow{PA} + n_e(\theta)\overrightarrow{AB} + \overrightarrow{BQ} \qquad \text{(Eq. 4.10)}$$

Under the approximations of Appendix B we take $\overrightarrow{AB}$ to be simply equal to *L* the crystal length, neglecting in particular the influence of the nonnormal incidence of the rays as well as the effects that are due to the crystal optics on the extraordinary ray; then:

$$l_e - l_o = n_e(\theta)\overrightarrow{AB} - n_o\overrightarrow{AB} = \Delta n_e(\theta)\overrightarrow{AB} \cong \Delta n_e(\theta)L \qquad \text{(Eq. 4.11)}$$

where $\Delta n_e(\theta)$ is the departure of the extraordinary refractive index at angle θ from the ordinary refractive index ($\Delta n_e(\theta) = n_e(\theta) - n_o$).

Using the approximation for the extraordinary refractive index developed in Appendix B (Eq. B.7), which we rewrite as:

$$\Delta n_e(\theta) \cong \Delta n \frac{n_o^{\,2}}{n_e^{\,2}}\theta^2 \qquad \text{(Eq. 4.12)}$$

and using the approximation of θ, which is the angle direction of the ray in the crystal (Eq. B.8),

$$\theta \cong r / z_c \qquad \text{(Eq. 4.13)}$$

we obtain:

$$\Delta\phi = 2\pi\frac{l_o - l_e}{\lambda} \cong 2\pi\frac{l_o - l_e}{\lambda} = 2\pi\frac{\Delta n L r^2}{\lambda n_c^{\,2} z_c^{\,2}} = \pi\frac{\kappa_o}{z_c^{\,2}}r^2 \qquad \text{(Eq. 4.14)}$$

where Δn is the birefringence $(n_e - n_o)$ defined in Appendix A (Eq. A.1) and L is the crystal length, n_c, the effective refractive index (close to the extraordinary refractive index n_e), is defined in Appendix A (Eq. A.2). r^2 is the projected radial coordinate, i.e., $r^2 = (x - x')^2 + (y - y')^2$. z_c the conoscopic corrected distance close to the longitudinal distance $(z - z')$, is defined in Appendix A (Eq. A.8).

For the positive on-axis conoscope the detected intensity Q, $I_+(Q)$ is given by:

$$I_+(Q) = (\varepsilon_o c/2)\left[A_o(Q,P) + A_e(Q,P)\right] \times \left[A_o(Q,P) + A_e(Q,P)\right]^* \qquad \text{(Eq. 4.15)}$$

or, with the modulus and phase delay value just introduced:

$$I_+(Q) = (\varepsilon_o c/2)2\left|A^2\right|\left[1 + \cos(\Delta\phi)\right] = \left[\frac{T}{z_c^{\,2}} + h_2(M)\right]I(P) \qquad \text{(Eq. 4.16)}$$

where $h_2(M)$ is the conoscopic point-spread function (CPSF) of the system, calculated below. The meaning of the subscript *2* is that this result correspond to

the second order (in angular coordinates) approximation. M is the vectorial difference between P and Q, i.e., M is represented by $(x - x', y - y', z - z')$. The shift-invariance properties discussed in the introduction are hidden in this derivation.

The first part Eq. 4.16 is the bias that is proportional to the intensity of the point. The bias is half the equivalent isotropic intensity of Eq. 4.5. The second term is the cosine hologram (of one point) given by:

$$H(Q) = h_2(M)I(P) = \frac{T}{z_c^{\,2}}\cos(\Delta\phi)I(P) \qquad \text{(Eq. 4.17)}$$

$$h_2(M) = \frac{T}{z_c^{\,2}}\cos(\Delta\phi) = \frac{T}{z_c^{\,2}}\cos\left[\frac{2\pi\Delta nLr^2}{\lambda n_c^{\,2} z_c^{\,2}}\right] = \frac{T}{z_c^{\,2}}\cos\left[\frac{\pi\kappa_o r^2}{z_c^{\,2}}\right] \qquad \text{(Eq. 4.18)}$$

where k_o the conoscopic parameter is defined in Appendix A (Eq. A.13).

For a negative on-axis conoscope the second circular polarizer is of inverse handedness. In this case, instead of cancelling the initial quarter-wave delay between the ordinary and the extraordinary rays it adds a supplementary quarter-wave delay, and so the extraordinary ray's phase is inverted relative to the ordinary one.

Experimental result are shown in Figure 4.2.

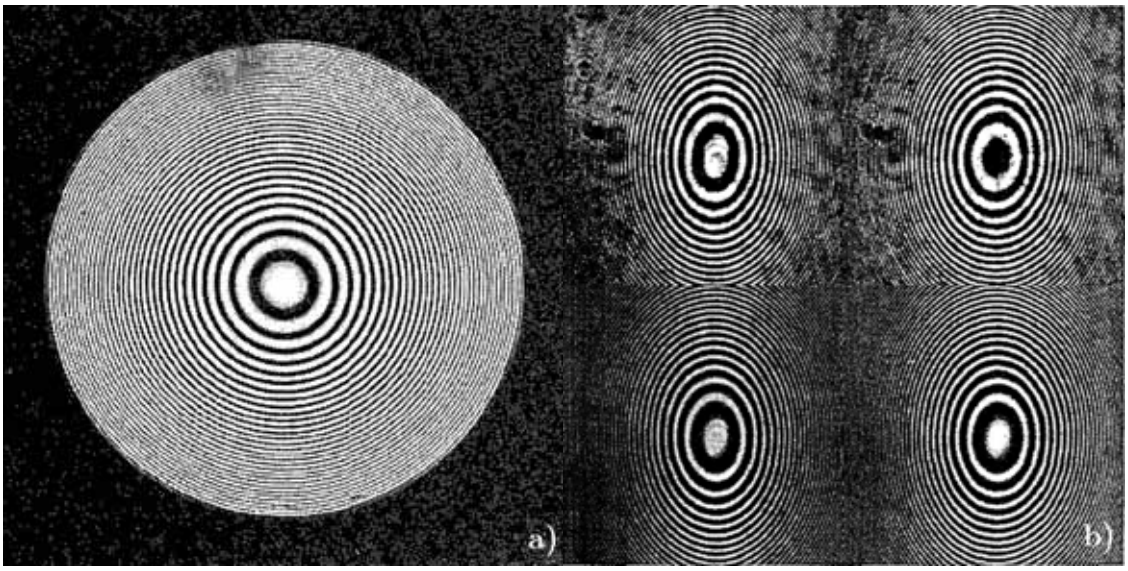


Figure 4.2: Experimental results. a) Positive conoscopic on-axis figure; a) Positive on-axis figure with numerical subtraction as described in Eq. 4.16.

The exponential conoscopic hologram is the complex signal obtained by replacing the cosine dependence with a complex exponential. It is analytical and equal to the self-coherence function. The CPSF, $h_2(M)$, can be thought as the conoscopic hologram of a point located at the origin (in which case $M = Q$) of unit intensity. Its Fourier transform is the conoscopic transfer function, and we define the exponential CPSF (ECPSF) and the exponential conoscopic transfer function accordingly.

The ECPSF is given by:

$$f_2(M) = \frac{T}{z_c^{\,2}} \exp(j\Delta\phi) = \frac{T}{z_c^{\,2}} \exp\left(j \frac{2\pi\Delta n L r^2}{\lambda n_c^{\,2} z_c^{\,2}} \right) = \frac{T}{z_c^{\,2}} \exp\left(j \frac{\pi\kappa_o r^2}{z_c^{\,2}} \right) \qquad \text{(Eq. 4.19)}$$

and the relation between the CPSF and the ECPSF is:

$$h_2(M) = \left[f_2(M) + f_2^{\,*}(M) \right] / 2 \qquad \text{(Eq. 4.20)}$$

4.2.1. Conoscopic transfer function

The conoscopic transfer function is given by the Fourier transform of the CPSF. For the ECPSF, using Eq. 4.19, we obtain:

$$\begin{aligned} \tilde{f}_2(u,v) &= FT_{xy}\left[\frac{1}{z_c^{\,2}} \exp\left(j\pi\kappa_o \frac{r^2}{z_c^{\,2}} \right) \right] = \\ &= \frac{-j}{\kappa_o} \exp\left(j \frac{\pi z_c^{\,2} \rho^2}{\kappa_o} \right) = \frac{-j}{f_R z_c^{\,2}} \exp\left(j \frac{\pi\rho^2}{f_R} \right) \end{aligned} \qquad \text{(Eq. 4.21)}$$

where FT_{xy} has the meaning of the Fourier transform relative to the x,y variables from the Goodman definition, ρ is the polar coordinate in the Fourier domain ($\rho^2 = u^2 + v^2$), and f_r, the Fresnel parameter, is defined in Appendix A (Eq. A.14).
The conoscopic transfer function is simply given as:

$$\tilde{h}_2(u,v) = FT_{xy}\left[\frac{f_2(x,y) + f_2^{\,*}(x,y)}{2} \right] = \frac{-1}{\kappa_o} \sin\left(\frac{\pi z_c^{\,2} \rho^2}{\kappa_o} \right) \qquad \text{(Eq. 4.22)}$$

It is shown in Appendix A (Eq. A.9) that the ratio between the differentials of z and z_e is close to 1; in this part we substitute z_e as the variable in the integrals instead of z. This substitution introduces only a intensity scale change, as a function of z_c, in the reconstructed image.

We define the three-dimensional conoscopic transfer function to be the three dimensional Fourier transform of the CPSF relative to the conoscopic corrected distance z_e, i.e.

$$\tilde{\tilde{f}}_S(u,v,\xi) = \int f(x,y,z_c)\exp\{j[2\pi(xu+yv+z_c\xi)]\}dxdydz_c \qquad \text{(Eq. 4.23)}$$

For the ECPSF the three dimensional transfer function is simple given:

$$\tilde{\tilde{f}}_2(u,v,\xi) = FT_{z_c}\left[\frac{-j}{\kappa_o}\exp\left(j\frac{\pi {z_c}^2\rho^2}{\kappa_o}\right)\right] = \frac{-(1+j)}{(2\kappa_o)^{1/2}\rho^2}\exp\left(j\frac{\pi\kappa_o\xi^2}{\rho^2}\right) \qquad \text{(Eq. 4.24)}$$

and for the CPFS the three-dimensional transfer function it is given by:

$$\tilde{\tilde{h}}_2(u,v,\xi) = FT_{z_c}\left[\tilde{h}_2(u,v,z_c)\right] \qquad \text{(Eq. 4.25)}$$

Unlike in the case of coherent holography, the case of the three-dimensional transfer function is not singular and so can be inverted. It is an important issue for three-dimensional numerical reconstruction.

4.2.2. Notational convention

Defining the shorthand *hol* to represent the dependence of the hologram on the phase delay; i.e., we rewrite Eq. 4.18 to be:

$$h_2(M) = hol\left(\frac{\pi\kappa_o r^2}{{z_c}^2}\right) \qquad \text{(Eq. 4.26)}$$

for on-axis conoscopic holograms, $hol = T(\cos / z^2)$; for exponential conoscopic hologram, $hol = (\exp / z^2)$.

4.3. Construction of the conoscopic hologram

Recording and reconstruction have not been differentiated, both being the same physical process. Given a set of three-dimensional emitting point, the conoscopic hologram is constructed by passing the emitted light through the conoscope; the hologram is the incoherent superposition of the elementary quasi-monochromatic conoscopic figures of each point, namely:

$$H(Q)=\sum_{allpo\,\mathrm{int}\,sP} I(P)hol\left[j\pi\kappa_o \frac{(x-x')^2+(y-y')^2}{(z_c-z')^2} \right] \qquad \text{(Eq. 4.27)}$$

where the sum is over all the emitting point *P*.

We also use the continuous version of the conoscopic hologram, given by:

$$H(x',y',z')=\int_V I(x,y,z_c)\times hol\left[j\pi\kappa_o \frac{(x-x')^2+(y-y')^2}{(z_c-z')^2} \right] dxdydz_c \qquad \text{(Eq. 4.28)}$$

The conoscopic hologram can also be represented for each plane, i.e., for each value of z_c, as a two-dimensional convolution:

$$H(x',y',z')=\int_{z_c} I(x,y,z_c)*\left\{ hol\left[j\pi\kappa_o \frac{x^2+y^2}{(z_c-z')^2} \right] Win[X,Y] \right\} dz_c \qquad \text{(Eq. 4.29)}$$

where *Win[X,Y]* represents the windows function in the lateral coordinates *x,y*. It is assumed [66] that the limits are imposed on the transfer and not on the hologram size to keep the shift-invariance properties and to simplify the numerical processing. In most cases it drops the literal windows-function dependence, and it writes simply:

$$H(x',y',z')=\int_{z_c} I(x,y,z_c)* hol\left[j\pi\kappa_o \frac{x^2+y^2}{(z_c-z')^2} \right] dz_c \qquad \text{(Eq. 4.30)}$$

This representation of the hologram is called the convolution representation.

Note the important difference between the *z* dependences of conoscopic and coherent holography. This difference is shown in section before to yield a

nonsingular three-dimensional transfer function. It can be understood intuitively in a simple way.

If the crystal length had been equal to the distance between the emitting and recording points, the term *L/z* would have been unity and the dependence would have been the usual *1/z* dependence. However, when the crystal length is smaller (or bigger) than this distance, the effect is scaled by the ratio *L/z*, which represents the portion of active media in the path of the light.

Note also that either coherent or incoherent light can be used. Conoscopic holography is the incoherent superposition of the elementary conoscopic figure. In the case of coherent light a second term, accounting for the intermodulation between the different voxels, has to be dealt with. However, it is equivalent to the coherent hologram of the object and is represented by much higher frequencies in most case. Moreover, for diffuse object it will be negligible and will appears as speckle. The case of conoscopic hologram whose scale is of the same order as the case of the corresponding coherent hologram is realizable and is of great interest.

4.4. Exponential conoscopic hologram

Several procedure exist for retrieving the analytic signal from the detected signal (phase retrieval). However, it is simpler and more accurate to measure the analytic signal directly when possible. To do this we need to record the quadratic term or the sine term of the exponential hologram.

The procedure that has been developed is the off-axis conoscopic holography, in which a carrier frequency is added.

4.4.1. Off-axis conoscopic holography

It is used a property of linear operator (such as the Hilbert transform) that the application of the linear operator on a convolution is equivalent to that of the linear operator on one of the cofactors of the convolution [67].

This mathematical transformation can be performed by numerical methods. However, great care has to be taken in the choice of the algorithms and windows when accurate results are required, because the hologram is limited in all direction and so is only an approximation to the Fresnel analytic function.

The different practical solutions for recording an off-axis hologram are developed. In Figure 4.3 three basic but different schemes are introduced [68].

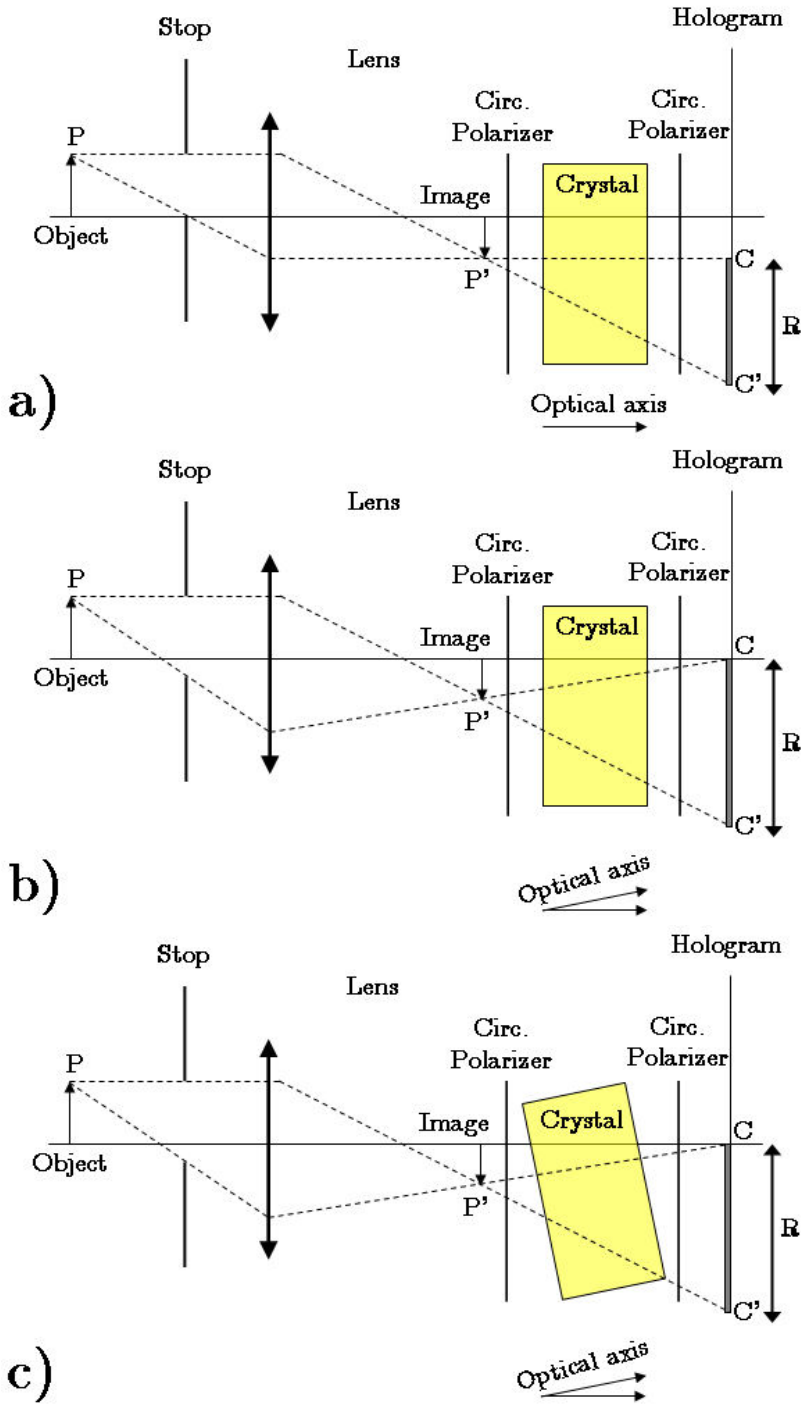


Figure 4.3: Off-axis conoscopic system: a) aperture-displacement system, b) optical-axis-tilted crystal, c) geometrically tilted crystal

They differ only in their aberrations.

4.5. Numerical reconstruction

In this section only the basic equations of the reconstruction are defined.

Given the conoscopic hologram defined in Eq. 4.27, the reconstruction process consists of finding the intensity and the position for each point *P*.
The numerical treatment by sampling the object volume in voxels and the recorded pattern in pixel is simplified. We assume that all the emitting points are contained in a given volume *V*, so that the reconstruction will provide the intensity emitted by each voxel.

Any additional information will strongly influence the retrieval algorithm. For example, for the conoscopic range finder, for which it is known that only one point is emitting, a specific algorithm has been developed. For a system involving a projected grid of point on the object another type of algorithm is developed. For measurement of deformations, interferometric coherent hologram algorithms are adapted.

4.5.1. Surface solution

A special case is the surface case, the most common case in imaging, for which ony one voxel in each full depth line is emitted. Then the position information can be represented by the unambiguous function zc(x,y) through Eq A.8, and the hologram, if it is planar, can be expressed as:

$$H(x',y') = \int_S \frac{I(x,y)}{z_c(x,y)^2} \times hol\left\{ \pi\kappa_o \frac{(x-x')^2 + (y-y')^2}{z_c(x,y)^2} \right\} dxdy \qquad \text{(Eq. 4.31)}$$

Solution based on expansions of $z_c(x,y)$ around a mean value z_0 were published in Refs. [69, 70]

4.5.2. Volume solution

Another approach is to keep the hologram in its three-dimensional form and to apply constrains to the shape of the object in one of the reconstruction process:

$$H(x',y',z') = \int_V \frac{I(x,y,z_c)}{(z_c-z')^2} \times hol\left\{ \left[\pi\kappa_o \frac{(x-x')^2 + (y-y')^2}{(z_c-z')^2} \right] \right\} dxdydz_c \qquad \text{(Eq. 4.32)}$$

Note that, unlike the three-dimensional transfer function of coherent holography, and owing to the z^2 dependence of conoscopic holography, the three-dimensional transfer function is not singular, and because of this the volume hologram three-dimensional relation Eq.4.32 can be inverted in the three-dimensional domain.

4.6. Resolution and accuracy: optical equivalents

If a point hologram is recorded, two different optical systems may be used to measure the parameters of the hologram and so to retrieve the position of the emitting point.

The first possibility is to illuminate a transparency on which the conoscopic hologram has been written with a coherent parallel and to measure the position of the focus. The Rayleigh criterion gives a fair, slightly conservative, measure of the positioning accuracy. With Δx_r as the Rayleigh lateral accuracy and Δz_r as Rayleigh longitudinal accuracy, we obtain:

$$\Delta x_R = 0.61\frac{R}{F} \qquad \text{(Eq. 4.33)}$$

$$\Delta z_R = 2\frac{R^2}{\lambda F^2} \qquad \text{(Eq. 4.34)}$$

where R is the radius of the conoscopic hologram and F is the number of fringes (see Appendix A).

We refer to this solution as the optical or the crude resolution. The lateral accuracy is roughly equivalent (without the 1.22 coefficient) to the fringe count, i.e., the difference corresponding to a one-fringe change, as calculated in Appendix A Eq. A.18.

The second possibility is to insert the same transparency into a Twyman-Green interferometer. First, with the use of fringe count, for example, the relative locations of the two arms are crudely positioned. The fine positioning is done by phase comparison of the reference elements and the conoscopic hologram. We refer to this as interferometric accuracy is found in Appendix C. However, as a rule of thumb, is not clear that the accuracy is limited by the signal-to-noise ratio of the measurement, and so detectable longitudinal difference will be roughly

proportional to the inverse of the signal-to-noise ratio measured in number of fringes, the factor being determined in Appendix C.

4.7. Temporal and spatial coherence and monochromaticy

The temporal delay between the ordinary and extraordinary ray is given in radians, by definition, by the phase delay, so that the maximum temporal difference in wavelength is simply given by:

$$\Delta t = F\lambda / 2 \qquad \text{(Eq. 4.35)}$$

Under the simplifying assumptions the ordinary and the extraordinary ray propagate along the same path and so are completely spatially coherent with each other if we assume that they are derived from the same beam. However, taking the walk-off into account, we obtain the angle differences as function of the lateral position:

$$\Delta\beta = n_o F\lambda / R \qquad \text{(Eq. 4.36)}$$

The monochromaticity, if we neglect the dispersion of the refractive indices, is simply given by:

$$\frac{\Delta\lambda}{\lambda} = \frac{1}{F} \qquad \text{(Eq. 4.37)}$$

Considering the dispersion may slightly correct the needed monochromaticity (this effect is not noticeable in calcite)

4.8. Conoscopic system

In this second part of the chapter, conoscopic system functional description, pointing out its distinctive attributes compared to other non-contact optical sensors is presented.

The conoscope is an optical-element build-up of standard optical components only, such as polarizer, wave plates, and crystal. Its alignment is straightforward and not critical, and it avoids the strong aberration of some incoherent holography systems, which are usable only in a narrow field close to the geometrical axis of the system. It is computer compatible, through its compatibility to television camera. But equally important is its compatibility with standard optical elements.

The basic limitation of all incoherent holographic systems, including conoscopic holography, are the strong requirement on the dynamic range of the detector because of the bias level. Today, with the emergence of the CCD as a mature and well-characterized technology, it would appear possible, by carefully choosing the detector and by tailoring the optical system to the detector performances, to design a complete three-dimensional camera.

The conoscopic technique has some of the drawbacks of incoherent light holography, it also possesses many advantages over classical holography:

- suitable for harsh environment;
- fringe spacing adapted to common CCD sensors pixel size;
- weak monochromaticity requirement; in practice, a 10nm spectral bandwidth is required permitting the use of laser diodes, superluminescent diodes and even LED.

In addition conoscopic systems offer specific benefits:

- simplicity in the set-up;
- increased stability of the set-up due to a common geometrical path of the two waves along the entire optical path;
- complementary interferograms recording, by recording separately the two polarizations, with no moving parts.

In comparison, a triangulation system measures the angle of a single ray, while the conoscopic system measures the angle of each ray in a complete solid angle. The

system will perform a measurement of 1400 angles and analyze the global behavior of all the rays. This procedure is much more precise, stable and robust, but requires more computations.

The basic conoscopic setup set-up is shown in Figure 4.4. The control box counts the fringes which are detected and imaged by the limited sensitive area of the CCD detector. The PC manages the overall measurement process. A large variety of devices (in speed, scale of measurement, precision, working conditions, matters of samples) can be developed while using the same technical basis.

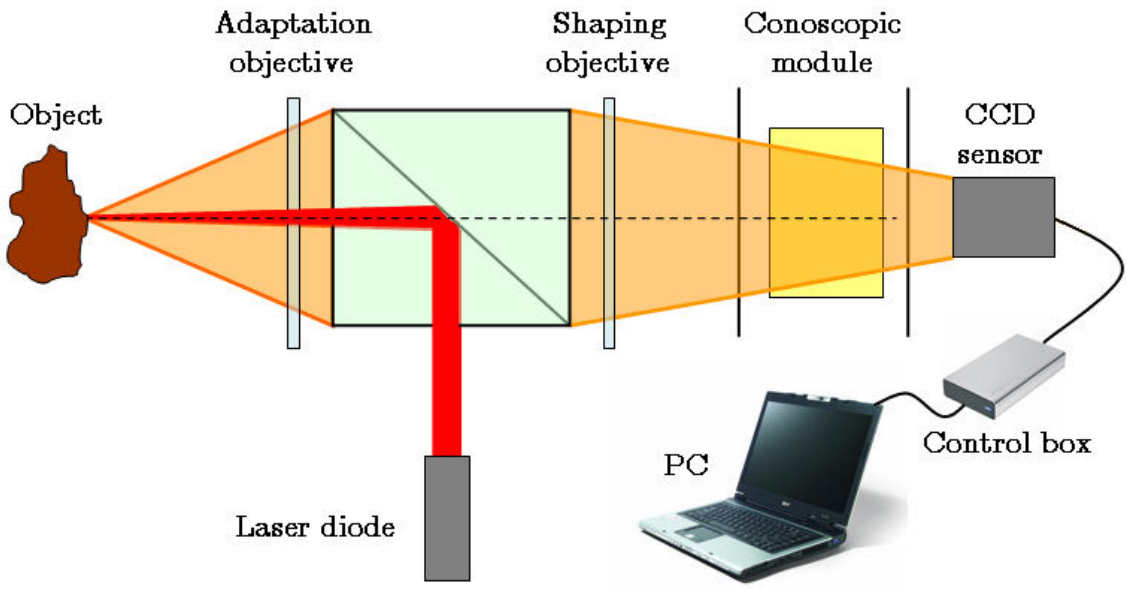


Figure 4.4: Basic conoscopic system set-up.

4.8.1. The distinction of conoscopic systems

Conoscopic systems differ from conventional triangulation systems by:

4.8.1.1. Measuring range

Conoscopic systems maintain high precision and reproducibility over a wide working range. In most cases, the reproducibility is better than 1/8000th of the working range. This is an improvement by a factor of 5 to 10 from standard systems.

4.8.1.2. Surface quasi independence

Conoscopic systems provide good performance while measuring on a wide variety of materials, which could hardly be measured with other non-contact sensors.

Machined metals, other relatively shiny objects and translucent materials as ear surface can be analyzed.

4.8.1.3. Grazing incidence measurement

Conoscopic systems are able to measure at angles very close to grazing incidence. Conoscopic systems measure angles up to 85° from normal incidence on diffusive surfaces from all directions.

4.8.1.4. Multi-use single head

A single conoscopic system can be used to measure ranges starting from hundredths of microns to tens of centimetres by only changing an objective lens.

4.8.1.5. Measurement of complicated structures

The conoscopic system is collinear, thus it has the unique capabilities to measure inside bore holes and complicated geometrical structures using simple optical bending elements.

4.8.2. Finite-point effect

The laser dot projected onto the surface of the object to be measured is not point-like with regard to the theoretical assumptions. This dot can, however, be divided into a continuous set of elementary points creating their own interference pattern (holographic property), thus building a lower contrast interferogram. Thus the subsequent calculation retrieves an “optically averaged distance” for the point. This particular point is useful, but it also affects the precision of the probe. This is intrinsically the case when using laser light as the light source. Laser diodes are used because of the extended depth of field which they yield to the probe thanks to their Gaussian beam characteristics. However, their coherence (much lower than solid-state or gas lasers) causes speckle noise to appear on the interferometric patterns. At very small scale (below the length of coherence and the surface roughness, ≈ 100 μm), elementary points of the laser dot can no longer be considered as independent. The “contrast” calculation process loses its ability to remove optical noise. The study of optical and algorithmic solutions to reach the full resolution capabilities of the conoscopic system is an important challenge for the conoscopic system manufacturer.

4.8.2.1. Possibilities in difficult cases of light–matter interaction

As a direct consequence of the way they handle finite points, conoscopic system present advantages in some situations which are tackled poorly by other techniques.

4.8.2.2. Matter transition.

The finite-point effect is exaggerated in internally diffusing and translucent materials. The distance can be measured "under" the surface. However, this error can be used for quality inspection where a visual solution gives no clear information. The conoscopic probe thus acts as a matter transition detector and can be sensitive to sub-surface defects.

4.8.2.3. Albedo transition.

On a black–grey–white transition, each elementary hologram is issued from the same distance and thus appears as their addition. The calculation yields the same distance as if there were no albedo changes. So the conoscopic probe shows no sensitivity to albedo changes.

4.8.2.4. Relief transition, steep slope.

When a surface relief transition or a steep slope is encountered, the probe gives a value near the intensity-weighted centre of the dot. In this case there is no finite-point effect because the point does not extend laterally, which better preserves the contrast of the interferogram (the amount of light, however, may go down dramatically). This quality associated with the minimal angular aperture needed makes the measurement process never stop until almost 90° slopes. This can find application in cylinder diameter evaluation. In contrast to shadow-casting techniques, it does not need to use two devices on either side of the piece to inspect it. The precision is around one-fifth of the projected light beam diameter.

4.8.3. Example devices

From 1988 to 1994, conoscopic probes have been proposed in many prototypes working from submicron resolutions (0.04 μm) to the order of metres (1 cm at 3 m).

The use of standard industrial components, such as laser diodes as used in CD players, a linear CCD array as used in a bar code reader has allowed the manufacturing of devices in large runs at very competitive prices. A standard optical probe has recently been successfully produced by the Optimet company, attaining state-of-the-art performance. Its performance is as follows: 4000 measuring points over a range of 50mm for a mean working distance of 100 mm at

a 1 kHz rate. By just replacing a single lens on the front face, the working distance and the depth of field is easily adjustable (Table 4.1). This conoscopic system can be completed by an on-axis colour TV camera, making full use of the collinearity property of the conoscopic probe.

The first sensor developed for industrial purposes was the ConoProbe (Figure 4.5a), which was made available commercially in early 1999. The ConoProbe is a high precision, non-contact point sensor. It is used today on many industrial applications, ranging from quality control of fine mechanics as turbine blades, gears, machine tools, molds and soft materials, to in-process inspection of thickness, shape and angles, and copy and milling on CNC machines. The commercial success of the ConoProbe is due to a combination of technical performances that distinguish it from other three dimensional point sensors. Recently the Smart ConoProbe (Figure 4.5b), which is the new version of the ConoProbe, was introduced. The weight and the volume have been reduced by a factor of 2. Its dimensions are small; therefore, it is distinguished by its small volume (55 x 85 x 65 mm) and weight (370 g). The performances in terms of the sampling rate have been improved by at least 50 – 70 %, reaching the 3 KHz mark instead of 850 Hz. It should be pointed out that all the controls and computational electronics are implemented and integrated inside this volume.

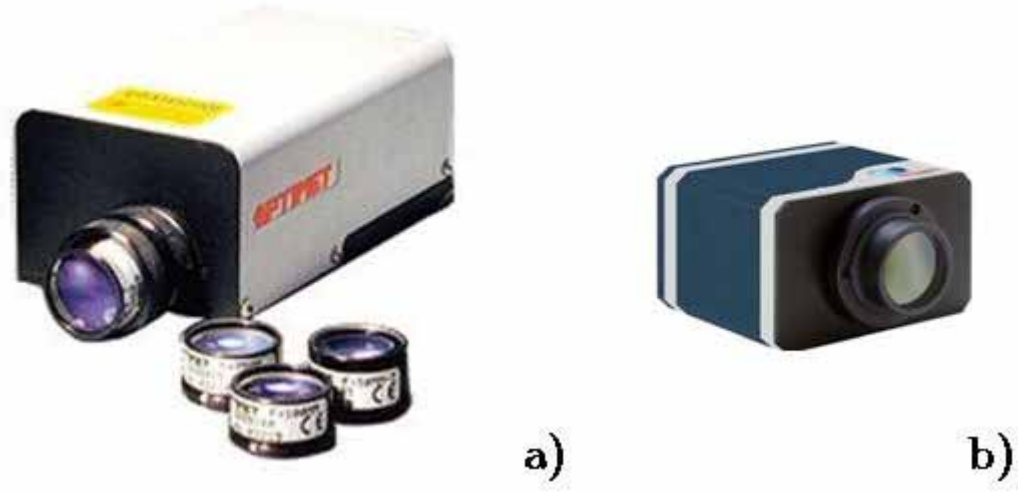


Figure 4.5: Commercial version of conoscopic system Optimet company a) ConoProbe and b) The Smart ConoProbe

The performance characteristics of the ConoProbe are summarized in the following Table 4.1:

Lens Assembly Type (by Focal Length in mm)							
	Standard					High Definition	
ConoProbe Technical Specifications	25	50	75	100	200	25	50
Precision (μm)	<1	<3	<6	<12	<40	<0.5	<1
Reproducibility 2δ (μm)	<0.3	<0.8	<1.5	<3.0	< 5	<0.2	<0.3
Working Range (mm)	1.8	8.0	18	36	125	0.6	1.8
Standoff (mm)	15	45	70	90	190	15	45
Laser Spot Size (μm)	20	40	65	100	200	8	20

Table 4.1: Performance characteristics of the ConoProbe commercial version.

5 Otoscan3D

Otoscan3D is shown in Figure 5.1. To the best of our knowledge, Otoscan3D is the first ever 3D laser scanner based on surface silicon micromachining techniques able to perform the direct scanning of the external auditory canal [71]. The 3D digital scanner proposed is composed by: the scanning probe comprising of MEMS micromirror integrated with electromagnet entering the inner part of the auditory canal, the conoscopic system arranged with the scanning probe, the motion and acquisition control system comprising of high precision hexapod robot having six degree of freedom, translational and rotational stages and 3D pattern reconstruction software for the elaboration of the signals to obtain a CAD format image of the external auditory canal.

Figure 5.1: prototype Otoscan3D: 3D laser scanner based on surface silicon micromachining techniques for shape and size reconstruction of the human ear canal.

All the parts of the 3D scanner laser and their integration are presented in detail in this chapter.

5.1. MEMS scanning probe and conoscopic system integration

For the prototype development, ConoProbe conoscopic system by Optimet has been chosen as noncontact single-point measuring sensor. The conoscopic holography technology is well-suited to the optical properties of the ear surface skin. When illuminate with a 633 nm monochromatic light, each point of the ear emitting (diffusing or reflecting) incoherent, unpolarized, quasi-monochromatic light. The spatial resolution required for the hearing aid fabrication is more than 100 µm, easy reachable with the conoscopic system. As shown in Table 4.1, the collection lens of the conoscopic system permits to choose the best compromise between accuracy and working range. The evaluated working range for the external auditory canal scanning is 30 mm. For a 100 mm lens assembly focal length, the calculated working range is 36 mm with a absolute accuracy < 15 µm and 90 mm standoff (distance between the lens and the middle of the working range) adequate for the ear geometry. The length of the MEMS micromirror scanning probe and the position of the micromirror are optimized in accordance with the standoff. The MEMS scanning probe is mounted on a rotational motor in front of the objective lens of the conoscopic system (by coinciding the centre plane of the micromirror with the optical axis of the detection laser beam). This has essentially ensured the maintaining the fixed distance between micromirror and the lens.

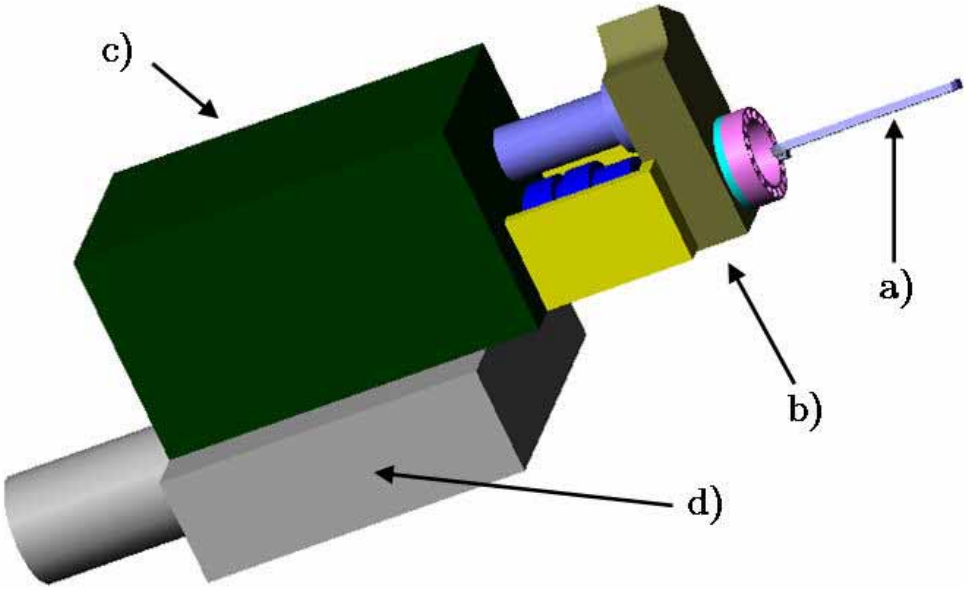


Figure 5.2: MEMS scanning probe and conoscopic system integration design project: a) MEMS micromirror scanning probe; b) rotational stage mounted on the reference plane of the conoscopic system; c) ConoProbe with 100 mm lens; d) translational stage for movement along the optical axis.

The first version of Otoscan3D prototype is developed following the design project in Figure 5.2. The MEMS micromirror scanning probe (Figure 5.2a) is mounded on the rotational stage (Figure 5.2b) and aligned to the optical axis of the conoscopic system. This configuration has the advantage that the rotational mass is limited only to the MEMS probe weight. The diameter of rotational motor mechanical aperture is increased avoiding measurement distortions. A version of the ConoProbe (Figure 5.3), expressly modified for this application, is connected to the rotational stage (Figure 5.2c). two pin-references are added to the frontal part of the ConoProbe for the fine aligned of the mechanical components. The conoscopic system, the rotational motor and the MEMS scanning probe are placed together on a translation stage moved by a linear motor(Figure 5.2d). The linear motor movement are along the optical axis of the system.

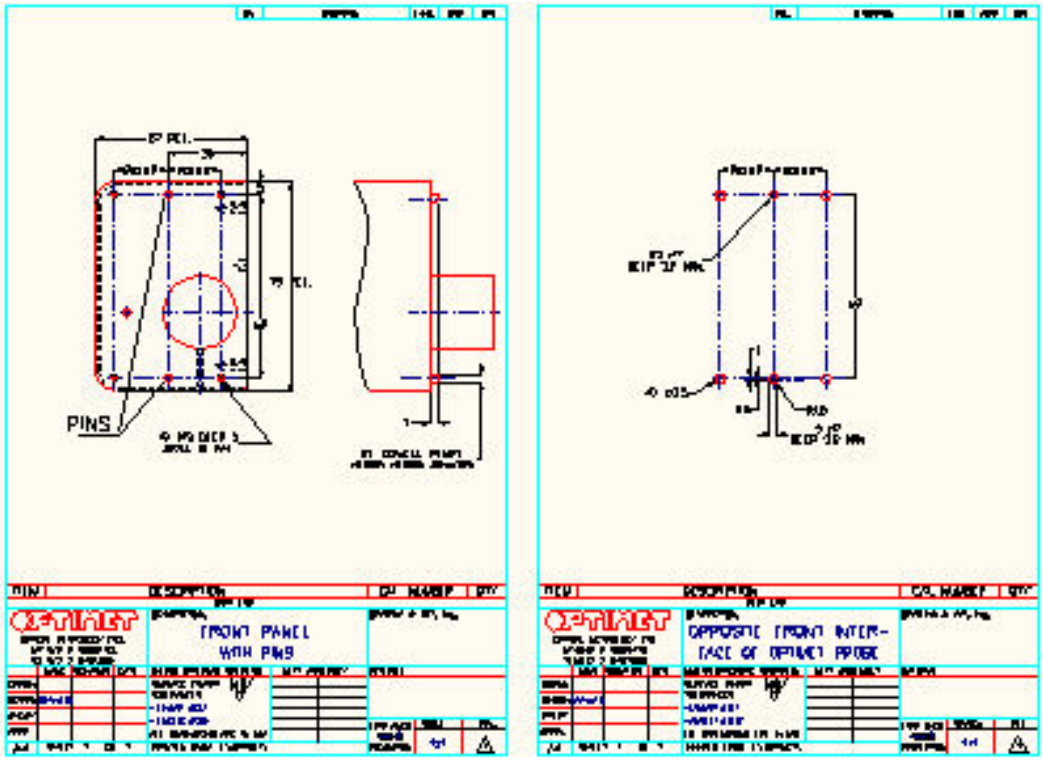


Figure 5.3: Design of the frontal part of the modified ConoProbe with 4 hole and 2 pin-reference.

The first Otoscan3D version presented high nonuniform angular behavior during the measurement calibration. The laser spot of the ConoProbe has an elliptical shape as discussed in Chapter 4. In the first Otoscan3D version the relative position between the MEMS micromirror and the laser spot changes because only the MEMS micromirror scanning probe rotated during the measurement. The flatness of the micromirror is not symmetrical on the x-y axis; in principle for each angle a different flatness correction is needed. The final flatness correction function resulting become a complex function, hard to calculate experimentally. For first Otoscan3D version the micromirror calibration is a long and complex procedure. The final correction function, in any case, is poor fitting of the experimental point calibration that introduced a no-negligible distance measurement errors.

To solve the angular nonlinearity a new mechanical system for the measuring head was developed. In Figure 5.4 the second Otoscan3D version is shown.

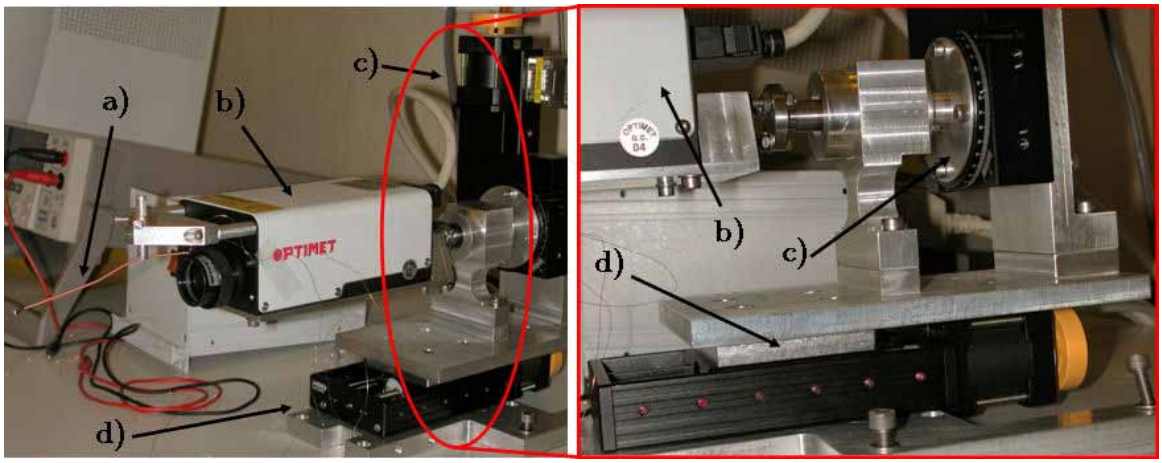


Figure 5.4: second Otoscan3D version: a) MEMS scanning probe; b) conoscopic system; c) rotational motor; d) linear motor. The MEMS micromirror is mounded directly on the reference frontal plane of the ConoProbe and the rotational motor turns the two part together.

The MEMS micromirror is mounded directly on the reference frontal part of the conoscopic system. The rotational motor is placed to the rear and turns the micromirror and the ConoProbe together (monostatic system). Rotational motor, ConoProbe and scanning probe are moved by a linear motor aligned with the optical axis. The advantage of this configuration is that the relative position between the MEMS scanning probe and measure spot coming from the conoscopic system cannot change. The calibration function is angular position independent. It is possible to calibrate the flatness of the micromirror only for a reference angle with fast and precise procedure. The final calibration function is simpler and the distance measurement offers an uniform angular signal to noise ratio (Figure 5.5).

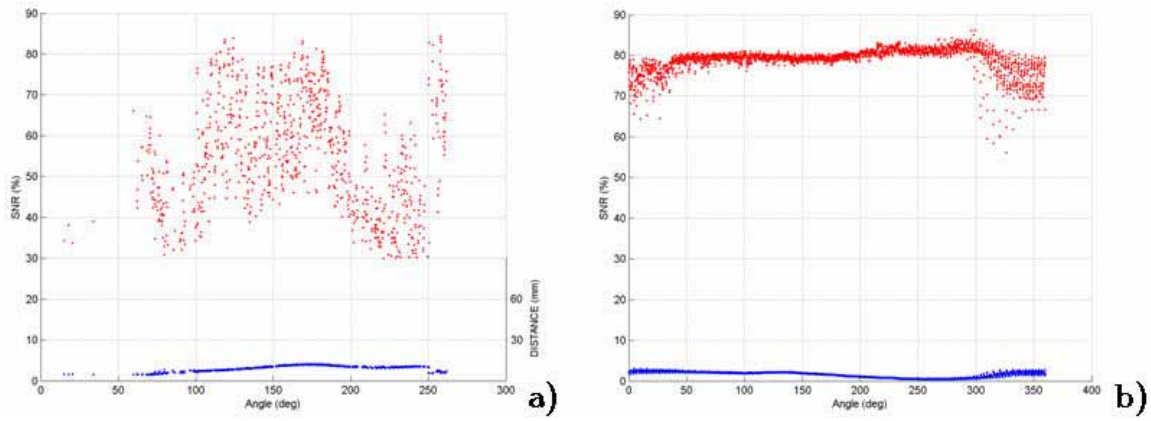


Figure 5.5: Comparison between signal to noise ratio (red point cloud): a) first Otoscan3D version with relative movement between MEMS scanning probe and Conoprobe; b) new Otoscan3D system (monostatic system).

5.2. Motion system

For the developing of the Otoscan3D motion system, anatomical consideration has been done. The best entrance angle for a direct scanning of a human ear canal can differ more than 30 degree from patient to patient. The mechanical stiffness of the motion system should be very high in order to avoid dangerous drift that can generated distance measurement errors or risks to touch the patient ear surface. The required positioning precision should be higher than the mechanical resolution required for the hearing aids fabrication. Motion system should be able to store and recover the instant position of the MEMS scanning probe in the 3D coordinates. The last feature gives the possibility to join two different scan of the same human ear canal for the complex geometry cases. A motion system that satisfy the requirements is the six degrees parallel robot (PAROS from Micos). The whole motion system is composed by the PAROS robot, the rotational motor and the linear motor. Hardware triggers are also implemented to control and synchronized the motion system movements to the flipping of the micromirror and the acquisition frequency of the conoscopic system.

5.2.1. PArallel Robot Operating System

PAROS is the abbreviation of PArallel Robot Operating System (Figure 5.6a). With one hexapod all six degrees of freedom can be moved without additional positioning elements. The low weight of the moving platform allows high dynamic positioning processes. The six struts of the PAROS, using a special designed joints, ensure extremely high stiffness, low backlash and accurate positioning. The travel ranges of the individual coordinates are interdependent. All six struts change their length even if motion in only one axis is required. The travel limits are maximum value where at least one strut is fully extended. For multi-axis motion, the travel range is shown in Figure 5.6b.

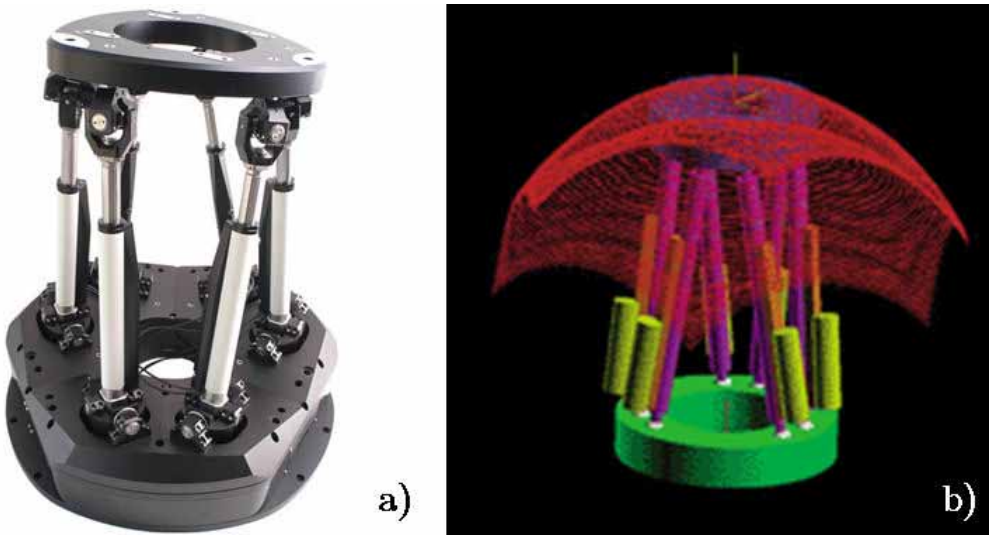


Figure 5.6: a) PArallel Robot Operating System(PAROS); b) travel limit of the six degrees parallel robot.

Especially for high resolution and high forces the robot can be equipped with a 1 mm pitch and a gear motor combination. A resolution up to 1 µm can be achieved.
Special software is available to determinate the system travel capabilities at any point in space and to modify the pivot point of the motion system. In the Otoscan3D prototype the pivot point is set on the MEMS micromirror position; in this way the orientation of the micromirror in the EAC can be changed without changing its position. This strategy permits to avoid the touching of the ear surface and increases the safety of the whole system.

5.2.2. Translation and rotation stage

The translation stage used is shown in Figure 5.7a. A backlash compensated lead screw produces a quiet smooth move. The translation stage is direct drives without gear and is equipped with integrated limit switches. The travel range is up to 300 mm with a speed up to 20 mm/s. The maximum load is up to 5 kg; the measure head of the system comprising the mechanical parts weights less than 3 kg.

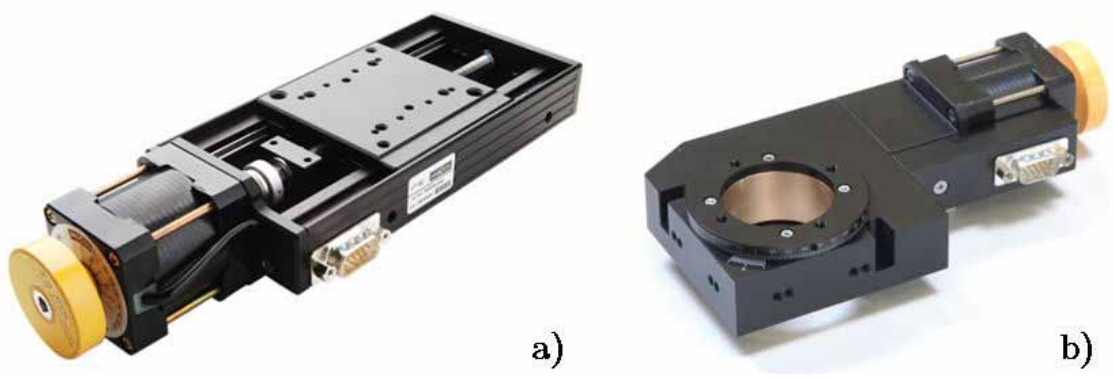


Figure 5.7: a) translation stage and b) rotation stage used for Otoscan3D prototype

The rotation stages used is shown in Figure 5.7b. The continuous aperture of 40 mm diameter is suitable for many applications in the range of microscopy. The screw worm gear wheel combination is pre-stressed with zero backlash and produces a smooth move. Motorized rotation stage is equipped with a mechanical reference limit switch. The rotation range is 360 degrees endless with a speed up to 45 degrees/s.

5.3. Software control

A homemade software is dedicated to the control and the synchronization of the different components of the Otoscan3D prototype (Figure 5.8). The software control is developed using Visual Basic language and tool for building applications for Microsoft Windows. The control software manages the scanning procedure, ensures the motor movements, synchronizes the micromirror flipping to the ConoProbe frequency acquisition and provides the storing of the 3D data points and the 3D point-cloud surface reconstruction.

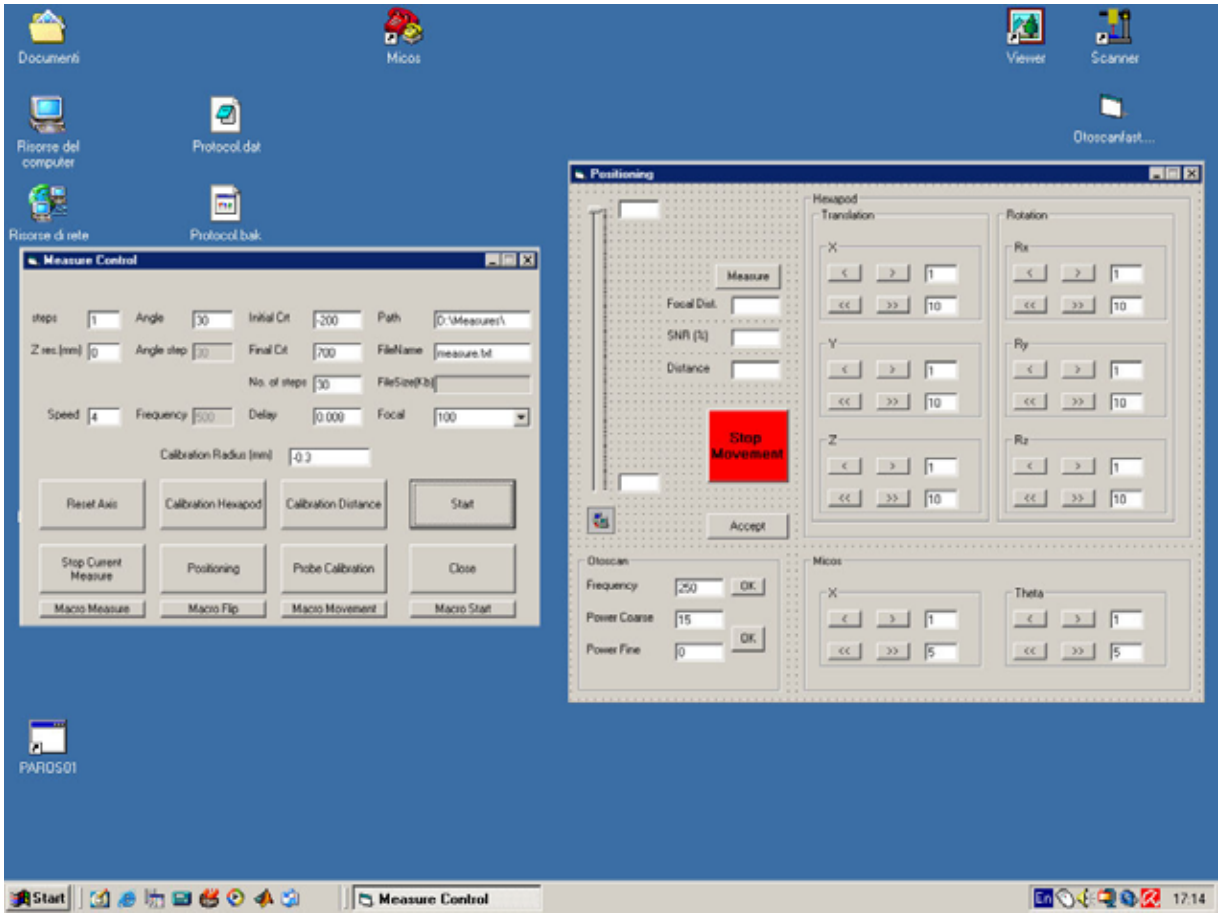


Figure 5.8: Screenshot of the Otosca3D program

A schematic description of the program functions is presented below.
The Otoscan3D program is structured into 3 main parts:

- The interface
- The modules
- The class modules

5.3.1. The interface

All needed parameters can be set. This window is divided in 4 parts. The first part is the one where the height data (number of steps and the dimension of one step) are introduced. In the second part, the angle data are introduced. Due to the fact that these data are correlated with the sampling rate, the dimension of the step angle is calculated accordingly with the sample frequency and motor speed. In the same time, this is not a limitation of the measure because the minimum step is

about 0.05 degrees. The user can choose the number of step sizes to be taken into consideration. The third part is the export file part. Here, the user chooses the path and the name of the exported file. To be noted that this file is a text file with 5 columns, representing: the height, the angle, the radius, the signal-to-noise ratio and total numbers. In the fourth part the user can see the sample frequency, can set the motor speed, the focal length and the calibration radius.
The interface contains the following functions:

- Reset axis button: sets the z-axis of the rotational motor to zero using the "cal" command and disabling the second axis;
- Stop current measure button: sets the StopFlag to true so that no command is transmitted to the motors anymore;
- Calibration button: makes a calibration measure for determining the distance between the objective and the mirror. For this, a calibrated cylinder is needed. The function makes a full 360 degrees measure of the cylinder, calculating an average of the measures and subtracting the known radius. The value obtained is saved in a file named CalLng.txt for further use;
- Probe calibration button: calls the "Probe Dialog Window" after reading the necessary parameters. The Probe Dialog Window is defined in the provided motor software and can be partially modified;
- Close button: closes the program;
- Start button:
  - Reads the value stored in CalLng.txt file;
  - Sets the values of frequency, velocity and acceleration for the probe and motors;
  - Moves the rotational stage so that the beginning of the sampling to be at the desired point. The sampling is uniform so, the first part of the movement, where the motor accelerates, is useless;
  - Sets the flags to the initial values;
  - Calculates the number of measures for each cycle so that not to overwrite them;
  - Begins the measurement process:
    - Starts the rotational motor;
    - Begins read the distances and stores them in the buffer;
    - After the motor is stopped, reads the measures and writes them in the specified file;
    - Moves the translational motor;
    - Restarts the rotational motor;

- Repeats the measurement process until the translational motor does the whole movement, except for the case in which the "Stop current measure" button is pressed, fact that leads to a change in the state of the StopFlag and finishes the measure.
- The ParityFlag determines the direction of the rotational stage. The rotational stage moves clockwise and counterclockwise depending of its value. This is made for giving the possibility to measure only a portion of the sample.

5.3.2. The modules

- Callback module (name of file: callback2.bas). Modifies the probe dialog by adding mechanical stage controls.
- FIFO module (name of file: fifo2.bas). Loads various functions needed by the probe. Provided by Optimet.

5.3.3. The class modules

- ConoControl (name of file: conocontrol2.cls). Contains various functions for the probe:
  - ConoInit: Initializes the probe.
  - ConoClose: Closes the probe.
  - ConoSettaPotenza: Sets the desired power of the probe
  - ConoSettaFrequenza: Sets the desired frequency
  - ConoSettaLente: Sets the desired lens
  - To be noted that the functions are called by the interface program

- MotorControl (name of file: motorcontrol2.cls). Contains functions for motor control:
  - MicosInit: Initializes the motors
  - MuoviAssi: Moves the motors
  - DisAsse2: Disables the rotational stage (used for the reset axis button, due to the fact that the rotational stage doesn't have an end of stage point.)
  - EnAsse2: Enables the rotational stage.
  - Calibra: Moves the axis until it reaches the "cal" switch (end of stage).
  - SetVel: Sets the desired velocity.
  - SetAccel: Sets the desired acceleration

- PosAsse1: Returns the actual position of the axis. (Not used in this version)
- Stato: Returns the state of the stages (0- all commands are finished, 1- commands are now executing)
- MicosClose: Closes the motors.

5.4. Scanning strategy

The scanning probe with the MEMS micromirror is able to enter the inner part of the ear and deflect the measure laser beam coming from the conoscopic system to the ear surface.

With a complete rotation along the optical axis of the system conoscope-scanning probe, a circular scanning path is achieved. Combining the rotation and the flipping of the micromirror, a cylindrical scanning surface suited for the EAC geometry is obtained.

As described in the next chapter, the external auditory canal has two bends. For the fabrication of the CIC hearing aid, the metrological information must be reach at least the second bend (the last before the eardrum).

The rest position of the micromirror allows a 130° measure beam reflection angle. As shown in Figure 5.9, the rest position of the micromirror permits to scan portions of the ear surface some millimeters above the MEMS scanning probe tip. During the EAC scan the second bend is reachable placing the micromirror between the first and second bend, in a safe and accessible position. On the contrary, increasing the rest angle of the micromirror, the shadow areas increases as shown in Figure 5.9a. Furthermore, with the micromirror flipping, the shadow parts of the surface are also highlighted (Figure 5.9b).

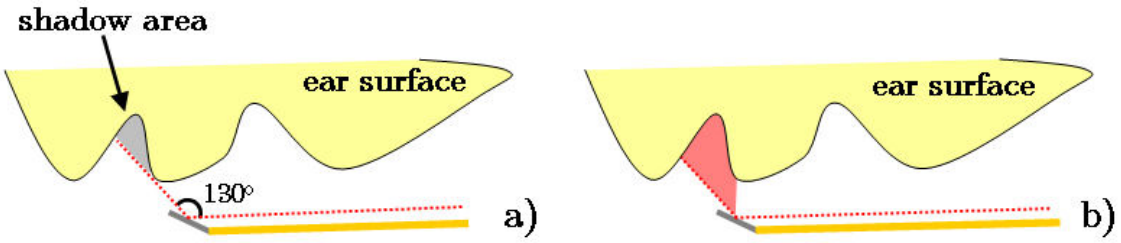


Figure 5.9: The rest position of the micromirror allows to scan portions of the ear surface some millimeters above the MEMS scanning probe tip. a) Without the micromirror scanning this creates shadow area; b) with the micromirror flipping, the shadow parts of the surface are also highlighted

The preliminary phase of the ear scanning involves the positioning of the MEMS probe after the first bend. The positioning is performed driving the MEMS scanning probe in the EAC using the PAROS robot.
In many case the best EAC entrance angle has obstructed by the tragus (a small pointed eminence). For this reason an additional mechanical element (Figure 5.10) is fabricated to displace the tragus.

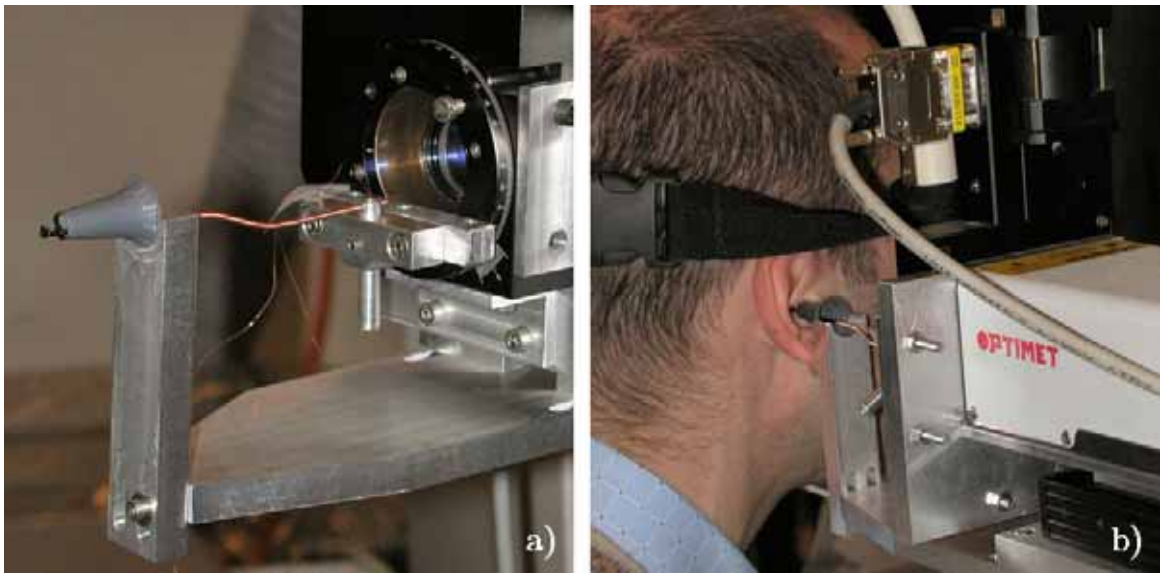


Figure 5.10: The mechanical system for the tragus opening: a) the MEMS scanning probe is free to move inside the otoscope cone; b) tragus opening during an in-vivo measurement

The additional element is based on a cone (the same used in a optical otoscope system) that can be changed after the in-vivo scan. The cone displaces the tragus and provides a geometrical reference for the positioning of the probe. During the positioning phase it has also a micromirror protection. The PAROS robot moves the tragus cone and the scanning probe together, the translational stage is dedicated for the relative movement between the tragus cone and the scanning probe. The displacement of the tragus does not change the shape of the EAC.

The 3D measurement starts with the mirror flipping and the rotation of the monostatic system probe-conoscope to which the rotational motor is dedicated. After a turn of the scanning probe, the system withdraws through the translation stage and restarts the rotation; five-six steps, each 4 mm long, are enough for a normal length EAC to cover all of the cavity.

EAC scanning is completed in less than 2 minutes. The software provides the 3D point-cloud surface reconstruction: Knowing the distance between the ear surface and the micromirror (from the conoscopic measurement) and the instant

position of the motion system, it is possible to place each acquired point in the 3D space. The final map is a point-cloud reconstruction of the ear auditory canal.
The maximum acquisition frequency of the conoscopic system is 870 Hz; working with an oscillation frequency of the micromirror 1/80 of the acquisition system and completing a turn in 30 seconds, a 30 µm final spatial resolution of a mesh point cloud is obtained. The required resolution for the fabrication of the hearing aid using the rapid prototyping techniques is less than 100µm. Degradation experiments performed over 2 weeks show that the performance of the mirror do not degrade.

5.5. 3D map reconstruction

At the end of the scanning and registration procedures, a set of 3D coordinates of the different points, $d_i = \{x_i, y_i, z_i\} \in R^3$ is obtained as shown in fig. 5.11. The dataset $D = \{d_1, \ , \ d_n\}$ consists of completely unstructured data. The ear surface is reconstructed by means of a triangular mesh model *M*. The triangular mesh *M* approximating the ear surface is obtained through an adaptive deformation of a geometric deformable model.

The approach used is based on the deformation of a geometric model minimizes an error function. At each iteration, the geometric model is deformed and re-sampled in a way that locally minimizes an error function, which gives a quantitative estimate of the accuracy of the reconstruction. The error function is used to obtain a nonuniform triangular mesh *M* characterized by a greater density of triangles in the higher spatial frequency regions. The algorithm is developed for the reconstruction of closed surfaces of genus zero, as is the case of ear canal impressions. By simply changing the shape of the geometric model ,the method can be generalized for use with open surfaces. The reconstruction is performed in three phases.

5.5.1. Coarse approximation of the ear surface

During the first phase, a regular geometric model is put inside the volume defined by the range data *D*. The geometrical model is then expanded iteratively until all its vertices reach the boundary of the surface in a way that any further expansion would bring its vertices outside the range data. At each iteration, the expansion is obtained by moving each vertex of a fixed amount along the vertex normal. Once the geometrical model has reached its maximum expansion, it is uniformly re-

sampled. This procedure gives at the end of the first phase a mesh M_1 that is a coarse approximation of the target ear surface.

5.5.2. Mesh refinement

During the second phase, the mesh M_1 is locally re-sampled to obtain a mesh M_2 that is richer in details where higher spatial frequencies are present. This non-uniform re-sampling is obtained by introducing new vertices in those triangles of the mesh M_1 having a relatively higher local approximation error. This error is the distance between the target ear surface and the surface represented by the mesh M_1. Local re-sampling takes place for those triangle which distance from the surface exceeds a predefined threshold value.

The re-sampling is considered valid if the variation of the local error is greater than a threshold value and the re-sampling produces a local significant improvement of the accuracy of the reconstruction. At each local re-sampling step, the resulting mesh is re-triangulated to maintain a correct connection. The refinement process of M_1 is stopped when there are no more triangles to be re-sampled. The result of the refinement phase is a non-uniform triangular mesh M_2

5.5.3. Mesh smoothing

To further improve the quality of the reconstruction, the mesh M_2 is filtered with the iterative low-pass filter proposed by Taubin [72]. In contrast to the classical Gaussian filter, Taubin's filter does not produce shrinkage of the mesh. While filtering, the vertices are moved without changing the connectivity, the number of vertices and triangles of the mesh. The number of iterations and the values of parameters of the filter affect the bandwidth of the filter.

5.5.4. Reconstruction of the external auditory canal

The reconstruction algorithm is tested on the 3D data points acquired. In particular, Figure 5.11a shows the set of unorganized range data obtained by in-vivo 3D scanning of the EAC ear canal with Otoscan3D, whereas Figure 5.11b shows the mesh obtained after the expansion and re-sampling of the geometrical model. This mesh is an initial, gross approximation of the original ear surface. The reconstruction of spatial details is achieved at the end of the second phase. The triangular mesh of the final reconstructed surface is obtained after local re-

sampling and smoothing of the previous mesh. As a result of the local refinement procedure, the mesh has a greater density of triangles in those regions with richer spatial details. The final digital surface (STL format) compatible with the rapid prototyping techniques is shown in Figure 5.11c.

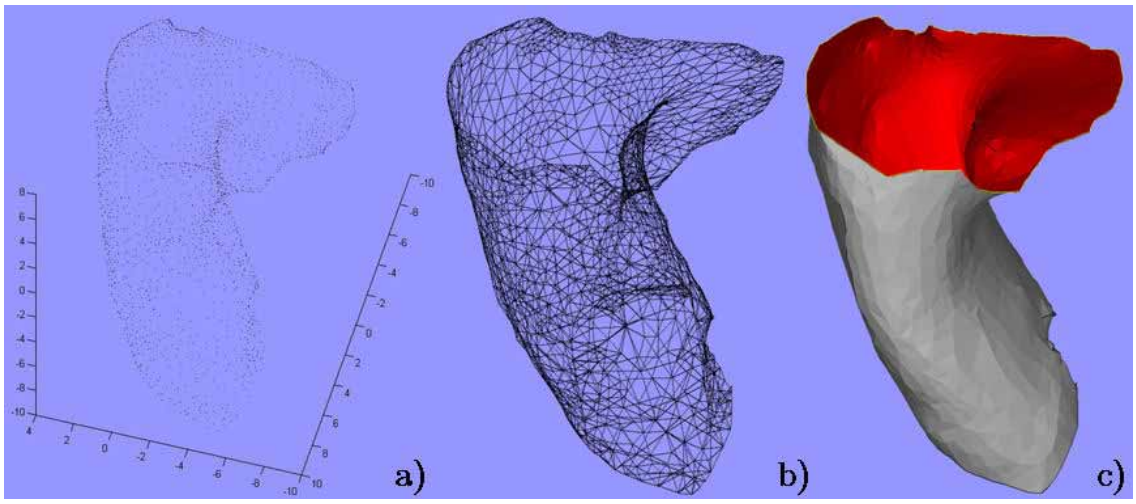


Figure 5.11: Representation of the main phases of the reconstruction algorithm: a) the range data used in this example, as obtained by laser scanning the surface of an ear canal impression; b) the geometric model at its maximal expansion and after three global re-samplings; c) the final reconstructed model after local re-sampling and smoothing.

6 Direct ear measurements

The direct ear canal scanning is the only way to perform accurate and repeatable measurements without the use of a physical ear impression. The aim of this work is to develop and validate a 3D laser scanning system for the direct scanning of EAC. In this chapter the actual scanning of the EAC by a MEMS scanning system that gives qualitative metrological information is demonstrated.

In the chapter beginning the anatomy of the human ear is presented. Following the scanning probe calibration and the in-vitro measurement are shown. The second part of the chapter is dedicated to the in-vivo measurement. The first in-vivo scan of the EAC is presented. Direct scan evaluation and digital comparison between the digital and the standard methods are discussed. Fabrication of two complete hearing aids from in-vivo digital scan and patient feedback are also accomplished.

6.1. Human ear anatomy

There are three sections of the ear: the outer ear, the middle ear, and the inner ear (schematic sketch in Figure 6.1) [73]. The section investigates in the hearing aid fabrication is the outer ear. It consists of the pinna or auricle (the visible projecting portion of the ear), the external acoustic meatus (the outside opening to the ear canal), and the external ear canal that leads to the ear drum. The external ear canal, the violet part in Figure 6.1, is the interest part for the CIC hearing aid fabrication. The outer ear concentrates air vibrations on the ear drum and makes the drum vibrate. The outer ear is also called the external ear.
In front of the concha, and projecting backward over the meatus, is a small pointed eminence, the tragus.

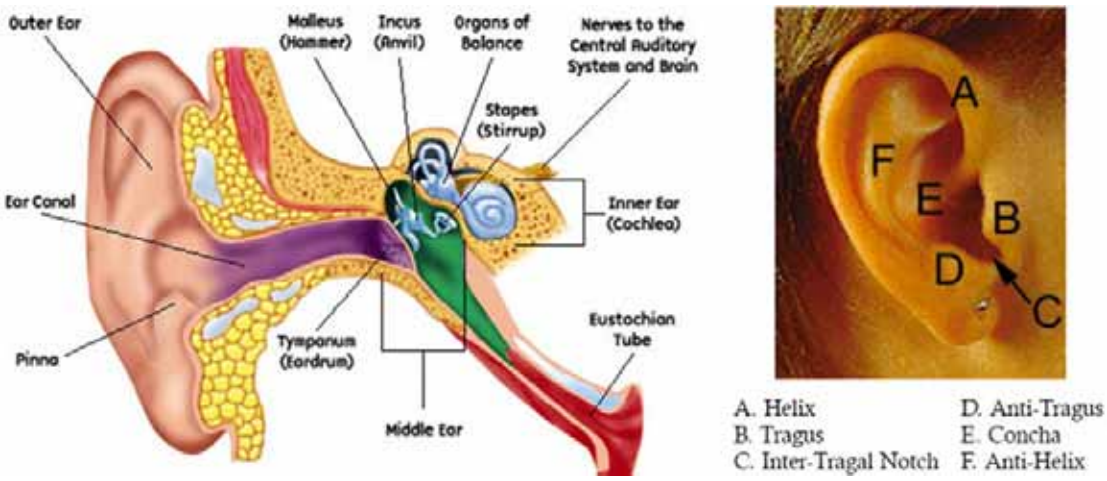


Figure 6.1: Human ear anatomy

The pinna is mostly skin and cartilage with some muscular attachments to the back side (this is what allows us to "wiggle" our ears). The twists and folds of the pinna enhance high frequency (pitched) sounds and also help us to determine where sounds are coming from. Sounds coming from the front and sides become enhanced as they are directed into the ear canal while sounds from behind us are reduced. This helps us to hear what we are looking at while reducing some of the distracting background noise. The CIC hearing aids preserves the pinna acoustic properties, miming the natural auditory process more closely then any other style of hearing aid.

The overall structure of the adult ear canal is a roughly S-shaped elliptical cylinder (2 bends), with a length ranging from 20 to 30 mm, and a diameter of the canal ranging from 2 to 12 mm. The walls of the ear canal are sensitive to touch and there is even a cranial nerve branch that passes just below the canal wall surface. If the ear canal wall is touched near this nerve, many people will react by coughing (Arnold's Reflex). The outer 2/3rds of the ear canal is surrounded by cartilage and contains glands that produce cerumen (earwax), while the inner 1/3 is surrounded by bone. Cerumen has many uses in the ear canal including maintaining a consistent level of humidity, preventing dust and other objects from going deeper.

In order to fabricate a CIC hearing aid, it is important to know the EAC morphology until the second bend. The first bend is the most significant part as far as feedback is concerned. A poor fit here will almost guarantee feedback. If the metrological data fit properly at this point in the ear canal no sound will leak out around the hearing aid.

During the fully manual fabrication process of the hearing aids starting from the ear impression, some geometrical symmetry of the different part of the ear canal are used. In the microphone positioning phase, for instance, the direction of the final part of the canal from the second bend to the ear drum can be deduce from the wall surface direction of the canal before the first bend. Also during the fabrication process, the parts of the ear canal lacking are reconstructed following the behavior of the surrounding parts of the impression. This craft approach based on the experience proves that the actual needed topological information of the ear canal can be reduced.

It seems that for the hearing aids fabrication only some key-point of the ear could be enough.

6.2. Scanning probe calibration.

Before the direct scan of the EAC, Otoscan3D is tested and calibrated on different reference mechanical components with a cylindrical symmetry. The position of an acquired point in the 3D space is a function of the optical distance between the conoscopic system and the measured surface provided by the Conoprobe, the absolute position of the motors (PAROS robot, translational and rotational stage) and the angular position of the micromirror. The optical distance and the motors positions are easier obtained from respectively controls; the unknown parameter for the point positioning is the angular position of the micromirror in function of the current pass through the copper coil generating the electromagnetic field. The Otoscan3D calibration consists in the calculation of the current/angle curve.

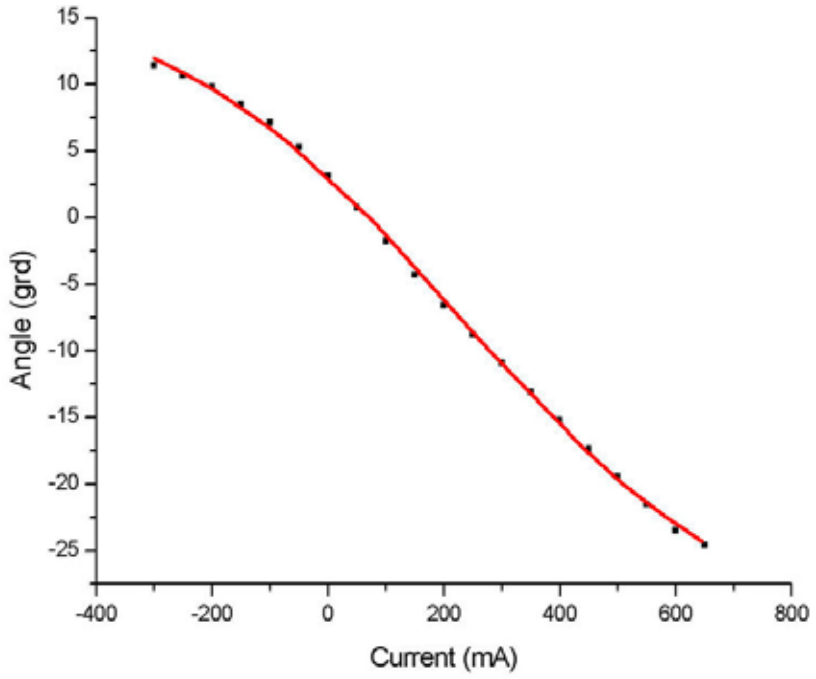


Figure 6.2: Angle/current curve experimental calculated

The angle/current characteristic shown in Figure 6.2 is experimental obtained. The electromagnetic scanning probe is powered with a sinusoidal current 200 mA bias and 800 mA amplitude peak to peak. A Keithley 2420 3A SourceMeter controlled and triggered by the Otoscan3D software provides the output current with 1 mA accuracy. The asymmetry of the driving current signal is due to the rest

position of the micromirror: the micromirror presents a reflective area to the conoscopic system that is zero when the micromirror is parallel to the optical axis, this limits the dynamic range of the micromirror in one angular direction. The oscillation frequency chosen for the micromirror scanning probe was 10.875 Hz which correspond to 1/80 of the ConoProbe acquisition frequency. This frequency is far from the resonance frequency of the micromirror calculated in the Chapter 3 so that the fatigue effect is negligible. During the digital scan the micromirror is keeping in oscillation in order to have a stable dynamic behavior and a linear angular response.

As shown in Figure 6.3a, the angle/current curve has linear response in the middle part but presents nonlinear behavior in the current inversions zone (red zones).

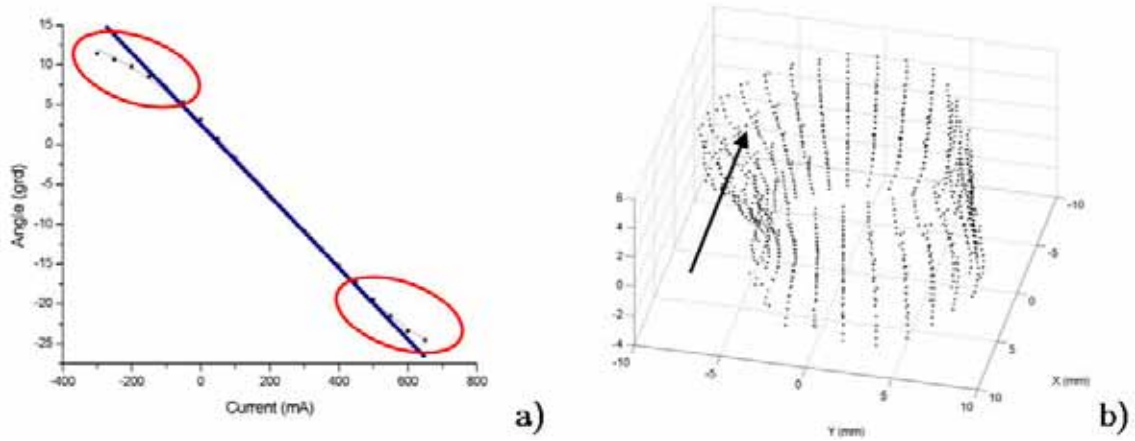


Figure 6.3: a) the angle/current curve presents linear response (blue line) in the middle part and nonlinear behavior during the current inversions (red circles); b) the nonlinearity creates point positioning errors (black arrow) in the jointed part between two translational motor steps.

The poor correction fit in the red circle zone creates positioning errors as shown in Figure 6.3b. The scanning procedure allows the measure overlap between two successive translational steps of the same digital scan (Figure 6.4). Measure overlaps ensure that all the parts of the object are illuminated including shadow zones. In the overlap region drops the points of the two opposite red zones. Due to the combination of the two poor fitting zones the point positioning error is high in the overlap parts as shown in Figure 6.3b.

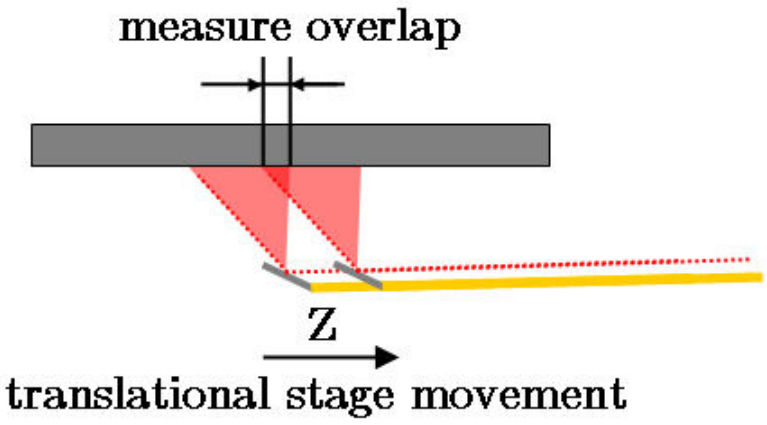


Figure 6.4: Measure overlap after the translational stage movements between the steps of the same digital scan

To solve this problem the points of the red zones are removed during the 3D map reconstruction and only the linear zone points are considered. To ensure the measure overlap and avoid the shadow parts, the length of the translational motor steps is reduced. The final length for the in-vivo measurements is set 4 mm. For narrow canal geometry the translation length can be furthermore reduced changing the value in the dialog box of the control software.

The performance of the Otoscan3D prototype has been tested on several mechanical components with different geometry. The results are shown in Figure 6.5.

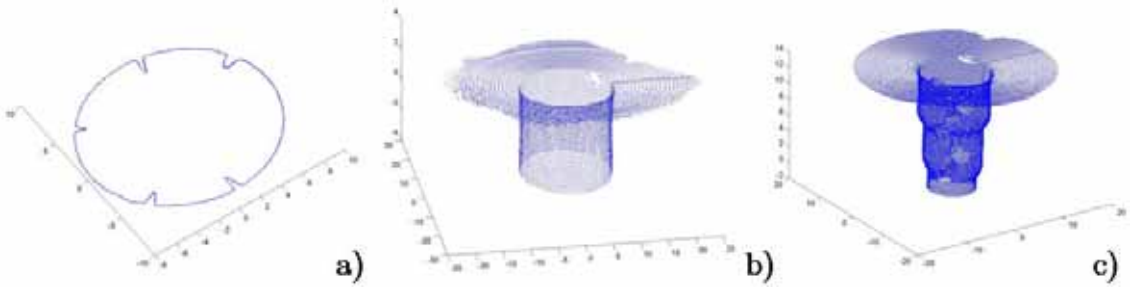


Figure 6.5: 3D scan of mechanical components: a) hole with angular marker for the angular displacement analysis; b) 16 mm diameter hole for the distance calibration; c) multiple diameters hole for the overlap calibration.

Mechanical hole with angular markers (Figure 6.5a) was measured in order to test the angular displacement between different turns of the same digital scan. The distance calibration was proved on the reference mechanical holes shown in Figure

6.5b and Figure 6.5c. On the multiple diameter hole (Figure 6.5c) the evaluation of the measure overlap was also demonstrated.

The data points are accurate to 30 μm in each of the three space directions, allowing ample resolution of surface detail

To understand the best strategy for the in-vivo measurement of the EAC, different 3D digital scans with Otoscan3D are performed on over 50 ear mold provided by Amplifon S.p.A. The ear mold is obtain through a photolithographic process starting from the ear impression. Usually is used for the manual fabrication of the hearing aid shell and is the negative replica of the ear canal (Figure 6.6).

Figure 6.6: the ear mold (on the left), obtained from the ear impression (on the right)

Geometrical consideration on the entrance angle and the tragus position has been done during the measuring test.

6.3. In-vivo measurements

In Figure 6.7 is shown a sequence of the procedure during the in-vivo measurement of the EAC with Otoscan3D prototype.

Before the direct digital scan the EAC, visual examination of the inside of the ears using an optical otoscope is needed. The otoscope analysis is devoted to the earwax and ear hairs inspection. An excess amount of earwax or ear hairs can obstruct the flipping of the micromirror and invalidate the digital measurement. The earwax and ear hair problems are common also for the traditional acquisition.

The preliminary analysis permit to evaluated the EAC dimensions and understand the entrance angle simplifying the positioning procedure. During the fine positioning of the MEMS scanning probe the EAC is not totally visible.

Figure 6.7: Otoscan3D in-vivo measurements: a) patient is immobilized with the help of a frame; b) positioning of the probe and tragus opening; c) scan of the EAC.

During the digital measurements the patient is immobilized with the help of a frame (Figure 6.7a). Figure 6.7b shown the positioning of the mechanical system for the tragus opening. The positioning procedure takes place with the PAROS robot motion. The pivot of the six-degree parallel robot is collocated on the MEMS micromirror, the best entrance angle can be adjusted remaining in front of the external acoustic meatus. Once the entrance angle is reached and the positioning of the tragus cone is finished, the scanning probe micromirror is driven inside the ear canal moving the translational stage in order to reach the deeper position between the first and the second bend. The 3D digital measurement starts with the rotation of the scanning probe and the flipping of the micromirror. After a turn of the scanning probe, the system withdraws and restarts the rotation. EAC scanning is completed in 2 minutes.

At the end of the scanning and registration, a set of 3D coordinates of the different points on the object surface is obtained with the help of dedicated software to reconstruct the 3D anatomical shape and size of ear canal. In Figure 6.8a, reconstructed 3D point cloud image of the EAC of volunteer is shown. In Figure 6.8b is displayed the digitally reconstructed shell of the EAC of volunteer.

We validated our process by fabricating the shell from the actual EAC measurement. The shell fabricated by laser sintering process from the 3D reconstructed image obtained from direct scanning of EAC is shown in Figure 6.8c.

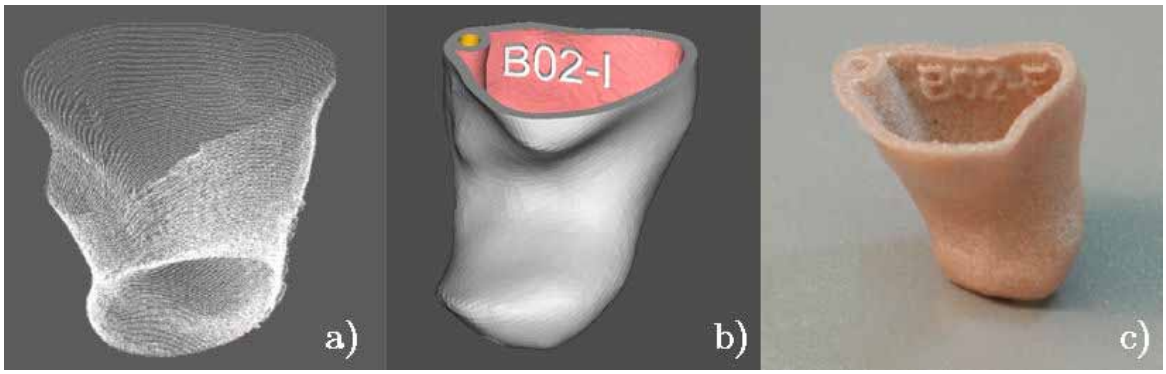


Figure 6.8: a) reconstructed point-cloud image of EAC; b) digitally reconstructed view of EAC and c) laser sintering fabricated shell from the 3D reconstructed digital model from direct scanning of EAC.

6.4. Comparative analysis

In collaboration with Amplifon S.p.A. a comparative test was performed in order to evaluate the real performance of the prototype Otoscan3D. The EAC of 15 individual volunteers and a hypoacusis patient, using a hearing aid for the treatment of hearing loss, have been measured in the TASC laboratory. Three different digital acquisitions of the same ear and two different standard impressions for each volunteer have been taken. Figure 6.9 shown the subjective feedback of the volunteers regarding the acquisition procedure.

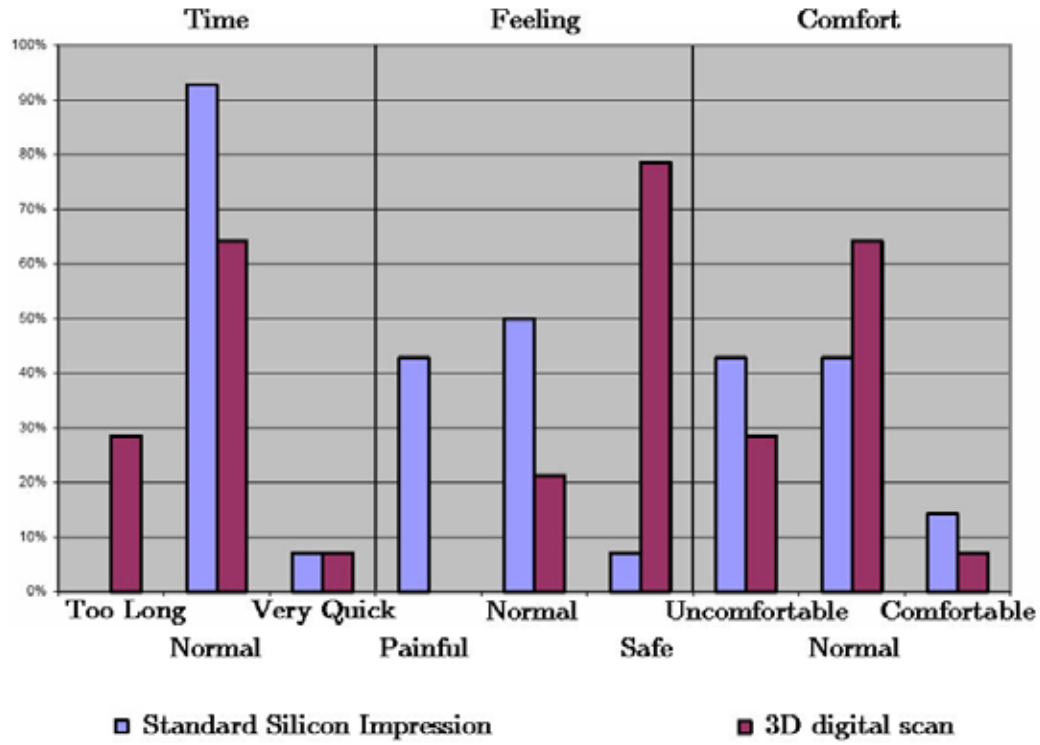


Figure 6.9: Comparative volunteers feedback after the digital and traditional acquisitions.

From the volunteers feedback appears that Otoscan3D feels more safe than the traditional acquisition method. In some cases the traditional method is painful. The measuring time for the 3D scan included the positioning of the scanning probe and the searching of the best entrance angle. For few volunteers the procedure has been repeated many times; this is the reason of some "too long" feedback. From the comfort point of view, the optometrist frame used was the reason of the negative feedbacks.

For the comparative evaluation all the standard impressions have been shipped to the hearing aid fabrication center and a 3D laser scanning of each impression has been performed. The Otoscan3D data has been upload to the hearing aids fabrication center in electronic format ready for the comparative evaluation.

The digital comparison between the two different data set has presented some problems. One of the problem that affect the digital comparison is the scan length. In Figure 6.10 are shown the two different typology of data set.

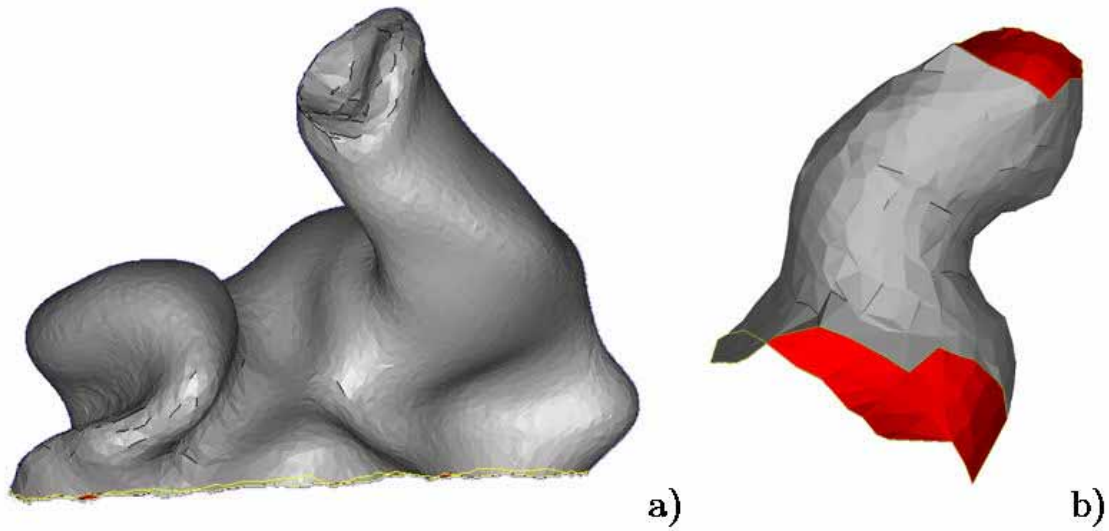


Figure 6.10: Two different digital scan: a) digital scan of the EAC impression; b) reconstructed digital surface from direct scanning of EAC

The 3D digital scan of the impression contains the complete geometrical information of the outer ear instead the direct scan contains only information about the ear canal. From the digital comparison point of view, the outer ear presents many reference point that can permit easily to overlap the digital impression scans. The overlapping of the Otoscan3D data is more complicated.

The edges of the open parts in the direct 3D maps are section of the ear canal. Small changes in the entrance angle between two different digital measurements of the same ear generate difference in the edge shape. The reference points has to be find in the central part of the ear canal scan but is not easy to find reference point in a cylindrical geometry.

The sampling density of the 3D reconstructed point-cloud is nonuniform. The Otoscan3D scanning strategy allows the measuring points overlap in order to avoid shadow regions as discussed in the former part of the chapter. For this strategy, the reconstruction algorithm used for triangulating the 3D impression scan data points has trouble distinguishing the ear canal surface in the overlap regions.

During the in-vivo measurements the MEMS scanning probe can illuminate zones on the ear canal surface with ear hair or earwax. This generated in the final 3D map the spatial noise shown in the Figure 6.11 (red circle).

Figure 6.11: High spatial frequency details due to the ear hairs or earwax.

The noisy points can be eliminated by filtering the high spatial frequencies in the 3D map with a digital filtering process during the surface reconstruction.
The obtained Otoscan3D maps do not present a “tip surface” at all. Therefore the scan is not “shells” but open “cones”. This difference is intrinsic of the measurement principle

The repeatability of a digital in-vivo scanning process is tested with a comparison analysis between different direct acquisitions of the same EAC. Figure 6.12 shows the comparison of two digitally reconstructed surface.

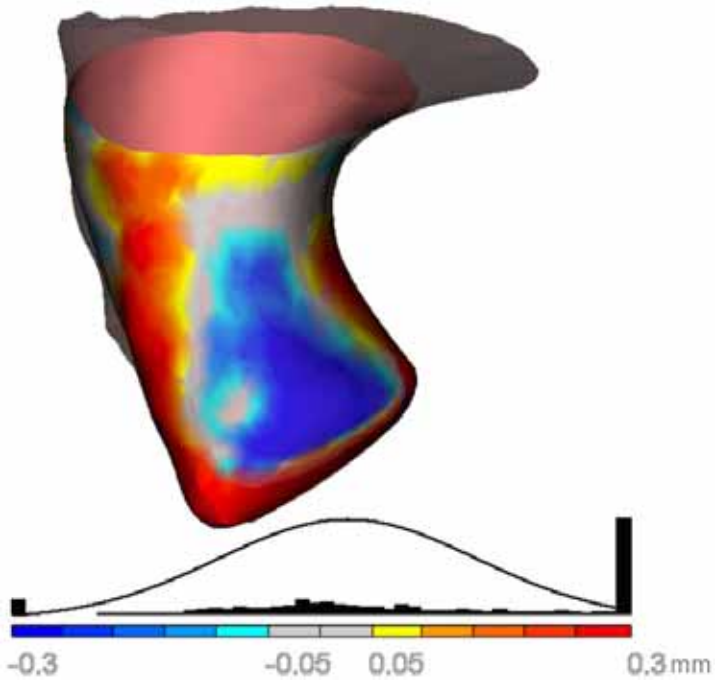


Figure 6.12: Comparison between two different digitally reconstructed surface of the same EAC.

The digital comparison suffers slightly of the overlapping difficulties former discussed, however the average distance and the standard deviation are comparable with the resolution limits of the rapid prototyping techniques used for the hearing aid fabrication

6.5. Hearing aid fabrication

The EAC of a hypoacusis patient, using a hearing aid for the treatment of hearing loss, have been measured. The Figure 6.13 shown the 3D point-cloud map and the digitally reconstructed surface of the hypoacusis patient canal.

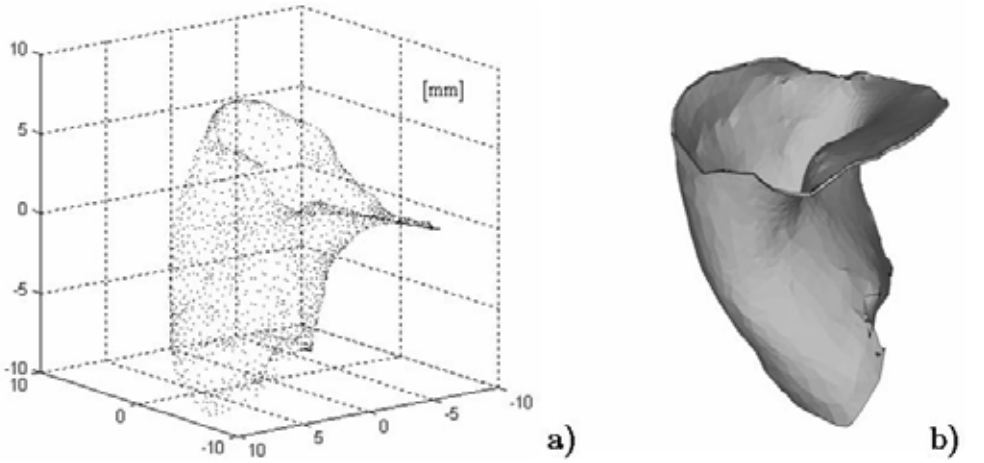


Figure 6.13: a) 3D point-cloud image from direct scanning of hypoacusis patient EAC; b) digitally reconstructed EAC in STL format

Using the 3D data, two hearing aids have been manufactured by Phonak. The modeling process is shown in Figure 6.14.

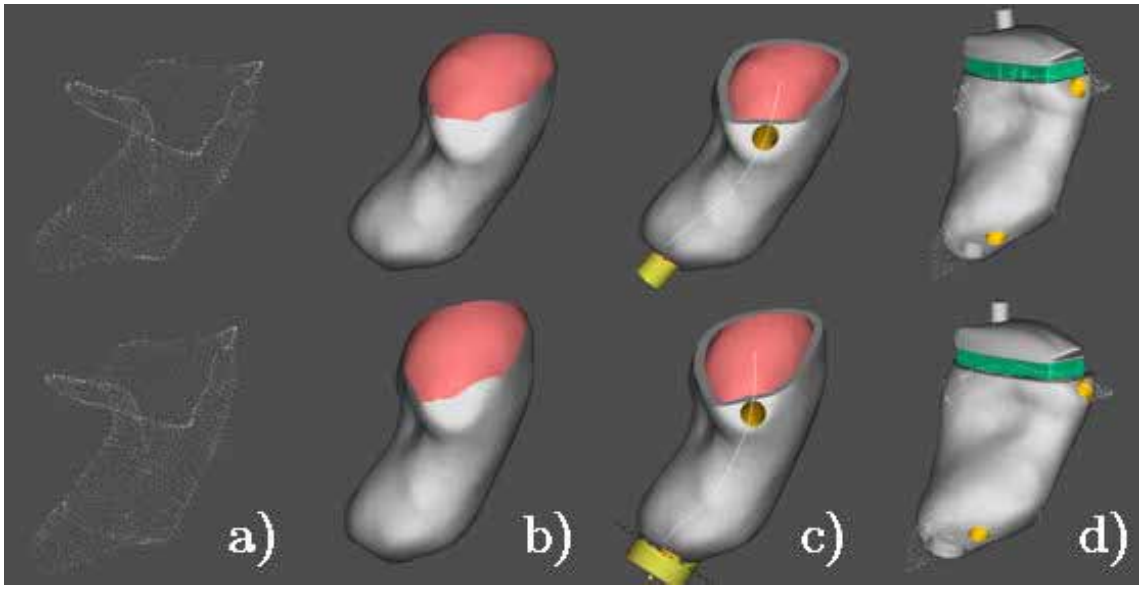


Figure 6.14: CIC hearing aid manufacturing starting out from an in-ear scan: a) in-ear scanned pointcloud; b) surface after electronic detailing, the tip and a large amount of the faceplate contour reconstructed; c) shell after adding the venting channel; d) finished hearing aid shell.

CIC hearing aid manufacturing starts from the 3D map point-cloud of the EAC (Figure 6.14a). From the point-cloud the digital surfaces of the final hearing aids are obtained (Figure 6.14b). The venting channels are added (Figure 6.14c) and the complete modeling of the hearing aid is finished Figure 6.14d).

The feedback is extremely positive. The hearing aids have been really appreciated by the patient. As per patient feedback, the shells wear very well: No over/undersizing of the hearing aid, no pain or discomfort (also chewing and yawning) was suffered, and it was easy to insert/take.

7 Conclusions

The 3D laser scanner based on surface silicon micromachining techniques for shape and size reconstruction of the human ear canal has been presented. The system allows noninvasive 3D imaging of the ear canal with spatial resolution in the 10μm range. To the best of our knowledge, this is the first ever demonstration of direct digital scanning of human external auditory canal by a single integral MEMS system. The complete digital process was validated by testing it on hypoacusis patient, using hearing aids for the treatment of hearing loss. The first working hearing aid fabricated with a completed digital process is accomplished and the patient feedback reported.

The micro-nanofabrication techniques used for the MEMS micromirror fabrication ensure the static and dynamic stability of the devices. The sub-micrometric fabrication precision and the high surface quality(low residual stress and roughness under 15nm) permit to neglect the fatigue effects usually due to the stress and surface crack initiation. The dynamic characteristics of the MEMS micromirror are correlated to the geometrical parameters of the proposed mechanical structure. The high resolution patterning allows the high reproducibility of the static and dynamic behavior of the MEMS micromirror.
The scanning strategy proposed for the MEMS micromirror scanning probe decreases the acquisition time. The rest position angle combining with the flipping of the micromirror permits to acquire deeper portion of the ear surface remaining in a safe and accessible position. The micromirror flipping and the measure overlap eliminate the shadow parts.

The distance measurement technique adopted is based on the conoscopic holography. The conoscopic system provide high performance with the harsh reflectivity characteristics of the ear surface. Moreover, the system offers the

possibility to adjust the resolution and the working distance by just replacing the focal of the acceptance lens.

Further Otoscan3D developments regarding the MEMS scanning probe, the conoscopic system and the motion system have been specified. Recently, a new version of the ConoProbe with reduced volume and weight has been introduced. The performances in terms of the sampling rate have been improved and the laser spot dimension made small. The final dimension of the scanning probe is limited from the conoscopic system laser spot size. Reducing the dimension of the laser beam of the Conoprobe the MEMS micromirror dimension could be scaled by a factor of 4. The micromirror holder on the top of the MEMS scanning probe could be fabricate with LIGA technique in order to obtain a micrometric precision micromirror housing and allow an easier substitution. Decreasing the weight and the volume of the conoscopic system, the sophisticated motion system comprising the PAROS six degree parallel robot could be replaced with a much more simple and cheap motion system.

An important development of the Otoscan3D system is shown in Figure 7.1. A video camera could be inserted between the ConoProbe and the objective lens. Through the camera a video picture of the ear canal and MEMS scanning probe are viewed. The preliminary otoscope analysis could be performed directly with the camera system. The MEMS micromirror scanning probe could be easily and safety drive in the EAC reducing the positioning time of the measurements.

Figure 7.1: Otoscan3D improvement: a video camera is inserted between the ConoProbe and the objective lens

Potential clinical benefits of this research include physician education, clinical-morphometric correlation research such as in diagnosis study of ear canal dimension variations during an inflammatory (such as Otitis externa) phase, refining of CIC fabrication approaches to patient personalized based. As an example, CIC shell placement is highly variable from individual to individual, possibly impacting hearing performance. The actual EAC measurement techniques provide a characterization of the EAC shape, which can help in the design of CIC shell that minimize such variability. It is conceivable that a variety of designs could be developed and optimally matched to a specific patient using this technique.

Appendix A: Conventions and definitions

The following conventions and definitions are used in the chapter 4.
The birefringent crystal is defined, as usual, as

$$\Delta n = n_e - n_o \qquad \text{(Eq. A.1)}$$

We use n_c as the effective value of the index of refraction:

$$n_c = \frac{2 n_o n_e^{\,2}}{n_o + n_e} \qquad \text{(Eq. A.2)}$$

which is shown in Appendix B to simplify the approximation of the extraordinary index of refraction close to the optical axis. The effective value of the index of refraction, n_c, is close to the extraordinary index of refraction, n_e. For example, for calcite at a wavelength of 633 nm, the ordinary index (n_o) is 1.658, the extraordinary index (n_e) is 1.486, the birefringence (Δn) is -0.172 (calcite is a negative crystal), and the effective index of refraction (n_c) is 1.526.

We also use in Appendix B the ratio of twice the birefringence to the square of effective index of refraction, which is:

$$\frac{2\Delta n}{n_c^{\,2}} = \frac{n_e^{\,2} - n_o^{\,2}}{n_o n_e^{\,2}} = \frac{1}{n_o} - \frac{n_o}{n_e^{\,2}} \qquad \text{(Eq. A.1)}$$

The description of the basis setup in its simplified version is given here by Figure 4.1. If we define z as being the optical axis direction of the system, an emitting point P positioned at (x,y,z), a recording point Q at (x',y',z'), a projected point Q' (x,y,z'), which is the projection of the emitting point along the optical axis, and, with a crystal of length L, then we define the ordinary corrected distance z_o, to be:

$$z_0 = (z - z') - L + (L / n_o) \qquad \text{(Eq. A.4)}$$

At first order, the ordinary corrected distance z_o is related to the extraordinary ray direction through:

$$\sin \beta_o = \overrightarrow{QQ'} / z_o \qquad \text{(Eq. A.5)}$$

In the same condition the extraordinary corrected distance, z_e, is:

$$z_e = (z - z') - L + (Ln_o / n_e^2) \qquad \text{(Eq. A.6)}$$

At first order, the extraordinary corrected distance z_e, is related to the extraordinary ray direction through:

$$\sin \beta_e = \overrightarrow{QQ'} / z_e \qquad \text{(Eq. A.7)}$$

We then define the conoscopic corrected distance z_c as the geometrical mean of the ordinary and the extraordinary corrected distance, i.e.:

$$z_c = (z_o z_e)^{1/2} \qquad \text{(Eq. A.8)}$$

In the chapter 4 we use the relaction

$$\frac{dz_c}{dz} = \frac{1}{2(z_o z_e)^{1/2}} \left(z_o \frac{dz_e}{dz} + z_e \frac{dz_o}{dz} \right) = \frac{(z_o + z_e)/2}{(z_o z_e)^{1/2}} \qquad \text{(Eq. A.9)}$$

We define also the following relactions:

$$z_l = (z - z') + (n_o - l)L \qquad \text{(Eq. A.10)}$$

$$l_o \cong \frac{z_l}{\cos\beta} \qquad \text{(Eq. A.11)}$$

We call Eq A.10 the approximation of the optical path length, with z_l as the lateral corrected distance and β (defined in Figure A.1) as the angular direction of the vector PQ. The optical path length is used in the equations involving the inverse square law (Eq. 4.6). We differentiate between β_o and β_e the direction of the ordinary and extraordinary rays, for the phase-delay measurement. However, in the approximation of the optical length for the inverse-square-law coefficient, the difference between β_o and β_e being small, we use a mean value. For the simplified analysis, a simpler, less accurate version of this equation, a crude approximation is:

$$l_o \cong z_c \qquad \text{(Eq. A.12)}$$

The conoscopic parameter is the unitless parameter defines as:

$$\kappa_o = \frac{2\Delta nL}{n_c^{\,2}\lambda} \qquad \text{(Eq. A.13)}$$

It is a measure of the strength of the phenomena in a particular system. It is a function of the birefringence (scaled by the effective index of refraction) of the length of the crystal and (inversely) of the wavelength of the light.
The Fresnel parameter, defined as:

$$f_r = \frac{2\Delta nL}{n_c^{\,2}\lambda}\frac{1}{z_c^{\,2}} = \frac{\kappa_o}{z_c^{\,2}} \qquad \text{(Eq. A.14)}$$

It is the parameter of the conoscopic transfer function Eq. A.20, rewritten as Fresnel transform, it is useful mostly in the two-dimensional developments.

The basic measure of the strength of the effect is the number of fringes. From the old, artificial convention that counts both bright and dark fringes, a fringe corresponds to a half-wave optical path difference. It is measured along the meridional ray (see Figure. A.1). The number of fringes, F, is given by:

$$F = \frac{2\Delta nL}{n_c^{\,2}\lambda}\frac{R^2}{z_c^{\,2}} = \kappa_o \frac{R^2}{z_c^{\,2}} = f_r R^2 \qquad \text{(Eq. A.15)}$$

where R is the radius of the conoscopic hologram.

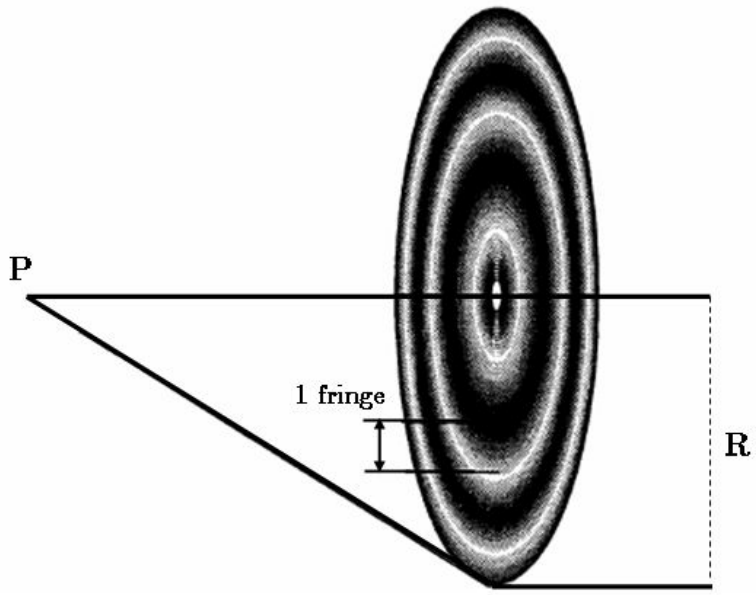


Figure A.1: Definition of a fringe

We also define x_G, the radius of the fringe G, to be:

$$x_G^{\,2} = \frac{G z_c^{\,2}}{\kappa_o} \qquad \text{(Eq. A.16)}$$

and the fringe-count difference to be:

$$\Delta x_F = x_F - x_{F-1} = \frac{x_F^{\,2} - x_{F-1}^{\,2}}{x_F + x_{F-1}} \cong \frac{\left[F-(F-1)\right] z_c^{\,2}}{\kappa_o (2R)} \qquad \text{(Eq. A.17)}$$

by using the fact that $x_{F-1} \cong x_F = R$, and, from Eq. A.15:

$$\Delta x_F = \frac{z_c^{\,2}}{\kappa_o (2R)} = \frac{R^2}{2FR} = \frac{R}{2F} \qquad \text{(Eq. A.18)}$$

The number of fringes is also a measure of the temporal coherence needed for the system. Indeed, if we neglect walk-off and geometrical effects that affect the spatial coherence, the maximum phase delay between the ordinary and the extraordinary

rays in the system is πF by definition. The temporal coherence is then $\pi F/2$ wavelengths.

In the case of the on-axis uniaxial crystal the extraordinary-wave polarization direction (the TM wave) is angle dependent. However, for the on-axis conoscope, for each angle the field is separated into ordinary and extraordinary fields, whose directions are angle dependent, and recombined by the second polarizer. After this polarizer, the extraordinary field for all angles has the same polarization direction and interfere.

For an on-axis conoscope with an aperture placed at the entrance face of the crystal at plane z_A, Bergstein and Zachos [74] showed that the ordinary field is given by:

$$u_o(x', y', z') = \frac{-jn_o}{\lambda} \int_R u(\xi, \eta, z_A) \frac{1}{R_o} \exp(jkn_o R_o) d\xi d\eta \qquad \text{(Eq. A.19)}$$

and the extraordinary field is given by:

$$u_e(x', y', z') = \frac{-jn_o}{\lambda} \int_R u(\xi, \eta, z_A) \frac{1}{R_e} \exp(jkn_e R_e) d\xi d\eta \qquad \text{(Eq. A.20)}$$

where $k(= 2\pi/\lambda)$ is the wave vector in vacuo and $u(\xi,\eta,zA)$ is the incident field amplitude at plane z_A at lateral position (ξ,η). R_o is the radius vector length and is given for an on-axis conoscope by:

$$R_o^{\,2} = (x'-\xi)^2 + (y'-\eta)^2 + (z'-z_A)^2 \qquad \text{(Eq. A.21)}$$

R_e is an effective distance, which replaces the distance in the isotropic Huygens-Fresnel principle. R_e is the radius vector length R_c, decreased by the projection of the radius vector on the optical axis scaled by a birefringent-dependent constant, and is given for an on-axis conoscope by:

$$R_e^{\,2} = R_o^{\,2} - \frac{n_e^{\,2} - n_o^{\,2}}{n_e^{\,2}} (z'-z_A)^2 = (x'-\xi)^2 + (y'-\eta)^2 + \frac{n_o^{\,2}}{n_e^{\,2}} (z'-z_A)^2 \qquad \text{(Eq. A.22)}$$

and we neglect the obliquity factor.

We take the aperture to be centered on the emitting point. A basic consideration of this research is the fact that the system is shift invariant. To keep this property and also following the discussion of three-dimensional optics by

Streibl [74-76] and Sitter and Rhodes [77], we need to design the imaging optics to be telecentric. Let a point positioned at *(0,0,0)* emit light. At point *(ξ,η,zA)* the light amplitude is given by

$$u(\xi,\eta,z_A)=\exp\left(jk\frac{\xi^2+\eta^2}{2z_A}\right) \qquad \text{(Eq. A.23)}$$

so that:

$$\begin{aligned} u_o(x',y',z') &= \frac{-jn_o}{\lambda}\int_R \exp\left(jk\frac{\xi^2+\eta^2}{2z_A}\right)\frac{1}{R_o}\exp(jkn_oR_o)d\xi d\eta = \\ &= \frac{-jn_o}{\lambda}\int_R \frac{1}{R_o}\exp\left[jk\left(\frac{\xi^2+\eta^2}{2z_A}+n_oR_o\right)\right]d\xi d\eta \end{aligned} \qquad \text{(Eq. A.24)}$$

$$\begin{aligned} u_e(x',y',z') &= \frac{-jn_o}{\lambda}\int_R \exp\left(jk\frac{\xi^2+\eta^2}{2z_A}\right)\frac{1}{R_e}\exp(jkn_eR_e)d\xi d\eta = \\ &= \frac{-jn_o}{\lambda}\int_R \frac{1}{R_o}\exp\left[jk\left(\frac{\xi^2+\eta^2}{2z_A}+n_eR_e\right)\right]d\xi d\eta \end{aligned} \qquad \text{(Eq. A.25)}$$

From the experimental results, the difference between the wave and the ray optics descriptions is small. Instead of developing a complete theory involving complicated integrals, the way of numerical evaluation of these integrals is chosen. Comparing the optical path difference, Δl_ω, calculated with Eq. 4.24 and Eq. 4.25, for the typical case developed there, with the ray optical path difference Δl_R and show that they are close.

Appendix B: Simplification

In this appendix summarize the simplifications used in the chapter 4.
Fresnel's law of light propagation in crystal can be written as:

$$\frac{1}{n_e^{\,2}(\theta_e)} = \frac{\sin^2\theta_e}{n_e^{\,2}} + \frac{\cos^2\theta_e}{n_o^{\,2}} \qquad \text{(Eq. B.1)}$$

where θ_e is the phase angle direction of the extraordinary ray.
For small angles the second-order approximation of the squared-sine and squared-cosine functions can be used to obtain:

$$\frac{1}{n_e^{\,2}(\theta_e)} = \frac{\theta_e^{\,2}}{n_e^{\,2}} + \frac{1-\theta_e^{\,2}}{n_o^{\,2}} = \frac{1}{n_o^{\,2}} + \theta_e^{\,2}\left(\frac{1}{n_e^{\,2}} - \frac{1}{n_o^{\,2}}\right) \qquad \text{(Eq. B.2)}$$

Rewriting Eq. B.2 with:

$$n_e(\theta_e) = n_o + \Delta n_e(\theta_e) \qquad \text{(Eq. B.3)}$$

we obtain:

$$\frac{1}{[n_o + \Delta n_e(\theta_e)]^2} = \frac{1}{n_o^{\,2}} + \theta_e^{\,2}\left(\frac{1}{n_e^{\,2}} - \frac{1}{n_o^{\,2}}\right) \qquad \text{(Eq. B.4)}$$

or

$$\frac{1}{\left[n_o+\Delta n_e(\theta_e)\right]^2}-\frac{1}{n_o{}^2}\cong\theta_e{}^2\left(\frac{1}{n_e{}^2}-\frac{1}{n_o{}^2}\right)\cong\frac{1}{n_o{}^2\left\{1+\left[2\Delta n_e(\theta_e)/n_o\right]\right\}}-\frac{1}{n_o{}^2}\cong$$
$$\frac{1-\left[2\Delta n_e(\theta_e)/n_o\right]}{n_o{}^2}-\frac{1}{n_o{}^2}\cong\frac{-2\Delta n_e(\theta_e)}{n_o{}^3} \qquad \text{(Eq. B.5)}$$

Thus we obtain:

$$\frac{-2\Delta n_e(\theta_e)}{n_o{}^3}\cong\theta_e{}^2\left(\frac{1}{n_e{}^2}-\frac{1}{n_o{}^2}\right) \qquad \text{(Eq. B.6)}$$

or

$$\Delta n_e(\theta_e)\cong\theta_e{}^2\left(\frac{n_o{}^3}{n_e{}^2}-\frac{n_e{}^3}{n_o{}^2}\right)\cong n_o\theta_e{}^2\left(\frac{n_o{}^2}{n_e{}^2}-1\right)\cong\theta_e{}^2\Delta n\frac{n_o{}^2}{n_c{}^2} \qquad \text{(Eq. B.7)}$$

by using the definition of the birefringence and the effective index of refraction as defined in Appendix A (Eqs. A.l and A.2 and their ratio, Eq. A.3). The approximation of θ_e, the angle direction of the ray in the crystal, is simply given by:

$$\theta\cong r/z_c \qquad \text{(Eq. B.8)}$$

The transmission coefficients of the ordinary and the extraordinary beams (assumed to be real numbers) are almost equal and can be taken to be so in most cases. We use this approach in the chapter 4, replacing both by t, the mean value. We take into account the slight difference between the transmission coefficients, t_o and t_e', owing to differences in transmission and absorption coefficients of the two modes and owing to the slight geometrical dependence of the walk-off effect. For a further simplification, the angular dependence of the transmission coefficients is neglected in the equations.

Appendix C: Interferometric accuracy

We have recorded a positive conoscopic hologram of a single point, as described in chapter 4(sect2), onto a two-dimensional sensor. The optical power density (in units of power per area, i.e., watts per square meter or watts per square centimeter) distribution of the hologram given by Eq. 4.15 and incident upon the detector array is rewritten as:

$$I_{+}(x',y') = I_o\left(1+\cos\left\{\pi\kappa_o\frac{\left[(x_o-x')^2+(y_o-y')^2\right]}{z_c^{\,2}}\right\}\right) \qquad \text{(Eq. C.1)}$$

The two-dimensional sensor is characterized by its lateral pixel sizes (Δx and Δy), taken here to be equal ($=\Delta$), and A_p, the pixel area ($=\Delta^2$); the number of pixels, N, in each direction (taken to be equal in both two dimensions for simplicity); η, the quantum efficiency of the detector; and τ, the integration time of the detector.
In discrete form, we rewrite Eq. C.1 as:

$$I_{+}(l',m') = I_o\left(1+\cos\left\{\frac{\pi F}{R_p^{\,2}}\left[(l'-l_o)^2+(m'-m_o)^2\right]\right\}\right) \qquad \text{(Eq. C.2)}$$

where x_o, y_o, l_o, and m_o are the lateral coordinates of the emitting point in metric and pixels units, related by:

$$x_o = l_o\Delta_x \qquad y_o = m_o\Delta_y \qquad \text{(Eq. C.3)}$$

and we make use of the relation developed in Appendix A (Eq. A.15) relating the conoscopic parameter and the number of fringes:

$$F = \kappa_o \frac{R^2}{z_c^{\ 2}} \qquad \text{(Eq. C.4)}$$

R being the hologram radius and R_p the hologram radius in pixels units, i.e., $R_p^2 = R^2/A_p$.

We record a negative conoscopic hologram (as explained in Appendix A) on the same two-dimensional sensor, and the optical power density is given in discrete form by:

$$I_-(l',m') = I_o\left(1-\cos\left\{\frac{\pi F}{R_p^{\ 2}}\left[(l'-l_o)^2+(m'-m_o)^2\right]\right\}\right) \qquad \text{(Eq. C.5)}$$

The hologram, $H(l', m')$, is given in discrete form as:

$$H(l',m') = \frac{I_+(l',m')-I_-(l',m')}{2} = I_o\cos\left\{\frac{\pi F}{R_p^{\ 2}}\left[(l'-l_o)^2+(m_o-m')^2\right]\right\} \qquad \text{(Eq. C.6)}$$

The optical power in the optical pattern (hologram and bias) is then detected by the array detector. Let us define the number of detected bias photons (or the number of created electrons) in a single detector to be:

$$c_o = I_o\frac{\eta\tau}{h\nu}A_p \qquad \text{(Eq. C.7)}$$

and the total bias count of photons (or electrons) to be:

$$C_T = I_o\frac{\eta\tau}{h\nu}\pi R^2 = c_o\frac{\pi R^2}{A_p} \qquad \text{(Eq. C.8)}$$

where ν is the optical frequency and h is the Planck constant.

We define the counts, $K(l',m')$, of the number of photons (or electrons) detected to be:

$$K(l',m') = \alpha_Q H(l',m') = c_o \cos\left\{\frac{\pi F}{R_p^{\,2}}\left[(l_o - l')^2 + (m' - m_o)^2\right]\right\} \qquad \text{(Eq. C.9)}$$

where α_Q, the proportionality constant, is given by:

$$\alpha_Q = \frac{c_o}{I_o} = \frac{\eta\tau}{h\nu} A_p \qquad \text{(Eq. C.10)}$$

To find the longitudinal resolution for a single point, assume that the lateral position is known a priori, either by prior knowledge or by using a standard imaging camera. Apply, as an estimation tool, the inverse Fresnel transform IF_rT; i.e., evaluate the complex sum:

$$F_r(F_1, l_o, m_o) = \sum_{l',m' \in R} K(l',m') \exp\left\{-j\frac{\pi F_1}{R_p^{\,2}}\left[(l' - l_0)^2 + (m' - m_o)^2\right]\right\} \qquad \text{(Eq. C.11)}$$

$$F_r(F_1, l_o, m_o) = K_r + jK_i \qquad \text{(Eq. C.12)}$$

where l', $m' \in R$, which signifies that l' and m' belong to a circle of radius R around (l_o, m_o) and K_r and K_i, are the real and the imaginary parts of the IF_rT, $|K|$ its module, and θ_K its phase. We denote their mean values by $\overline{K}_r$ and $\overline{K}_i$ and their variances by σ_r^2 and σ_i^2.

For a known lateral position, we evaluate only the fringe parameter F_1. For simplicity we take the point to be positioned at the origin (l_o, $m_o = 0$). The IF_rT is then:

$$F_r(F_1) = c_o \frac{1}{2}\left(\sum_{l',m' \in R}\left\{\exp\left[-j\frac{\pi(F_1 - F)}{R_p^{\,2}}\left(l'^2 + m'^2\right)\right] + \exp\left[-j\frac{\pi(F_1 + F)}{R_p^{\,2}}\left(l'^2 + m'^2\right)\right]\right\}\right) \qquad \text{(Eq. C.13)}$$

For F close to F_1 the second term is negligible compared with the first-it can also be removed by recording the exponential hologram as described in Chapter 4.

Finally we define ΔF, the difference in fringes, as:

$$\Delta F = F_1 - F \qquad \text{(Eq. C.14)}$$

The IF_rT at F, is given approximately by:

$$F_r(F_1) \cong \frac{c_o}{2} \sum_{l',m' \in R} \exp\left[- j \frac{\pi \Delta F}{R_p^{\,2}} \left(l'^2 + m'^2 \right) \right] \qquad \text{(Eq. C.15)}$$

We replace the sum by an integral, using:

$$\begin{aligned} dx &= \Delta dl' \\ dy &= \Delta dm' \\ r dr d\theta = dx dy &= \Delta^2 dl' dm' = A_p dl' dm' \end{aligned} \qquad \text{(Eq. C.16)}$$

to obtain:

$$\begin{aligned} F_r(F_1) &\cong \frac{c_o}{2A_p} \int_{\theta=0}^{2\pi} d\theta \int_{R=0}^{R} \exp\left(- j \frac{\pi \Delta F r^2}{R^2} \right) r dr \cong \\ &\cong \frac{2\pi c_o}{2A_p} \int_{R=0}^{R} \exp\left(- j \frac{\pi \Delta F r^2}{R^2} \right) r dr \cong \frac{C_T}{R^2} \int_{R=0}^{R} \exp\left(- j \frac{\pi \Delta F r^2}{R^2} \right) r dr \end{aligned} \qquad \text{(Eq. C.17)}$$

Note that, in the exponential argument, we return to continuous notation instead of a discrete notation (R_p and (l', m')) as in previous equations to simplify the comprehension of the already complex calculus.
If we set $u = r^2$ (i.e., $du = 2rdr$),

$$F_r(F_1) \cong \frac{C_T}{R^2} \int_{R=0}^{R} \exp\left(- j \frac{\pi \Delta F u}{R^2} \right) du \qquad \text{(Eq. C.18)}$$

$$\begin{aligned} F_r(F_1) &\cong j \frac{C_T}{2\pi \Delta F} \left[\exp(- j\pi \Delta F) - 1 \right] = \frac{C_T}{\pi \Delta F} \exp\left(\frac{j\pi \Delta F}{2} \right) \sin \frac{\pi \Delta F}{2} = \\ &= \frac{C_T}{2} \exp\left(\frac{j\pi \Delta F}{2} \right) \left[\frac{\left(\sin \frac{\pi \Delta F}{2} \right)}{\frac{\pi \Delta F}{2}} \right] = \frac{C_T}{2} \sin c\left(\frac{\Delta F}{2} \right) \exp\left(- j \frac{\pi \Delta F}{2} \right) \end{aligned} \qquad \text{(Eq. C.19)}$$

from the definition *sinc(x) = sin(πx)/(πx)*. We then obtain, using the definitions of Eq C.12:

$$K_r = \frac{C_T}{2}\sin c\left(\frac{\Delta F}{2}\right)\cos\left(\frac{\pi\Delta F}{2}\right) \qquad \text{(Eq. C.20)}$$

$$K_i = \frac{C_T}{2}\sin c\left(\frac{\Delta F}{2}\right)\sin\left(\frac{\pi\Delta F}{2}\right) \qquad \text{(Eq. C.21)}$$

We also obtain:

$$|K| = \frac{C_T}{2}\sin c\left(\frac{\Delta F}{2}\right) \qquad \text{(Eq. C.22)}$$

$$\phi_\kappa = \frac{\pi\Delta F}{2} \qquad \text{(Eq. C.23)}$$

so that:

$$\Delta F = \frac{2\phi_\kappa}{\pi} \qquad \text{(Eq. C.24)}$$

For ΔF different from 0 the IF_rT is small. Far from the origin, the sinc term will be small and the modulus of the IF_rT small. We use this information (Eq. C.22) to find the rough longitudinal position (within one fringe) and the phase information of Eq. C23 for the accurate value.

Let us perform the following sequence of events for a static range finder, i.e., system in which a point is projected onto an object; we want to measure its distance. The conoscopic corrected distance-in fringes (see Eq. A.6 for the definition of conoscopic corrected distance and Eq. A.15 for the definition of the number of fringes) is between F_{min} and F_{max}:

1. the IF_rT is calculated for values of F_1, each fraction of a fringe;
2. from these values we can find the values between which the actual value of F_1 falls. Indeed, these two values have the highest modulus and inverse phase (from Eqs. C.22 and C.23);
3. a numerical zero-finding algorithm is then applied to find the value for which the phase (or the imaginary part) of the IF_rT is 0.

We denote the values F_m. If we neglect the contribution of the real part of the IF_rT to the noise, the rms error in the value of F_m is simply given by:

$$\Delta F_{rms} = \frac{2\sigma_i}{\pi K_r} \cong \frac{2\sigma_i}{\pi \overline{K}_r} \qquad \text{(Eq. C.25)}$$

The results for the discrete Fourier transform can be applied directly to the discrete Fresnel transform. With the assumption of thermally generated light, we obtain:

$$\sigma_i^{\,2} = \frac{C_T}{2} \qquad \text{(Eq. C.26)}$$

$$\overline{K}_r = \frac{C_T}{2} \qquad \text{(Eq. C.27)}$$

We finally obtain:

$$\Delta F_{rms} = \frac{2\sigma_i}{\pi \overline{K}_r} = \frac{2\sqrt{2}}{\pi \sqrt{C_T}} \qquad \text{(Eq. C.28)}$$

The theoretical error in the longitudinal position is proportional to the inverse square root of the total number of photons (electrons).

Acknowledgments

I thank all the persons who have contributed to this work: Enzo Di Fabrizio, my tutor and my mentor, Radu Malureanu, Stefano Cabrini, Dan Cojoc, Alessandro Carpentiero, Luca Businaro, Arrigo Gosparini, Rakesh Kumar and Gianluca Terragni for their fundamental help and contribution.
I also thank Amplifon SPA for assistance and financial support, and PHONAK for comparison analysis and hearing aid fabrication.
I thank all my friends and colleagues for ear-rents and patience. Without your contribution no scans could have been acquired
I thank Mamma, Papà and Ale that finally can calm down: I finished, maybe.
Unfortunately from now on my nervous behavior will not be justifiable anymore.
And Giulia you're the first (page)

Bibliography

1. NIDCD, *Statistics about Hearing Disorders, Ear Infections, and Deafness*. 2006, National Institute on Deafness and Other Communication Disorders.
2. Wallhagen, M.I., et al., *Comparative impact of hearing and vision impairment on subsequent functioning.* J Am Geriatr Soc, 2001. **49**(8): p. 1086-92.
3. Ross, M., *The occlusion effect.* Hearing Loss, 2004.
4. Jorzick, J., et al., *Spin Wave Wells in Nonellipsoidal Micrometer Size Magnetic Elements.* Physical Review Letters, 2002. **88**(4): p. 047204.
5. Nakajima, Y., et al., *Fabrication of a silicon quantum wire surrounded by silicon dioxide and its transport properties.* Applied Physics Letters, 1994. **65**(22): p. 2833-2835.
6. Schwarz, M.P., et al., *Micromechanical cantilever magnetometer with an integrated two-dimensional electron system.* Applied Physics Letters, 2000. **76**(24): p. 3564-3566.
7. Mooij, J.E., et al., *Josephson Persistent-Current Qubit.* Science, 1999. **285**(5430): p. 1036-1039.
8. Nakamura, Y., Y.A. Pashkin, and J.S. Tsai, *Coherent control of macroscopic quantum states in a single-Cooper-pair box.* Nature, 1999. **398**(6730): p. 786-788.
9. Cianci, E., et al., *One-dimensional capacitative micromachined ultrasonic transducer arrays for echographic probes.* Microelectronic Engineering, 2004. **73**: p. 502-507.
10. Giles, C.R., et al., *A silicon MEMS optical switch attenuator and its use in lightwave subsystems.* Selected Topics in Quantum Electronics, IEEE Journal of, 1999. **5**(1): p. 18-25.
11. Tuantranont, A., et al., *Phase-only micromirror array fabricated by standard CMOS process.* Sensors and Actuators A: Physical, 2001. **89**(1): p. 124-134.
12. Cui, L., T. Zhang, and H. Morgan, *Optical particle detection integrated in a dielectrophoretic lab-on-a-chip.* Journal of Micromechanics and Microengineering, 2002. **12**(1): p. 7-12.
13. Weigl, B.H., R.L. Bardell, and C.R. Cabrera, *Lab-on-a-chip for drug development.* Advanced Drug Delivery Reviews, 2003. **55**(3): p. 349-377.

14. Chou, S.Y., P.R. Krauss, and P.J. Renstrom, *Imprint of sub 25 nm vias and trenches in polymers.* Applied Physics Letters, 1995.
15. Tormen, M., et al., *3D patterning by means of nanoimprinting, X-ray and two-photon lithography.* Microelectronic Engineering, 2004. **73**: p. 535-541.
16. Thong, J.T.L., *Electron beam testing technology*. 1993: New York, Plenum Press.
17. Valiev, K.A., *The physics of submicron lithography*. 1992: New York, Plenum Press.
18. Akkerman, A.F., Y.M. Nikitushev, and V.A. Botvin, *Solving problem in the transmission of fast electrons in matter by the Monte Carlo method*. 1972: Nauka, Alma-Ata.
19. Hawryluk, R.J., A.M. Hawryluk, and H.I. Smith, *Energy dissipation in a thin polymer film by electron beam scattering*. Journal of Applied Physics, 1974. **45**(6): p. 2551-2566.
20. Parikh, M., *Corrections to proximity effects in electron beam lithography. I. Theory.* Journal of Applied Physics, 1979. **50**(6): p. 4371-4377.
21. Chow, D.G.L., et al., *An image processing approach to fast, efficient proximity correction for electron beam lithography.* Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures, 1983. **1**(4): p. 1383-1390.
22. Carroll, A.M., *Proximity-effect correction with linear programming*. Journal of Applied Physics, 1981. **52**(1): p. 434-437.
23. Rai-Choudhury, P., *Handbook of Microlithography, Micromachining, and Microfabrication. Volume 1: Microlithography*. 1997: SPIE press.
24. Lifšič, M. and L. Landau, *The Classical Theory of Fields*. 1980: Elsevier.
25. Guo, J.Z. and F. Cerrina. *Absorber roughness effect in XRL image formation*. in *Electron-Beam, X-Ray, and Ion-Beam Submicrometer Lithographies for Manufacturing III*. 1993. San Jose, CA, USA: SPIE.
26. Guo, J.Z.Y., et al. *Experimental study of aerial images in x-ray lithography*. in *Proceedings of the 16th international symposium on electron, ion, and photon beams*. 1993. San Diego, California (USA): AVS.
27. Gentili, M., et al. *Fabrication of controlled slope attenuated phase-shift x-ray masks for 250 nm synchrotron lithography*. in *The 38th international symposium on electron, ion, and photon beams*. 1994. New Orleans, Louisiana (USA): AVS.
28. McCord, M.A., A. Wagner, and D. Seeger. *Effect of mask absorber thickness on x-ray exposure latitude and resolution*. in *Proceedings of the 16th international symposium on electron, ion, and photon beams*. 1993. San Diego, California (USA): AVS.
29. Businaro, L., et al., *High-resolution complex structures for two-dimensional photonic crystals realized by x-ray diffraction lithography.* Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures, 2003. **21**(2): p. 748-753.
30. Abbe, E., Archiv. f. Mikraskopische Anat., 1873. **9**(413).
31. Köhler, M., *Etching in Microsystem Technology*. 1999: Wiley-VCH.
32. Chen, F.F., *Introduction to plasma physics*. 1974: New York, Plenum Press.
33. Koenig, H. and L. Maissel, *Electrical properties of RF sputtering systems.* IBM J. Res. Develop, 1970. **14**(168).

34. Brodie, I. and J.J. Muray, *The physics of microfabrication*. 1982: New York, Plenum Press.
35. Maluf, N., *An Introduction to Microelectromechanical Systems Engineering*. 2000: Boston, Artech House.
36. Romankiw, L., et al., *Advantages and special considerations in fabricating bubble circuits by electroplating and sputter etching.* Magnetics, IEEE Transactions on, 1974. **10**(3): p. 828-831.
37. Hughes, H.G. and L.T. Romankiw, *Etching for pattern definition*. 1976: Princeton.
38. Harsch, S., A. Maner, and W. Ehrfeld, *Mass Production of Microdevices With Extreme Aspect Ratios by Electroforming*. Plat. Surf. Finish., 1988. **75**(3): p. 60-65.
39. Helvajian, H., *Microengineering Aerospace Systems*. 1999: The Aerospace Press.
40. DeVoe, D.L. and A.P. Pisano, *Surface micromachined piezoelectric accelerometers (PiXLs).* Microelectromechanical Systems, Journal of, 2001. **10**(2): p. 180-186.
41. Lemkin, M.A., et al. *3-Axis surface micromachined sigma-delta accelerometer*. in *Digest of Technical Papers - IEEE International Solid-State Circuits Conference*. 1997.
42. Nasby, R.D., et al., *Application of chemical-mechanical polishing to planarization of surface- micromachined devices.* Proc. Solid-State Sensor and Actuator Workshop, 1996: p. 48-49.
43. Sniegowski, J.J., *Multi-level polysilicon surface-micromachining technology: Applications and issues.* American Society of Mechanical Engineers, Aerospace Division (Publication) AD, 1996. **52**: p. 751-759.
44. Nathanson, H.C. and R.A. Wickstrom, *A resonant gate silicon surface transistor with high Q band pass properties.* Applied Physics Letters, 1965. **7**(4): p. 84-86.
45. Becker, E.W., et al., *Fabrication of microstructures with high aspect ratios and great structural heights by syncrotron radiation lithography, galvanoforming, and plastic moulding (LIGA process).* Microelectronic Engineering, 1986. **4**(1): p. 35-56.
46. Smith, J.H. and S.T. Walsh, *Selecting a process paradigm for an emergent disruptive technology: evidence from the emerging microsystems technology base.* IEEE International Engineering Management Conference, 1998: p. 7-10.
47. Wiegerink, R., et al., *Quasi-monolithic silicon load cell for loads up to 1000 kg with insensitivity to non-homogeneous load distributions.* Sensors and Actuators, A: Physical, 2000. **80**(2): p. 189-196.
48. Timoshenko, S. and S. Woinowsky-Krieger, *Theory of Plates and Shells*. 1959: McGraw-Hill.
49. Stankevic, V. and C. Simkevicius, *Use of a shock tube in investigations of silicon micromachined piezoresistive pressure sensors.* Sensors and Actuators, A: Physical, 2000. **86**(1-2): p. 58-65.
50. Liu, C. and Y.W. Yi, *Micromachined magnetic actuators using electroplated permalloy.* IEEE Transactions on Magnetics, 1999. **35**(3 PART 2): p. 1976-1985.

51. Toshiyoshi, H. and H. Fujita, *Electrostatic micro torsion mirrors for an optical switch matrix.* Journal of Microelectromechanical Systems, 1996. **5**(4): p. 231-237.
52. Di Fabrizio, E., et al., *Microphotonic devices fabricated by silicon micromachining techniques.* Japanese Journal of Applied Physics, Part 1: Regular Papers and Short Notes and Review Papers, 1997. **36**(12 SUPPL. B): p. 7757-7762.
53. Prasciolu, M., et al., *Electromagnetically Actuated Surface Micromachined Free Standing Torsion Beam Micromirror Made by Electroplated Nickel.* Japanese Journal of Applied Physics, Part 1: Regular Papers and Short Notes and Review Papers, 2004. **43**(1): p. 418-423.
54. Ohno, I., *Electrochemistry of electroless plating.* Materials Science and Engineering A, 1991. **A146**(1-2): p. 33-49.
55. Scheeper, P.R., et al., *Investigation of attractive forces between PECVD silicon nitride microstructures and an oxidized silicon substrate.* Sensors and Actuators, A: Physical, 1992. **30**(3): p. 231-239.
56. Takeshima, N., et al. *Electrostatic parallelogram actuators*. in *Transducers '91*. 1991.
57. G. T. Mulhern, D. S. Soane, and R.T. Howe. *Supercritical carbon dioxide drying of microstructures*. in *Tech. Dig. 7th Int. Conf. Solid-State Sensors and Actuators (Transducers 93)*. 1993. Yokohama, Japan.
58. Lee, Y.I., et al., *Dry release for surface micromachining with HF vapor-phase etching*. Journal of Microelectromechanical Systems, 1997. **6**(3): p. 226-233.
59. Patterson, P.R., et al. *Scanning micromirrors: An overview*. in *Proceedings of SPIE - The International Society for Optical Engineering*. 2004.
60. Kim, W.H. and C. Laird, *Crack nucleation and stage I propagation in high strain fatigue machanism.* Acta Metall, 1978. **26**(5): p. 789-799.
61. Gohn, G.R. and C.H. Greenall, *Fatigue properties of non-ferrous sheet metals.* Proc ASTM, 1937
62. E. Di Fabrizio, S. Cabrini, and A.D. Cojoc, *An apparatus for determining the internal outline of a duct or cavity*, in *Int. Patent*. 2004.
63. Sirat, G. and D. Psaltis, *Conoscopic holography.* Optics Letters, 1985. **10**(1): p. 4.
64. Goodman, J.W., *Statistical Optics.* 1985: John Wiley & Sons.
65. Leith, E.N. and B.J. Chang, *Space-invariant holography with quasi-coherent light.* Appl. Opt. , 1973. **12**(8): p. 1957-1963.
66. Dufresne, E., P. Chavel, and G. Sirat, *Holographic apparatus using incoherent light*, in *US Patent*. 1990.
67. Soroko, L.M., *Holography and coherent optics*. 1980: New York, Plenum Press.
68. Dufresne, E.Y. *Off-axis conoscopic holography*. in *Conference on laser and electro-optics*. 1989.
69. Charlot, D. and D. Psaltis. *Realization of a conoscopic television system*. in *Conference on laser and electro-optics*. 1989.
70. Charlot, D., et al., *Conoscopic Holography. Numerical Reconstruction.* Holographie conoscopique. Reconstructions numeriques, 1988. **43**(7-8): p. 460-466.

71. Prasciolu, M., et al., *Three-dimensional digital scanner based on micromachined micromirror for the metrological measurement of the human ear canal.* Journal of Vacuum Science and Technology B: Microelectronics and Nanometer Structures, 2005. **23**(6): p. 2990-2994.
72. Taubin, G. *Signal processing approach to fair surface design*. in *Proceedings of the ACM SIGGRAPH Conference on Computer Graphics*. 1995.
73. Frohse, F., et al., *Atlas of Human Anatomy*. 1985: Barnes & Noble.
74. Bergstein, L. and T.H. Zachos, *Diffraction by plane apertures in uniaxially anisotropic Media.* J Opt Soc Amer, 1971. **61**(11): p. 1477-1486.
75. Streibl, N., *Fundamental restrictions for 3-D light distributions.* Optik (Jena), 1984. **66**(4): p. 341-354.
76. Streibl, N., *Depth transfer by an imaging system.* Optica Acta, 1984. **31**(11): p. 1233-1241.
77. Sitter, D.N. and W.T. Rhodes, *Three-dimensional imaging: a space invariant model for space variant systems.* Appl. Opt., 1990. **29**: p. 3789.

List of publications

- Prasciolu, M., Malureanu, R., Cabrini, S.,Cojoc, D., Businaro, L., Carpentiero, A., Kumar, R., Di Fabrizio, E. *Three-dimensional digital scanner based on micromachined micromirror for the metrological measurement of the human ear canal.* Journal of Vacuum Science and Technology B: Microelectronics and Nanometer Structures, 2005. **23**(6): p. 2990-2994.
  also selected for
  Virtual Journal of Nanoscale Science & Technology, December 19, 2005, Volume 12, Issue 25
  and
  Virtual Journal of Biological Physics Research, December 15, 2005, Volume 10, Issue 12
- Prasciolu, M., Carpentiero, A., Kumar, R., Cojoc, D., Cabrini, S., Businaro, L., Romanato, F., Di Fabrizio, E., Recchia, D., Parmigiani, G. *Electromagnetically Actuated Surface Micromachined Free Standing Torsion Beam Micromirror Made by Electroplated Nickel* Japanese Journal of Applied Physics, Part 1: Regular Papers and Short Notes and Review Papers 2004. **43** (1) , 418-423
- Di Fabrizio, E., Fillipo, R., Cabrini, S., Kumar, R., Perennes, F., Altissimo, M., Businaro, L., Cojoc, D., Vaccari, L., Prasciolu, M., Candeloro, P. *X-ray lithography for micro-and nano-fabrication at ELETTRA for interdisciplinary applications* Journal of Physics Condensed Matter 2004. **16** (33) , S3517-S3535

- Prasciolu, M., Cabrini, S., Cojoc, D., Malureanu, R., Kumar, R., Businaro, L., Di Fabrizio, E. *Fabrication through silicon micromachining of 3D scanner for optical determination of the ear canal*. IFMBE Proceedings, Volume 6, Nov. 2004, MEDICON and HEALTH TELEMATICS 2004 "X Mediterranean Conference on Medical and Biological Engineering"
- Di Fabrizio, E., Prasciolu, M. *La nanolitografia come strumento per il controllo metrologico di materiali/strutture artificiali.* NanotechIT Newsletter, Number 2, Oct 2004 , 35-38

Other publications

- Cojoc, D., Ferrari, E., Garbin, V., Cabrini, S., Carpentiero, A., Prasciolu, M., Businaro, L., Kaulich, B., Di Fabrizio, E. *Wave front engineering by means of diffractive optical elements for applications in microscopy* Proceedings of SPIE - The International Society for Optical Engineering 6254, art. no. 625416
- Cabrini, S., Businaro, L., Prasciolu, M., Carpentiro, A., Gerace, D., Galli, M, Andreani, L.C., Riboli, F., Pavesi, L., Di Fabrizio, E. *Focused ion beam fabrication of one-dimensional photonic crystals on Si3N4/SiO2 channel waveguides* Journal of Optics A: Pure And Applied Optics, 2006, **8**, S550–S553
- Mohanty, K.S., Liberale, C., Mohanty, S.K., Degiorgio, V., Cabrini, S., Carpentiero, A., Garbin, V., Prasciolu, M., Cojoc, D., Di Fabrizio, E. *Fiber optic trapping of low-refractive-index particles* Proceedings of SPIE - Progress in Biomedical Optics and Imaging, Volume 6088, 2006, Article number 608812
- Cabrini, S., Liberale, C., Cojoc, D., Carpentiero, A., Prasciolu, M., Mora, S., Degiorgio, V., De Angelis, F., Di Fabrizio, E. *Axicon lens on optical fiber forming optical tweezers, made by focused ion beam milling* Microelectronic Engineering, Volume 83, Issue 4-9, April-September 2006, Pages 804-807
- Cabrini, S., Carpentiero, A., Kumar, R., Businaro, L., Candeloro, P., Prasciolu, M., Gosparini, A., Andreani, C., De Vittorio, M., Stomeo, T., Di Fabrizio, E. *Focused ion beam lithography for two dimensional array structures for photonic applications* Microelectronic Engineering, Volume 78-79, Issue 1-4, March 2005, Pages 11-15

- Schiappelli, F., Kumar, R., Prasciolu, M., Cojoc, D., Cabrini, S., Proietti, R., Degiorgio, V., Di Fabrizio, E. *Design and fabrication of diffractive optical element-microlens with continuous relief fabricated on-top of optical fibre by focused ion beam for fibre-to-waveguide coupling* Japanese Journal of Applied Physics, Part 1: Regular Papers and Short Notes and Review Papers **43** (6B) , June 2004, 3772-3778
- Schiappelli, F., Kumar, R., Prasciolu, M., Cojoc, D., Cabrini, S., De Vittorio, M., Visimberga, G., Gerardino, A., Degiorgio, V., Di Fabrizio, E. *Efficient fiber-to-waveguide coupling by a lens on the end of the optical fiber fabricated by focused ion beam milling* Microelectronic Engineering 73-74 , 397-404
- Cojoc, D., Ferrari, E., Cabrini, S., Malureanu, R., Danailov, M.B., Carpentiero, A., Prasciolu, M., Kumar, R., Businaro, L., Di Fabrizio, E. *Design and implementation of optical tweezers arrays using diffractive optical elements* Proceedings of SPIE - The International Society for Optical Engineering 5477 , 281-292