Integration of Ultrathin Silicon Chips

Von der Fakultät für Elektrotechnik und Informationstechnik
der Rheinisch-Westfälischen Technischen Hochschule Aachen
zur Erlangung des akademischen Grades eines
Doktors der Ingenieurwissenschaften
genehmigte Dissertation

vorgelegt von
Diplom-Ingenieur
Kaspar Hungar
aus Köln

Berichter: Universitätsprofessor Dr. rer. nat. W. Mokwa
Universitätsprofessor Dr. ir. R. W. De Doncker

Tag der mündlichen Prüfung: 15.05.2009

Danksagung

Diese Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Angestellter am Institut für Werkstoffe der Elektrotechnik, Lehrstuhl I der RWTH Aachen.

Herrn Professor Mokwa möchte ich ganz herzlich für die sehr gute Betreuung dieser Arbeit danken. Ich danke Herrn Professor De Doncker für die Übernahme des Koreferats und das meiner Arbeit entgegengebrachte Interesse.

Mein Dank gilt auch Dr. Uwe Schnakenberg, der durch seine Anregungen zum Gelingen dieser Arbeit beigetragen hat.

Für die gute Zusammenarbeit danke ich meinen Kollegen Dennis Ellersiek, Gerald Ganske, Mirko Grabowski, Dr. Mirko Hofmann, Dr. Gökhan Kizilirmak, Christian Koch, André van Ooyen, Roland Schlierf, Dr. Gerd Spanier, Dr. Börge Wessling, Jan Watzlaw und Ingo Klammer. Besonders möchte ich mich bei den Kollegen des AVT-Labors Holger Fassbender und Uli Nolten für die intensive Unterstützung bedanken. Mein besonderer Dank gilt Andreas Buchenauer für die hilfreichen Tipps zu allen Simulationsfragen.

Auch danke ich Dr. Jochen Bauer vom Fraunhofer Institut IMS für die Unterstützung bei der Durchführung von Lötversuchen.

Ich danke allen Studienarbeitern und Diplomanden, die unermüdlich zum Erfolg dieser Arbeit beigetragen haben. Insbesondere danke ich Frank Fuders für die Optimierung der Chipdünnung im Rahmen seiner Diplomarbeit.

Besonders möchte ich mich bei den Kolleginnen und Kollegen des Zentrallabors für Technologie für die gute Unterstützung bei der Probenherstellung bedanken, insbesondere bei Dorothee Breuer, Renate Butterweck, Petra Grewe, Ewa Sekula, Georg Dura, Reinhard Körfer und Achim Malzahn.

Auch möchte ich den Mitarbeitern der mechanischen und elektrischen Werkstätten für ihre schnelle und hilfreiche Unterstützung danken, vor allem danke ich Karl-Heinz Stellmach, Jochen Heiss und Hartmut Pütz.

Von Herzen möchte ich meiner Familie und meinen Freunden für die Unterstützung während meiner Promotionszeit danken. Mein allergrößter Dank gilt meiner Freundin Beate, die mir trotz der vielen Entbehrungen stets zur Seite stand und die nötige Kraft gegeben hat, diesen außergewöhnlichen Lebensabschnitt der Promotion erfolgreich abzuschließen.

Kelsterbach, Mai 2009 *Kaspar Hungar*

TABLE OF CONTENTS

1 Introduction

1.1 Motivation

Modern integrated systems, for example those used in telecommunications, medical and consumer electronics, are characterized by high functionality and complexity. Even though the complexity of these products is increasing steadily, they tend to become smaller and smaller at the same time [1].

As a result, there is a growing need for smaller components that are used to assemble such systems. Alternatively, the components may be attached to flexible substrates (resulting in flexible electronics), thus allowing the system to fit into unoccupied volumes of the product that would otherwise remain empty. Instead of placing rigid components onto a flexible substrate, the circuits themselves may be flexible, like in polymer based thin electronic circuits that achieve their flexibility by combining thinness with soft material properties (for example a small Young's modulus and a large elastic limit). The term *polytronics* (which is short for *polymer electronics*) is generally used to refer to this kind of circuit. An example of such a product is an A3-size electronic paper (or e-paper) that was fabricated on roll-to-roll processing equipment, has a thickness of only 290 µm, and is capable of displaying 4,096 colors [2].

There is increasing research activity in the area of flexible devices that is also reflected in the number of filed patents for flex displays, which rose from only 10 in 1991 to nearly 200 in 2006 [2]. In the future, flexible electronic assemblies will enable new products in the field of mobile telecommunication such as wrist-watch communicators, wearable medical control units, concealed protective electronics, flexible identification and tracking devices for goods, and anti-counterfeit devices [3]. The International Roadmap Committee (IRC), that publishes the International Technology Roadmap for Semiconductors (ITRS) on a yearly basis, presented the following vision on flexible electronics in its 2007 Edition:

> *"Flexible electronics is projected to grow into a multibillion-dollar industry over the next decade and will revolutionize our view of electronics. The unique properties of flexible electronics, such as its compliant structures, ultra-thin profiles, low weight, and potential low cost and high reliability could have enormous impact on consumer electronics, aviation and space electronics, life sciences, military applications and telecommunications. Flexible electronics will enable a broad range of devices and applications not possible today. Smart clothing with integrated electronics and displays will have many consumer, medical and military applications. Realizing the possibility of smart active bandages, and other medical devices such as reconfigurable systems and sensors, micro drug-delivery systems, active/integrated prostheses and massively parallel secure and fault free distributed environmental field sensors will have the potential to change the way people detect and deal with disease and pathogen exposure." [4]*

By going one step further and adding flexible components to flex circuits, entirely flexible systems may be achieved [1]. Even brittle devices like silicon chips can be bent, if their thickness is sufficiently reduced. For example, a silicon thickness of 30 µm results in good flexibility, similar to polymer films, and may lead to the development of new techniques such as *chip in polymer, chip in multilayer printed circuit board (PCB), chip-on-chip, ultra-thin assembly, thin chip embedded in redistribution layer*, and *chip on bent surface* [5], [6].

Therefore, thin semiconductor devices play a major role in the steadily progressing development of highly integrated systems [3], and the use of ultrathin chips with a thickness below 10 µm and of flexible substrates are current areas of research [1]. Chip thicknesses between 40 µm and 200 µm have already been used in smart card fabrication for about 5 years [7]. According to the 2007 ITRS Roadmap [4], the minimum wafer thickness for general products will be reduced from today's 50 µm to 40 µm in 2012 and then remain at this level. For special applications, such as smart cards, however, the current minimum wafer thickness is 20 µm and will decrease to 8 µm in 2015. For 10 µm wafers, manufacturable solutions are either in practice or known, while for 8 µm wafers only interim solutions are known. Some of the major challenges lie in the areas of wafer thinning technology, stress relief, wafer handling technology, singulation, die handling, and thinning of bumped wafers.

Since in standard complementary metal oxide semiconductor (CMOS) chips, the active areas typically have a depth of 1 µm to 2 µm, chip thinning is generally not expected to influence chip performance to a large degree [8]. Thin silicon chips may even perform better than thick silicon chips, because parasitic capacities are reduced (similar to silicon on insulator (SOI) substrates, which typically have an active silicon thickness of 0.2 µm to 5 µm [3]). However, the transistor characteristics in CMOS chips are sensitive to bending-induced stress [9]. Since the magnitude of this effect is known, it may be compensated by adjusted CMOS design rules. Intrinsic stress in layers deposited during the fabrication process and stress resulting from assembly may have a similar influence as the bending stress.

Generally, the chip thickness needs to be reduced as required by the application and viable packaging technologies have to be developed in parallel [10]. Flip chip technology may be used in combination with thin chips, if the pressure exerted by the chip handling tool is reduced to avoid chip breakage [10]. However, flip chip assembly of thinned chips has to overcome a number of problems [11]:

- no flux dip possible due to bumps with small height,
- capillary flow underfill after reflow difficult (due to very small gap height),
- underfill may squeeze out and contaminate chip handling tool,
- die may float after placement or during reflow,
- standard solder masks (100 µm for stencil printed masks) and gold stud bumps (typical height: 40 µm to 80 µm) are too thick for thin chip applications.
- chips may be warped due to intrinsic stress from processing.

Also, wafer bow needs to be minimized in thin silicon applications, because manual and automatic handling of bent wafers during processing is very difficult. For thin chips, the major problem associated with bowing is the placement of the chips onto the lead frame in plastic housings [12].

At the *Institute of Materials in Electrical Engineering, Chair* 1 (IWE1) at *RWTH Aachen University*, where this thesis was prepared, a major research focus is on flexible, telemetric, medical implants. One such implant, called *EPI-RET*, is used to electrically stimulate the retina in blind humans [13]. The implant is continuously being improved and recently 3rd generation prototypes were successfully implanted in several humans for up to 4 weeks [14]. The implant consists of a polyimide (PI) tape (thickness $t_{PI} = 5\ \mu m$) with electroplated gold wires, stimulation electrodes, and a receiver coil. The polyimide used to fabricate the implant (type PI2611) is considered biocompatible [15]. A diode and a capacitor are attached to the flexible polyimide substrate using isotropically conductive adhesive (ICA) and 2 CMOS chips are flip chip attached using gold stud bumps. The device is encapsulated in silicone and then implanted into the human eye ball with the 5 x 5 array of stimulation electrodes placed onto the retina. The receiver coil is placed in the front of the eye and receives information and energy via radio frequency (RF) coupling from a transponder unit outside the eye. The transferred energy and information are used to generate bipolar current pulses that stimulate the retinal nerve cells and lead to visual sensation [16].

A 2nd generation prototype of the implant prior to encapsulation is shown in Figure 1.1. By placing such an implant into a cylindrical form, having about the same radius of curvature as the human eyeball, the inflexibility of the silicon chips becomes well visible (see Figure 1.2).

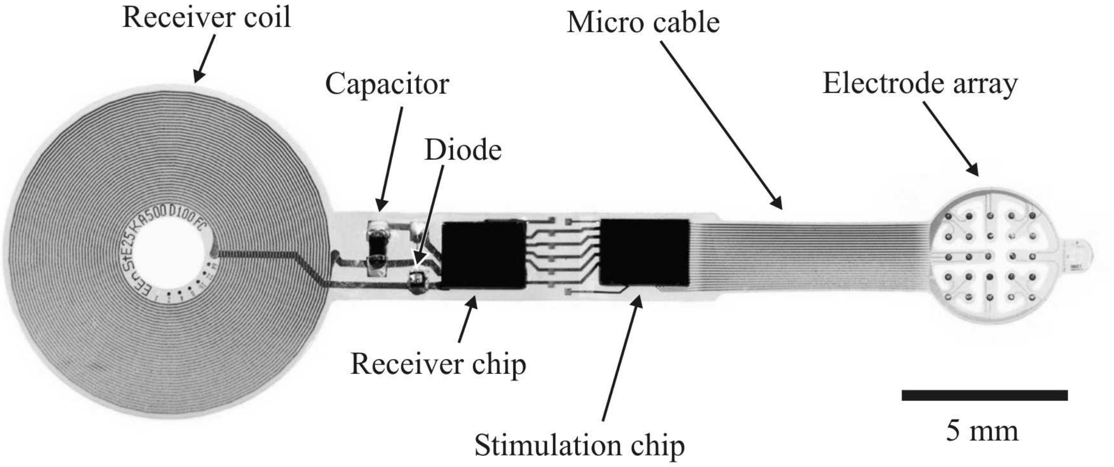

Figure 1.1: 2nd generation prototype of the EPI-RET implant (modified from [17]).

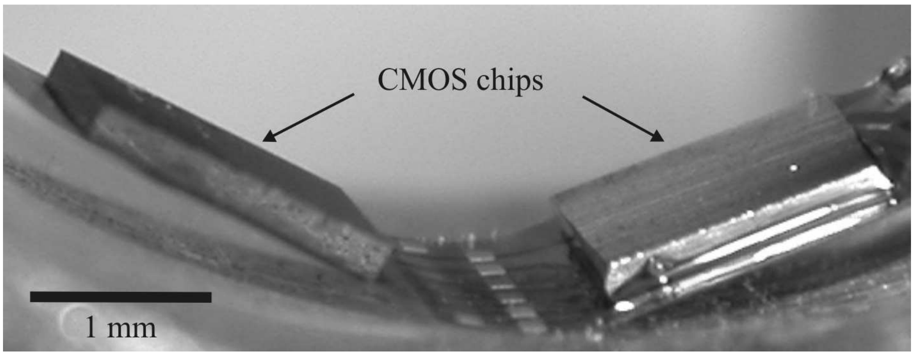

Figure 1.2: Middle section of the implant with two rigid CMOS chips, placed into a cylindrical form (modified from [18]).

By using ultrathin chips for the assembly of the implant, a more flexible device may be obtained. However, when using ultrathin chips, the standard gold stud bumping technology can no longer be used for assembly, because of the substantial breaking risk associated with the large force required for stud bump interconnects. Eutectic gold/tin solder, which is composed of 80 wt.% gold and 20 wt.% tin, is a well known biocompatible solder. Replacing the gold stud bumps with soldered gold/tin interconnects would retain the implants biocompatibility, but may significantly reduce mechanical stress in the chips during assembly. Therefore, thin chip fabrication and gold/tin soldered flip chip attachment of these chips to flexible polyimide tapes is an important area of research. It is the purpose of this thesis to advance the knowledge of the assembly process by fabricating ultrathin dummy chips (using the Dicing-by-Thinning (DbyT) process) and flip chip soldering them to polyimide tapes, similar to the ones used in EPI-RET implants. Because solder deposition in this thesis was based on two consecutive electroplating steps, a bump reflow was required to obtain eutectic solder caps prior to flip chip attachment. However, this bump reflow process results in deformations of the chips and/or substrates which must be minimized to facilitate the subsequent flip chip process. Understanding the relationship between the reflow process and measurable deformations is the main focus of this thesis. A model is devised and verified by comparing simulation results with measured deformations during and after the bump reflow process. Complete systems with functional test structures, such as Daisy chains and Kelvin contacts, were assembled and the mechanical stability of these systems under forced bending with a radius of curvature similar to that of the human eye was successfully demonstrated.

1.2 State of the art

This section gives a detailed overview on research activity relevant to this thesis.

The first part (Section 1.2.1) deals with different aspects of thin chip technology and flexible systems in general. Different areas of application for thin chips are described, with a focus on their production processes and assembly techniques. This part also touches on related areas like polytronics.

The second part is dedicated to the current gold/tin soldering technology. Among the topics are

- advantages of gold/tin solder,
- typical applications for gold/tin solder,
- solder deposition techniques,
- soldering processes, and
- soldering techniques for flex tapes.

The third and final part of this section includes a summary on simulations involving thin silicon chips and/or interconnects.

1.2.1 Ultrathin silicon chips and flexible systems

Applications using thin silicon chips

Thin silicon chips are used in a wide variety of applications with different silicon thicknesses t_{Si}, like memory cards (t_{Si} = 100 µm), integrated circuit (IC) cards (t_{Si} = 50 µm), and IC tags (t_{Si} = 20 µm to 30 µm) [8]. Other major fields of application for thin chips are portable devices, such as mobile phones and notebooks [19]. Thin chips may also be used in medical applications such as retina implants [20]. According to [7], typical chip thicknesses are 70 µm to 200 µm for power applications, 50 µm to 70 µm in stacked memories, and 40 µm to 200 µm in smart cards. Thin silicon chips are also of interest in solar cell fabrication, where the silicon material is the main expense factor. In [21] functional solar cells on 34 µm thick silicon wafers were fabricated. In lab-type environments silicon chips with thicknesses around 10 µm have been achieved [8].

In many fields (such as automotive, microcontrollers, power management, communications, chip card controllers, and memory) a trend towards integration of several chips into one package can be observed, resulting in so-called systems in a package (SiPs) [22].

While in conventional multi chip modules (MCMs) the chips are placed directly onto the substrate and are all connected to the substrate with the same technique (e.g. wire bonding or flip chip technology), in more advanced SiPs the chips may be stacked in order to save floor space and increase the chip density. However, new connection technologies are required for systems with stacked chips. The chips may have different sizes and be stacked such that the bottom chip is largest and the top chip smallest, leaving a small area near the edge of each chip accessible for wire bonding. Another approach uses equally sized chips with spacers between the stacked chips for wire bonding. The bottom chip may be attached by a regular flip chip connection while the other chips of the stack are wire bonded. Or the complete stack may be flip chipped, requiring through vias in all chips except the topmost chip. A good overview of these and other stacking techniques is given in [23]. In all these applications thin chips offer significant advantages, because the total stack height is greatly reduced, reducing packaging volume and facilitating the connection technology. Stacked chips may be used in applications such as the *Global Positioning System* (GPS), personal digital assistants (PDAs), visual phones, gigabyte memory and micro hard drives [8]. A good example of a multi-chip stack is the Toshiba memory MCM, which is composed of 6 chips, each having a thickness of 70 µm. Interposers are placed between the chips to leave space for wire bonding to the substrate.

Thin chip and wafer fabrication

Because front side processing of thinned wafers is extremely difficult, thinning of wafers is typically performed after having created the active CMOS areas [19].

A widespread thinning technique for silicon wafers is grinding with a ceramic wheel embedded with diamonds in combination with water cooling and tape protection of the CMOS on the front side. This technique results in removal rates of up to 5 µm/s and a total thickness variation (TTV) below 3 µm (for 150 mm wafers). The target thickness can be reached with an accuracy of a few microns in an industrial style production. Because of the surface damage resulting from the large pressure between the diamond wheel and wafer surfaces, the silicon surface is stressed, which may lead to substantial wafer bow. To remove this stressed layer, a so-called stress-relief process is carried out, like, for example, a wet etching process (often based on a $HF+HNO_3+H_3PO_4$ etching solution), a chemical-mechanical polishing (CMP) step, or a dry etching step [3].

Although wet etching and CMP offer better surface quality than dry etching [19], plasma etching has been shown to reduce warpage in ground chips by a factor of 6 and increase die strength by a factor of almost 7 [24]. Other sources found that dry etching can lead to increased die strength compared to wet etched surfaces, because dry etched surfaces are smoother, reducing stress concentration effects [25].

Plasma etching is also beneficial at the end of the thinning sequence of the Dicing-by-Thinning process, if the trenches are created by mechanical means, like predicing with a diamond dicing blade. In this case the plasma not only removes damage on the chip's back, but also on its four edges [24]. Dry etching may induce damage, which can, however, be minimized by using an inductively coupled plasma (ICP) with a large ion density and low ion energy [21].

Stress relief etching not only reduces bow, but also increases wafer and die strength [26]. This is especially important if samples are bent, because residual defects in conjunction with stress often lead to breakages [19]. In some applications the removal of intrinsic damage reduces leakage currents and increases minority carrier lifetimes [26].

A conventional slurry based CMP process may lead to rounding of the chip edges. By using fixed abrasion pads with a water based lubricant, the total thickness variation (TTV) and surface roughness can be reduced substantially [27].

Spin-etching is a wet etching technique, where the wafer is rotated quickly while the etchant is sprayed onto the upper side of the wafer. In [10] spin etching of 30 µm of silicon at the end of the DbyT process was used for stress-relief and resulted in very good damage removal and a tensile strength of about 1.1 GPa. Spin-etching has a fairly good homogeneity of typically ± 5% (for removal of 20 µm of silicon) [26].

However, silicon sidewalls of DbyT trenches etch inhomogeneously in spin etching. Therefore, some sidewalls are mechanically stronger after spin etching than others and the breaking strength may vary largely from chip to chip. The average breaking strength increases with increasing stock removal by wet etching after grinding. For example, removing 45 µm compared to only 30 µm increases the average breaking force by over 50% (for chips with a final thickness of 20 µm) [3].

Spin-etching has the disadvantage of being very slow on an industrial scale and should therefore be preceded by a much faster grinding process [26].

Very recently a new production process for thin silicon wafers called *PolyMax* was developed at Silicon Genesis Inc. that is not based on thinning of initially thick wafers, but creates thin wafers directly from the silicon ingot. The process is based on ion implantation and subsequent wafer cleavage without the need for a carrier substrate. Wafers as thin as 50 µm were fabricated, but the process may potentially be used to produce wafers with a thickness of only 20 µm. Without further treatment, the wafers exhibited exceptional strength of about 1500 MPa [28].

Singulation technology

Singulation of thin wafers is another major challenge. Mechanical dicing is traditionally used, but chipping at the side of the die may lead to cracks at a later stage. Therefore, gentler singulation methods, such as laser dicing or plasma etching, are required to obtain strong dies [4].

Wafers with a thickness of only 30 µm may be diced with a mechanical saw. However, mounting of these wafers on dicing tape is very critical, because uniaxial tension may lead to wafer breakage. Also, a stress-relief process is required after dicing to remove damage [11].

In the scope of the Flex-Si project, silicon chips with a thickness of 50 µm were fabricated by wafer thinning and mechanical dicing using standard film frame carriers and dicing equipment. The resulting chips were bent mechanically by attaching them to a cylinder with a radius of 25 mm using double-sided sticky tape. Although all chips initially survived this forced bending test and were electrically functional afterwards, breaking would often occur when the chips were left in the bent state overnight [3].

Other singulation techniques induce less defects in the first place, as for example water jet guided laser, laser-induced stress cleavage, laser cutting, abrasion, and wet or dry etching [8].

Dicing-by-Thinning

In this thesis, the Dicing-by-Thinning (DbyT) process was used for thin chip fabrication. This process is a singulation and thinning technology, which uses a combination of front side dry etching of trenches and backside wafer thinning (for a detailed description of the process sequence see Section 3.2.2). The chip separation is achieved, when the front side trenches are reached during the backside thinning process. The DbyT concept was investigated intensively in [12], where the trenches were created by different techniques, for example by laser ablation, mechanical dicing, or plasma etching. The plasma etched trenches yielded the best results with respect to chip strength. After the mechanical thinning of the 8" wafers, a stress relief process was required. Although wet etching is a well-suited method, it may lead to contamination of the active areas of the CMOS chip. To avoid this disadvantage, stress relief by plasma etching was tested. The wafer was attached to a carrier by a temperature release adhesive tape ($T_{release}$ = 70°C). To reduce the temperature, a remote-plasma source was used, which resulted in very inhomogeneous removal rates across the wafer. By adding an aperture, the homogeneity could be improved, but at about ± 80% still fell short of the required homogeneity of ± 10%. Locally, the process resulted in good quality chips, which could be picked and further processed. It was found that air bubbles enclosed during taping of the wafer expanded during the vacuum plasma process, resulting in local bowing and hot spots, leading to locally increased removal rates. By vacuum lamination, these air bubbles may be eliminated. To avoid contamination of the active areas of the chips during stress relief wet etching, the trenches can be filled with resist. In a technique described in [12] a wafer was attached face-down onto a carrier substrate by a two-layered stack consisting of a temperature release and an ultraviolet (UV) release tape. After thinning, the tape was removed from the carrier substrate by applying heat. The tape with the attached thin chips was then flipped onto a standard dicing tape and the chips were transferred face-up to the dicing tape by removing the UV tape. If the chips had a topography larger than 1 µm (for example because bumps were added in advance), an additional soft tape was used on the active side of the chips during thinning. Two alternative temperature resistant wafer attachment techniques were also tested and shown to be compatible with temperatures of 430°C. The first technique was based on a water soluble silicate adhesive and the second technique used polyimide dots for gluing the wafer to the carrier. To release the wafer after thinning, the silicate adhesive was dissolved in water, while the polyimide dots were removed by UV laser ablation through the quartz carrier substrate. All these wafer handling techniques did not allow for electrical backside contacting, which is often required during wafer level chip testing. By replacing the carrier substrate with a support ring, glued to the front side of the wafer by silicate adhesive, successful handling and electrical testing of thinned wafers was demonstrated [12].

When dry etching of trenches was used during the DbyT process, the chip outline did not have to be rectangular as in mechanically diced chips. Round chip corners and other chip shape variations could be fabricated [3].

The DbyT concept may also be used in conjunction with mechanical dicing by creating the front side trenches with a dicing blade that does not penetrate the entire thickness of the wafer [19]. This technique is called *Dicing Before Grinding* (DBG) and requires additional treatment, such as wet etching to remove damage on the chip edges [8].

Recently, the DbyT concept was successfully implemented to fabricate 12 µm thick CMOS image sensor chips. For thinning and handling purposes, the wafers were attached to a carrier by thermoplastic glue [29]. And in [30] a so-called *silicon e-cube* was presented that has a size of 30 x 30 x 35 μm^3 and was fabricated using the DbyT concept.

Adhesive-based thin chip interconnects

A major contributor in the field of thin chip technology has been the InnoSi project [12]. The project covered many topics such as wafer thinning, singulation, handling of thinned wafers and chips, and assembly. Thin chips were contacted to a substrate by different techniques. In a technique called *isoplanar contacting* the chip is attached to the substrate face-up and the conductor lines are fabricated by screen or stencil printing silver-filled conductive adhesive along a path from the chip pads, across the chip edge and to the tracks on the substrate. By using a 50 µm poly ethylene terephthalate (PET) substrate, 10 µm die attach adhesive, a 20 µm chip, and 10 µm conductor lines, a total system thickness below 100 µm was achieved. Isoplanar contacting was found to be a well suited technique for thin chip applications, because of the limited step height between the surface of the thin chips and the substrate [5].

Adhesives may also be used for flip chip interconnects. Mainly three types of adhesives are available [6]:

- isotropically conductive adhesive (ICA),
- anisotropically conductive adhesive (ACA),
- non-conductive adhesive (NCA).

ICA typically consists of a thermosetting polymer filled with silver particles, which results in sufficient electric conductivity. The application methods range from screen and stencil printing to dispensing and dipping techniques [6].

In ACA the conductive particles usually consist of pure nickel, gold-coated nickel, gold-coated polymer or silver-coated glass. The particles typically have a diameter of 3 µm to 15 µm and form part of the current path between the substrate and chip pads, when the chip is pushed against the substrate. In the sideways direction the conductive particles do not touch and are hence electrically isolated from each other. ACA is often used in low cost electronics and may be snap cured resulting in only a small thermal load [6].

NCA is not conductive by itself and must be penetrated by conductive structures (such as stud bumps) to achieve an electrically conductive connection [6].

Regarding flip chip technology in combination with conductive adhesives, 3 different techniques were shown to work well for contacting thin chips [12]:

- ACA filled with 8 µm gold spheres: a thin chip ($t_{Si} \approx 50$ µm) was attached to a substrate, resulting in a contact resistance of about 1 Ω to 2 Ω.
- ICA on the pads and NCA sideways between the pads: a fairly low contact resistance of about 1 Ω was achieved.
- ICA in conjunction with a geometric effect: a conductive adhesive with fairly low conductivity was used to coat the substrate area, where the chip was placed. If the minimum distance between two pads on the chip was fairly large and at the same time the thickness of the conductive adhesive was very small, the contact resistance would be 3 to 4 magnitudes smaller than the isolation resistance for purely geometric reasons.

In [10] a 25 µm thick flip chip was presented that was attached onto a polyimide tape using copper stud bumps and ACA.

Since polymers are naturally flexible, they allow for very small bending radii in thin chip applications. In [3], for example, a 25 µm thick transponder IC, housed between two polyimide tapes ($t_{PI} = 25$ µm each), was presented. The chip was attached to the polyimide tape by various adhesives and mechanical bending with a minimum radius of 3 mm was achieved.

In [9] a strain gauge based on the strain-induced change of the transistor characteristics in ultrathin silicon chips ($t_{Si} = 10$ µm, size ≈ 3 x 4 mm²) was developed. After CMOS processing, individual thin chips were obtained by mechanical dicing of the wafers and a subsequent thinning process based on lapping, wet etching, CMP, and dry etching. Although the CMP step resulted in a very good surface quality, thicknesses below 20 µm could not be achieved due to chip breakages. Therefore, a dry etching step was carried out on the 20 µm thick chips to remove an additional 10 µm. The surface quality of the dry etched chips is less good, but, as a result of the reduced thickness, the flexibility was significantly increased. 4 chips were attached face-down to the backside of a polyimide tape ($t_{PI} = 9$ µm) using an underfill. The polyimide tape was prepared with holes underneath the contact pads of the chips, so that the chips could be connected by ICA to the gold tracks on the front side of the polyimide tape. Since a non-conductive native oxide formed on the aluminum pads, they could not be directly contacted by ICA. Therefore, the contact pads were previously coated with a conductive 3-layer under bump metallization (UBM) of zinc, nickel and gold (using the so-called *zincate process*). The mechanical properties of the fabricated strain gauges were sufficient for gluing to a torsion bar with a radius of 15 mm. The system withstood a pressure of 5 N/cm² that was required for gluing and also survived strains of up to 0.2% during operation. It was found that the threshold voltage of the transistor did not change during thinning, but that the drain current was reduced by about 5%. After assembly, the drain current had decreased by another 20%, which was explained by the higher operating temperature resulting from the thermal insulation of the attachment glue and polyimide tape.

Intrinsically conductive polymers (that means polymers which are naturally conductive and do not require filling with conductive particles) have fairly large resistivities of about 5 Ωcm [3]. Due to the small interconnect length in flip chip connections and resulting small contact resistance, they may still offer a viable alternative to metal-filled adhesives.

Thin chip assembly and integration

In [20] a silicon chip with a thickness of 20 µm, that was fabricated using the DbyT process, was bonded onto copper tracks on a polyimide tape (t_{PI} = 25 µm) by thermode bonding of immersion solder bumps. By laminating a second polyimide layer over the chip and the first polyimide layer, a chip-in-polymer package was fabricated. Sandwiching the silicon chip between two equally thick polyimide tapes was beneficial with respect to reliability during bending, because the chip lay in the neutral axis and was exposed to less stress [11].

In [31] the fabrication of CMOS chips with a thickness of 65 µm using a similar thinning process as in this thesis is described. The chips were bent using a 4-point bending unit and the transistor characteristics were measured, showing that the transistors could withstand stress as large as 350 MPa with only a small shift in the transistor characteristics.

In the scope of the *European Flex-Si project* [3], [11] an integrated module board (IMB) was fabricated that was based on a flexible FR4 board (t_{FR4} = 100 µm) and a silicon chip (t_{Si} = 50 µm) with electroless Ni/Au (e-Ni/Au) bumps. A hole having the size of the chip was drilled into the board and, using standard processes, copper tracks were added to the board. An adhesive tape was laminated over the hole and copper tracks of the FR4 board. From the opposite side, the chip was placed through the hole, face-down onto the sticky surface of the tape. By encapsulating the chip into a molding polymer that remained elastic after curing, a completely flexible system was fabricated. Finally, the tape was removed and the tracks on the board were connected to the input/output (I/O) bumps of the chip by an additive printed wiring board process based on electroless copper plating.

Using another strategy, thin silicon chips (t_{Si} = 50 µm) were successfully attached by solder flip chip technology to a flexible substrate [3], [11]. The wafer bumping was carried out before the thinning process. The bumps had a height between 5 µm and 20 µm and were based on electroless Ni/Au, followed by a dip solder transfer process. Although this process achieved overall thin solder thicknesses, it had the disadvantage of a large bump height variation, resulting in inhomogeneous pressure and removal rates during the subsequent grinding. Therefore, a mechanical buffer tape was attached to the front of the wafer before mounting onto the chuck of the grinding machine. The thinning process was carried out on 6" silicon wafers and consisted of a rough grinding step down to 150 µm, followed by a fine grinding and polishing step to a final thickness of 50 µm. The wafers with a thickness of 50 µm were handled with manual tweezers after thinning. Because it was difficult to fill the narrow gap between the chip and substrate with capillary flow underfill after assembly, a pre-applied no-flow fluxing underfill was used instead.

In the European project SHIFT (Smart High-Integration Flex Technologies) an ultrathin chip package (UTCP) was presented. It consisted of a thinned silicon chip (t_{Si} = 20 µm to 30 µm, edge length: 4.9 mm), which was attached to a rigid carrier supported polyimide tape ($t_{PI} \approx 20$ µm, type: PI2611) using benzocyclobutene (BCB). Then, a second polyimide layer was added by spin coating and this layer was opened by laser drilling. Electrical contacts were fabricated by photolithographic patterning of a sputtered metal layer consisting of 50 nm TiW + 1 µm Cu [32], [33].

In another European project called *Hidden Dies*, the integration of thin silicon chips (t_{Si} = 50 µm, max. edge length: 10 mm) in polymer was analyzed [33]. The aluminum pads of the chips were coated with 5 µm copper stud bumps. The chips were then attached face-up onto a standard FR4 board and covered with resin coated copper (RCC), consisting of 80 µm of resin on an 18 µm copper carrier. The RCC resin and copper were opened at the chip pads by laser drilling and electroless and electroplated copper was added to fill these vias. Finally, the RCC copper was patterned. To minimize deformations, an identical RCC laminate was simultaneously added on the back of the FR4 board, resulting in a total thickness of the device of 300 µm. More detailed descriptions of the production processes for both the SHIFT and Hidden Dies projects are given in [34].

A possible solution for reducing stress in soldered interconnects without using underfill is the use of flexible electrical leads. Also, pillar-like bumps may be used to relieve mechanical stress [4]. An example, where such technology was implemented, is given in [34]. A matrix of 5 x 5 metal pillars (diameter = 3 µm, pitch = 6 µm, height = 10 µm) was fabricated on a single 30 x 30 µm² bond pad for bonding through pre-applied epoxy glue. This technique was demonstrated to work well for connecting a 50 µm thick silicon chip to another 140 µm thick chip.

Thin silicon in other applications

In the following paragraphs applications and research areas are presented, which go beyond standard thin chip to flex substrate attachment techniques.

Wire bonded chip stacks have been used in recent years in cells phones. By using a technique called *cascade wire bonding*, stacks consisting of 8 chips have been fabricated. Each chip was wire bonded directly to the substrate or to a lower chip, which in turn was connected to the substrate or a lower chip, and so on [4].

In [36] a monolithic 5 x 5 array of CMOS compatible silicon chips connected by flexible ultrathin silicon bridges was presented. The individual chips were about 1 mm across and 300 µm thick. The silicon bridges were either made of meandering or straight, perforated silicon, had a thickness between 1 µm and 20 µm and were up to 1.2 mm in length and 0.8 mm in width. The structures exhibited remarkable bending capabilities of 90° per bridge for a bridge thickness of 5 µm and of 180° per bridge for a bridge thickness of 1 µm. The devices were fabricated using SOI wafers, which were patterned by plasma etching on the front and anisotropic wet etching on the back. After removal of the buried oxide layer, the described flexible silicon chip matrix was obtained [36].

Instead of connecting rigid device islands by silicon bridges, stretchable metallic conductor lines may be used. Such conductor lines were fabricated on an elastomeric polydimethylsiloxane (PDMS) substrate (t_{PDMS} = 1 mm) by evaporation and patterning of a thin gold layer (t_{Au} = 25 µm to 500 nm). The gold wires survived stretching of the PDMS of up to 60%, however, resulting in an increase of the resistance by a factor of about 3. By mechanically pre-stretching the PDMS by 10% to 25% prior to gold evaporation and releasing the PDMS from the fixture after evaporation, wavy gold wires could be formed that exhibit increased stretchability of about 90% without a significant increase in resistance [37].

For packaging of microelectromechanical systems (MEMS) a technique based on a 20 µm thick silicon caps was demonstrated in [38]. A thin silicon wafer was temporarily attached to a pyrex glass carrier that was pre-structured with cavities. Then, cavities and tethers were wet etched into the silicon and lead/tin solder was deposited onto the silicon fields around the cavities by electroplating. The resulting silicon cap was soldered onto a substrate, thereby enclosing a MEMS device. The silicon cap was released from the glass carrier by simply breaking the silicon tethers [38].

In [39] wavy stretchable silicon structures were fabricated by transferring ultrathin silicon ribbons (l_{Si} = 15 mm, w_{Si} = 5 µm to 50 µm, t_{Si} = 20 nm to 320 nm) onto pre-stretched PDMS (t_{PDMS} = 1 mm to 3 mm) and then relaxing the PDMS. The monocrystalline silicon ribbons were fabricated from SOI wafers by patterning the top silicon and then removing the buried oxide using concentrated hydrofluoric acid. To prevent the silicon ribbons from being washed away in the etchant, the ends of the ribbons were not patterned, but were left connected to the wafer. After bonding the top silicon to a pre-stretched PDMS substrate the carrier wafer was removed, leaving the silicon ribbons attached to the PDMS. Upon releasing the pre-strained PDMS a sinusoidal silicon structure with periodicities between 5 µm and 50 µm and amplitudes between 0.1 µm and 1.5 µm formed. Active components, such as pn photo-diodes and metal oxide semiconductor field effect transistors (MOSFETs) based on Schottky-barriers, may be integrated in the top silicon of the SOI wafer prior to ribbon patterning. To keep the active side on top, the active component ribbons were first transferred to an unstrained intermediate PDMS carrier and then transferred to a pre-stretched PDMS layer as before. The MOSFET and pn diode remained functional even at strains of about ± 10%. In [40] the process was further developed to fabricate free-standing wavy silicon structures with defined buckling characteristics. To achieve this, the PDMS surface was treated to obtain thin lines with strong adhesion interspersed with wide lines of weak adhesion. Upon relieving the pre-stretched PDMS a wavy silicon structure formed with a periodicity defined by the distance between the lines of strong adhesion. The samples fabricated with this process allow forced bending with a radius of curvature of 5 mm, a stretchability of 100%, and compressibility of 25%. Similar processes were also used to create two-dimensional wavy surfaces (having a herring bone pattern) that could be stretched in 2 directions and hence allowed, for example, spherical deformations.

Radio frequency identification and smart cards

There is an ongoing trend towards ever thinner smart cards [41]. Regarding minimum card thickness, one has to differentiate between production on an industrial scale and in the lab. The thinnest mass-produced smart card available to date was developed by Texas Instruments Inc. and had a total card thickness of 345 µm [42]. For radio frequency identification (RFID) applications much thinner systems are available as, for example, an RFID label with a total thickness of 100 µm (including a thin silicon chip) [10].

Hitachi has a history of fabricating ultra small RFID chips. In 2001 a 400 x 400 µm² RFID chip with an external antenna was presented and two years later, the antenna was integrated into the chip [43]. In 2003 a 300 x 300 µm² RFID chip with a thickness of 60 µm was demonstrated and in 2006 RFID chips with an area of 150 x 150 µm² and a thickness of only 7.5 µm followed [44]. A year later the chip size was reduced even further to 50 x 50 µm² [45]. However, these examples were only prototypes. The smallest RFID chip manufactured on a large scale was 400 x 400 x 60 µm³ and was used in about 20 million electronic tickets during the World Exposition in Aichi, Japan in 2005 [44].

Polytronics

An alternative to thinned silicon chips is the use of printable electronics [46]. Many different components such as thin film transistor circuits (TFTCs), displays, batteries, sensors, and microphones could in principle be printed onto a single substrate, resulting in low cost devices. The reel-to-reel (R2R) processing technique that is used in the manufacture of polymer electronics offers high throughput at very low operating cost [47]. The major disadvantage of polytronics is the small carrier mobility (typically 0.2 cm²/Vs for printable soluble organic compounds, and a maximum of 3.5 cm²/Vs achieved by Merck in 2007) leading to fairly low operating frequencies [48]. For example, in [49] a polytronic ring oscillator with an operating frequency of 119 Hz was presented. The oscillator was fabricated in a reel-to-reel process and the transistors had a feature size of 15 µm, allowing the integration of up to 12 transistors per cm² at a cost of about 0.03 €/transistor. A feature size reduction to 5 µm is under development. Recently, the fabrication of batteries in an R2R process with a capacity of 2 mAh/cm² was demonstrated [50].

RFID tags are a major potential market for polytronics, and estimates for RFID demand are as large as 10 trillion (10^{13}) devices p.a., assuming a price below 0.01 US$ per tag. An example for polymer electronics is the pressure sensitive skin presented in [51]. The integrated pressure sensors and transistors could withstand a bending radius of 2 mm. However, having a cycle time of 30 ms, these polymer transistors were far slower than conventional silicon CMOS transistors. RFID tags can be fabricated with complete biocompatibility as, for example, the edible RFID tag that was recently patented by Kodak [48].

One particular field of application for flexible circuits is wearable electronics [4]. However, due to the nature of textile use, simple bending functionality does not suffice to create reliable electronics. A certain stretchability is also required. Electrical conductors made of meandering copper tracks embedded in an elastic matrix are shown to be a suitable technology for interconnect lines, allowing cyclic stretching in one direction of up to 40% [2]. However, for stretchability in two directions, as required for use in textiles and in other complex three-dimensional (3D) shapes (like spheres), new challenging designs are needed. Possible solutions are using a conductive polymer or a non-conductive polymer matrix highly filled with conductive particles [4]. A detailed overview on current technology for smart textiles is given in [52]. A very good source for further reading on flexible circuit design and technology is [53].

1.2.2 Gold/tin soldering

General notes on solder connections

Solder interconnects typically consist of a solder bump with an under bump metallization (UBM) and a substrate metallization. Apart from its main purpose to form an electrical connection, the bump also serves as a structural link and dissipates heat during operation. Apart from solder, the bump may be made of metallic studs or conductive adhesives. The UBM prevents corrosion of the chip metallization due to ions in the encapsulation or migrating solder (primarily migrating tin). Electroless nickel UBMs typically have a thickness of 5 µm and result in mechanically stable interconnections by the formation of intermetallic compounds (IMCs) between nickel and tin [6].

To achieve long term reliability, brittle Ni/Sn alloys should be minimized and the UBM should exhibit little internal stress [21]. Underfill is often used to locally constrain the thermal mismatch caused by the difference of the coefficients of thermal expansion (CTEs) between chip and substrate, further increasing reliability [54].

Creep, which is defined as a plastic deformation occurring after prolonged exposure to stress, is often encountered in soft solders. Since the amount of creep in different solders depends on stress, strain rate, and temperature, one solder may perform better in one application and less good in another application compared to other solders [55].

Instead of using a solder alloy, a single material like gold may be used to form solid state bonds. Gold stud bumps on the chip are pressed against the top gold layer of the substrate pad. Thermal and/or ultrasonic energy is added to complete the bonding process. The resulting contact height is typically in the range of 50 µm to 100 µm [6]. This fairly large connection height is a clear disadvantage of stud bump connections in thin chip applications [19]. Also, the force required during bonding may be fairly large, for example

0.5 N/bump (for 10 minutes at 150°C) [56]. These mechanical forces and constraints are locally concentrated around the bumps and may break the chip.

In flip chip connections, the chip is placed face-down with the chip pads onto the substrate pads. Soldering of the bumped pads of the substrate and/or chip is the most common connection technique (as compared to adhesives or stud bumps) [3]. One advantage of flip chip connections is that chip mounting and electrical interconnection are achieved simultaneously. Other advantages are compactness, high interconnect density, and improved electrical performance (due to the small lead length) [56]. Flip chip connections can be used in chip scale packages (CSP), leading to substantial weight and package size reductions [57]. Flip chip technology may also be used in conjunction with flexible substrates (for example in liquid crystal display (LCD) drivers), allowing the assembled device to be bent and hence to occupy otherwise unused space in the final product [57]. Saving space is important in portable electronic devices such as mobile phones and pagers, where small and light-weight features are preferred by consumers.

In reflow soldered flip chip connections, the bumps are usually electroplated or stencil-printed onto electroless Ni/Au UBMs at the wafer level and, after dicing and placement onto the printed circuit board, soldered in a belt furnace [3].

Due to differing CTEs of the chip and substrate materials, mechanical shear stress occurs at the flip chip interconnects after cool-down from the soldering temperature. To reduce the stress in the bumps, an underfill may be dispensed to fill the empty space around the interconnects. Having a CTE similar to that of the bump material, the underfill evenly distributes the thermal stress over the entire chip area, reducing shear stress in the interconnects. The most common underfill is capillary flow underfill, which is dispensed close to the gap between chip and substrate and fills the entire gap by capillary flow. Afterwards a thermal curing step is required to harden the underfill [6].

Apart from reducing thermal stress, the underfill also increases shock and vibration resistance in applications such as automotive, mobile phone, and other portable devices [58]. In order to achieve good reliability, the underfill must have good adhesion to the substrate and chip [57]. Flux residue can be critical for the adhesion of the underfill, reducing reliability of the flip chip connections.

Advantages of gold/tin solder

For a solder to be acceptable for industry applications, it has to exhibit desirable material characteristics, for example regarding melting temperature, wettability, electrical and thermal conductivity, mechanical strength, creep and thermal fatigue resistance, corrosion resistance, manufacturability, and cost [59]. Gold/tin solder, which is mostly used at its gold-rich eutectic composition (80 wt.% Au + 20 wt.% Sn, corresponding to 71 at.% Au + 29 at.% Sn), has many favorable properties like:

- flux free bonding, good wetting, low creep, good corrosion resistance, large tensile strength (270 MPa compared to 40 MPa for eutectic lead/tin solder), slow IMC phase growth at elevated temperatures, and high melting point [60],
- large yield strength at elevated temperatures (165 MPa at 150°C compared to 4 MPa for lead/tin solder) [61],
- good fatigue resistance [62] (e.g. 75 cycles between -20°C and +100°C were achieved in silicon chips that were flip chip soldered to polyimide tapes; by adding an underfill the number of failures was reduced to zero even after 2000 cycles [63])
- compatibility to different UBM materials (e.g. Ni, Au, Pt, Pd, Pd/Ag [60], or Cr [64]), low intermetallic growth rates when used over Ni, Pd, or Pt [65],
- good thermal shock resistance (e.g. 40 cycles between -196°C and +160°C were achieved in silicon chips soldered onto alumina substrates) [64],
- good thermal conductivity of 0.57 W/mK [64] (this is important in applications where the solder connection serves as a heat sink; in laser bar soldering for example a temperature increase of 8°C during operation may halve the expected operating life [66]),
- good electrical conductivity [67], [68] (in [69] a value of $\sigma = 17.35 \cdot 10^{-6}\ \Omega\text{cm}$ is given),
- good self-alignment capabilities in flip chip soldering [70],
- biocompatibility [42],
- conforming to the European Directive 2002/95/EG (RoHS, Directive on the Restriction of the Use of Certain Hazardous Substances in Electrical and Electronic Equipment), which prohibits the use of lead containing solders in most applications since July 2006 [71],
- a thin gold coating may be deposited onto the underlying nickel layer prior to solder deposition to avoid oxidation of the nickel; this is an advantage compared to lead/tin interconnects, where a gold coating resulting in a gold content between 2% and 7% in the solder may lead to the formation of brittle gold-tin intermetallics (specifically $(Au_{0.5}Ni_{0.5})Sn_4$) during thermal aging [72],
- when using a nickel barrier underneath gold/tin solder, a stable IMC, Ni_3Sn_2, forms at the gold/tin to nickel interface during reflow [57],
- similarly, gold may be deposited on top of the tin to avoid oxidation [73].

The fairly high eutectic temperature of 280°C is primarily an advantage, resulting in interconnects that are stable at elevated temperatures. This allows higher operating temperatures or an additional soldering process with a different, lower melting solder on the same substrate (see [42] for an example of a two-step soldering process based on gold/tin and tin/copper solder). On the other hand, the thermal load on the device is larger during soldering compared to many other lead-free solders such as tin/copper (99.3Sn0.7Cu, T_{melt} = 227°C), tin/silver (96.5Sn3.5Ag, $T_{melt} = 221°C$), or tin/silver/copper (96.5Sn3Ag0.5Cu, $T_{melt} = 218°C$) [59]. However, compared to other widely used hard solders (for example eutectic gold/silicon, $T_{melt} = 363°C$), gold/tin can even be considered low melting [64].

As a hard solder, gold/tin mainly shows elastic strain when deformed, often leading to larger stress than in soft solders (which typically show plastic deformation with small stress, but increased fatigue and creep). Because mechanical stress concentrates at voids, leading to weakened bonds and solder cracking, voids pose a problem in gold/tin bumps and need to be minimized. Also, interconnects with voids have a larger thermal insulation resistance, leading to increased operating temperatures. Voids may be caused by the segregation and formation of materials such as oxides, carbon, and silicon on the molten solder, which prevent the liquid solder from forming bonds with the pads. Scrubbing may be used to break down the surface film, however, this also disturbs the molten solution and consequently reduces the homogeneity of the melt and resulting bond [64]. Besides the mentioned mechanical and thermal disadvantages, voids also lead to the degradation of electrical characteristics [68].

Applications for gold/tin solder

Gold/tin solder is widely used in applications such as optoelectronics, high frequency, MEMS, hermetic packaging, and in high temperature applications [60]. Gold/tin is the preferred solder for precision optical components (because of their sensitivity to high temperature creep), MEMS devices (because of their intolerance to fluxes and their residues), and biomedical devices, which require corrosion resistance along with good electrical, thermal, and mechanical properties [74]. Package lids and heat sinks are also often solder attached using gold/tin [75]. Specific examples of gold/tin solder applications are:

- mounting of laser bars [76],
- hermetic wafer level packaging of RF device [75],
- MEMS packaging: soldering of a cap wafer with Ni+Au+Sn+Au metallization onto a substrate wafer with Ni+Au metallization [77],
- biocompatible chip mounting [42],
- flip chip attachment of a 21 x 12 mm² GaAs chip with 2608 I/Os onto ceramic substrates as part of an optical switch [78],
- flip chip attachment of silicon chips onto BCB substrates with gold/tin interconnects with diameters between 15 µm and 50 µm; the larger diameter bumps (50 µm) were reflowed before soldering to obtain good bonds, while the smaller diameter bumps (15 µm to 30 µm) were aged instead by high temperature storage prior to soldering [79],
- laser soldering through transparent flex tapes [57],
- tape automated bonding (TAB) inner lead bonding and die bonding [63],
- flip chip on flex in commercial hearing aids [63].

The latter example is fairly similar to the application targeted in this thesis and is described in more detail in the following. The chip attachment in the hearing aids was based on gold/tin flip chip connections to copper tracks on a polyimide tape. The chips were electroplated with gold bumps. A thin tin layer was added to the gold bump and/or the copper tracks. The optimal thickness for tin on the gold bumps was found to be between 2 µm and 3 µm. Using tin on both the gold studs and the copper tracks increased the wettability and resulted in fewer voids. Gold/tin was used, because it did not remelt in the subsequent (lower temperature) soldering of passive devices to the same substrate. The soldering was carried out consecutively for 7 flip chips on the same flex print by pulsed-heat thermode bonding with a heating rate of 100°C/s. Only the top part of the gold bump formed an intermetallic compound with the melted tin, leaving a defined stand-off height between chip and substrate. Because the tin and gold needed to interdiffuse to form reliable bonds based on the eutectic composition, a minimum temperature of 309°C was required, resulting in fast solid-liquid interdiffusion and preventing the formation of the brittle δ and ε phases. Below 309°C, slow solid-state diffusion took place with formation of the δ and ε phases (for more details on the intermetallic reactions during soldering of gold/tin see Section 2.1). In order to improve wetting, a non-residue flux (2% adipic acid in isopropanol) was used [63].

Au/Sn solder deposition techniques

Gold/tin solder may be applied in various forms and using different methods, for example preforms, solder paste, evaporation, and electrodeposition [62]. For micro bump fabrication, thick resist lithography ($t_{resist} = 20$ µm) followed by gold/tin solder evaporation and a lift-off may be used [80]. Preforms are the most common gold/tin solder application method [81], however, with the downsizing of connection areas, preform handling becomes increasingly difficult [67], [68].

Historically, preforms are used in hermetic seals. For wafer level packaging, solder pastes are more suitable. However, due to the potential for contamination, cleaning may be required during processing [75].

Solder pastes may be applied by printing, dispensing, or pin transfer and generally have low production costs, but often exhibit problems with void formation. The number of voids can be reduced by minimizing the generation of gases during soldering through optimized preheat temperatures and hold times. Also, reducing the oxygen content in the metal powder used to fabricate the solder paste and optimizing the amount of activator in the flux may reduce voids. Finally, avoiding tin-rich phases by careful control of the solder paste composition can reduce voids [67].

In [29] solder paste was successfully stencil printed onto a 55 µm thick silicon wafer that was attached to a mobile electrostatic chuck and then reflowed. One disadvantage of stencil printing is that certain design rules must be fulfilled to achieve good paste transfer (examples: stencil mask opening area / wall area > 0.66 and line width / stencil thickness > 1.5) [82]. When screen printing solder spheres, a metal content of 60 wt.% to 90 wt.% is typically used in a thixotropic carrier and a fluxing agent is added. Out-gassing of volatile compounds may condense on the device surface or be trapped inside the package [65]. Among the solder paste processes, stencil printing is the most cost effective and widely spread. However, in 2004 minimum pitches of 200 µm were feasible in mass production and pitches of 100 µm to 150 µm in the lab [6].

Another technique for solder application is reactive ion sputtering. In [83] a sequence of Cr-Sn-Au was deposited by sputtering and used in wafer level MEMS packaging. Gold/tin may also be co-sputtered (i.e. simultaneously sputtered) as shown in [84]. An alternative to sputtering is evaporation, which is a very clean process and offers good thickness and compositional control [62]. The deposition may be achieved by alternating electron beam evaporation of gold and tin or by evaporation of a gold/tin composite [67], [68]. Pulsed laser deposition (PLD) has also been used successfully to deposit gold/tin [85].

Electroplating of gold/tin can be done using different approaches. Either the elemental materials are plated sequentially in a two-step process [78] or they are plated simultaneously from a single electrolyte [62]. A detailed analysis of the layer growth dynamics for gold electroplating of gold on tin is given in [86]. The initial gold deposit reacts with the tin grains resulting in an $AuSn_4$ coating, followed by gold clusters forming on the grain ridges that keep growing and eventually connect to each other. According to [57], the height variation across a 4"-wafer for electroplated layers is typically around ± 1 µm, allowing this technique to be used for flip chip connections.

In simultaneously plated gold/tin layers, the tin does not oxidize when exposed to air [62]. However, this technique requires very precise control of the plating parameters to obtain the correct eutectic composition [87]. The use of complexing agents is imperative to narrow the reduction potential gap between Au and Sn ions and to stabilize the plating bath [88]. The eutectic plating bath lifetime may also be very short (for example 3 days in [62]). The simultaneous plating technique may be modified by alternatingly plating thin layers of Au_5Sn and AuSn from the same bath by varying the current density [89]. In this case, the final composition of the stack only depends on the plating time ratio for the gold-rich and tin-rich phases. When electroplating tin, the bath temperature has to be kept below 35°C to avoid oxidation of the stannous tin (Sn^{2+}) to stannic tin (Sn^{4+}) [62].

In general, electroplated gold/tin offers a superior dimensional control at small feature size, a tight pitch, and a high feature density compared to solder preform placement or solder paste printing. However, electroplated gold must be free of organic contaminants that could adversely affect solderability. When plating gold over tin, the gold plating bath may be contaminated with tin ions, because tin readily dissolves in acids and bases. When using the opposite plating order, the gold dissolution in the tin plating bath is negligible. For tin plating, methane sulphonic acid baths are preferred, because the carbon content is lower, resulting in good reflow characteristics on a gold top layer. A small increase of the gold content in the solder, for example due to a gold surface finish on the UBM, also yields a strong solder joint, however with a slightly increased melting point [65].

A problem often encountered in pure tin deposits is the formation of tin whiskers, which may lead to short circuits and device failures [90]. The mechanism behind whisker formation is the deposit's tendency to decrease internal energy (for example mechanical stress) through recrystallization and grain growth. Whiskers can range in length from a few microns up to several millimeters [90].

Meniscus bumping is yet another technique for applying gold/tin solder. The wafer with electroless Ni coated aluminum (Al) pads is dipped into the liquid solder, which adheres only to the UBMs. Since the bump height variation is fairly large (± 5 µm), this technique is not suitable for general flip chip applications, but well suited for flip chip on flex [57].

In [91] an ink jet technique was presented, which was used to apply melted solder balls with a minimum diameter of about 90 µm by piezo actuation. The maximum device temperature was 360°C, making it suitable for gold/tin eutectic solder ball deposition.

Gold/tin solder heating methods and soldering profiles

The peak oven temperature for gold/tin soldering typically lies between 330°C and 350°C, with dwell times at maximum temperature between 15 seconds and 3 minutes [83], [61].

In [78] electroplated gold/tin bumps were reflowed by dipping the whole wafer in a heated liquid medium at 180°C and then heating the liquid to 285°C, before removing the wafer from the bath.

In thermode bonding, thermal energy is transferred from the heated chip carrier to the chip by physical contact, resulting in fast heating rates and low thermal load for the substrate and chip [6]. Typical process cycle times are between 5 s and 20 s [3]. Therefore, this technique can be used for cheap temperature sensitive flex substrates [57].

Even faster heating rates and a thermal load confined to the solder joint area can be achieved by laser heating [57].

Another laser-based gold/tin soldering technique is the SB2-Jet process [92]. A eutectic gold/tin solder sphere is held inside a tapered capillary and melted by a laser pulse. Propelled by a slightly larger pressure inside the capillary, the solder sphere slips out of the capillary and lands on the target pad, solidifying upon contact with the pad.

In [93] a eutectic gold/tin bond was achieved by bonding a gold-tin-gold coated chip at 240°C. At this temperature, the molten tin dissolved the adjacent gold, increasing the gold concentration in the melt and thereby raising the melting temperature of the gold/tin composite which consequently solidified at 240°C (so-called isothermal solidification). Due to the large interdiffusion speeds of gold and tin, the gold concentration was further increased by solid state diffusion, leading to a remelting temperature above 280°C.

Under bump metallizations

A typical under bump metallization may consist of 3 layers: an adhesion layer, a diffusion barrier, and a layer, which prevents oxidation of the underlying barrier layer. Nickel is often used as a diffusion barrier for tin-containing solders [94]. The intermetallic compound (IMC) growth characteristics for gold/tin on nickel at elevated temperature are favorable compared to other barrier metals such as platinum and palladium [95].

The mechanical properties of solder joints not only depend on the composition and thermal history, but also on the reactions at the solder-to-substrate interface. For example, in a sequential Au_5Sn and AuSn electroplating process, the interaction between the solder and nickel UBM depends on the sequence of deposition (Au_5Sn deposited after or prior to AuSn). The reason for this lies in the different solid solubilities of nickel in the Au_5Sn and the AuSn phases, which are 1% and 20%, respectively [96].

When using a thick gold base underneath the tin, a separate barrier metal underneath the gold base may not be required. In [78] the thick gold base was only partly consumed by the tin during reflow. The amount of consumed tin, and hence the solder composition and solder cap height, depended on the geometry of the bump and the reflow conditions. If the tin layer was fairly thick, a high temperature storage had to precede the reflow to obtain a planar interface between the solder and gold base after reflow. This pre-reflow aging step lead to the growth of the Au_5Sn phase. The growth characteristics of the Au_5Sn layer as a function of storage temperature and time were analyzed in [57]. Under certain circumstances the Au_5Sn layer may even act as a natural diffusion barrier [97]. However, one disadvantage of this approach is that the Au_5Sn layer may consume the entire eutectic gold/tin layer on top during bump reflow [61]. Because of its high melting point of 519°C, the Au_5Sn phase does not remelt in the subsequent gold/tin soldering process, rendering the reflowed bumps useless.

By using a defined diffusion barrier underneath the solder, the available solder volume, resulting cap height and solder composition are predefined, independently of the reflow profile. By depositing a gold base underneath the diffusion barrier, shear forces can be reduced, because pure gold is a fairly soft material [98].

Atmospheric conditions during soldering

In [99] different atmospheric conditions during gold/tin soldering were compared with respect to process time, yield, and wetting. It was found that an active atmosphere works far better than either a pure nitrogen or hydrogen atmosphere. For example, a residual oxygen concentration of 8 ppm (parts per million) in a nitrogen atmosphere resulted in a bond yield of only 75%. By using pure hydrogen, the yield was increased to 98%. The authors argue that the thickness of the oxide layer is determined by the atmosphere and is responsible for the bond yield. In [100] hydrogen was found to be suitable for removing oxides in gold/tin soldering, if the temperature exceeded 350°C. In [101] the oxidation of gold/tin during soldering was analyzed. While the native oxide formation on gold/tin is a self-limiting process, the oxidation during soldering depended on the partial oxygen pressure. Good bonds were achieved in a vacuum soldering oven at 0.1 Pa. However, in order to achieve good alignment, the oxides had to be reduced, for example by introducing hydrogen during the soldering process. The hydrogen may also be activated in a plasma. In [102], for example, bumps made of tin-rich gold/tin eutectic were successfully reflowed in a hydrogen plasma. Fluorine-based plasmas (such as CF_2Cl_2, CF_4, or SF_6) are not as suitable, because reaction products may remain on the solder surface and become corrosive upon contact with atmospheric humidity [80].

In [63] the wettability of a tin pad metallization was improved by adding 2% adipic acid in isopropanol prior to soldering. Formic acid may also be added to reduce tin oxidation and promote alignment. In [103] a detailed analysis of the alignment performance of lead/tin bumps as a function of formic acid concentration in the nitrogen carrier gas is presented. Very good performance was achieved for formic acid concentrations above 0.7 vol.%.

Tin oxides in the melt lead to voids in the bump [104]. The tin oxide thickness t_{SnOx} can be calculated using

$$t_{SnOx} = \sqrt{D_0 \cdot t \cdot e^{\frac{-Q}{RT}}}\,, \tag{1.1}$$

where $D_0 = 3.7 \cdot 10^{-18}$ m²/s is the volume diffusion coefficient, t the time, $Q = 33 \cdot 10^3$ J/mol the activation energy, $R = 8.314$ J/mol·K the molar gas constant, and T the temperature in units of Kelvin [63]. For storage at room temperature for one day, a tin oxide thickness of 0.65 nm is calculated. At the soldering temperature of 330°C, the same oxide thickness is reached after only 80 s.

Bonding with gold/tin in an inert atmosphere without using any flux is also possible [105]. The bonding was carried out between two complete 2"-wafers and the tin layer was protected from oxidation by a thin electroplated gold layer. However, about 2% of the soldered area contained voids.

Chip alignment

The distance between the chip and substrate after soldering is the result of the geometry based minimization of the surface energy of the molten solder [99]. The self-alignment effect is caused by the same effect of surface energy reduction. Using solder deposited in Vernier patterns (lines of varying distances), the self-alignment capabilities of gold/tin solder were analyzed in [99]. A chip placement accuracy of ± 10 µm was found to be sufficient for a self-alignment accuracy of ± 3 µm, when a hydrogen or active atmosphere was used and the oxygen concentration was around 8 ppm. In [101] a self-alignment accuracy of 2 µm was achieved, by using hydrogen as a flux substitute during soldering. However, a time-span of 2 minutes was required for sufficient oxide removal. When using conventional flux in combination with punched out ribbon preforms, an accuracy of ± 1 µm could be achieved in all directions (x, y, and z) [106]. The self-alignment capability decreased with decreasing aspect ratio (height/width) of the bumps.

Alternative interconnect technologies

In [107] a flip chip solder connection based on a plastic core solder ball (PCSB) was presented. The connection was based on a small plastic sphere with a diameter of 170 µm that was coated with a 5 µm copper layer and 10 µm of plated lead/tin solder. Because of their flexible plastic cores, these interconnects could survive up to ten times as many thermal cycles as regular solder connects. Although tested with lead/tin solder, these interconnects might also work well with gold/tin solder. By replacing the polymer core with air, the so-called *aircore bump* was fabricated [108]. The air bubble inside the solder had a diameter of about 300 µm.

For small pitches (below 150 µm), conventional solder printing is not suitable. A cost effective alternative technique called *immersion solder bumping* (ISB) was presented in [6]. The complete wafer was moved through the liquid solder, which adhered only to the pads. The solder height of a few microns depended on the pad size and surface tension of the molten solder. A pitch of 40 µm has been realized with this technique.

In [15] an alternative connection technique between polyimide tapes and silicon chips (with standard thickness) was presented. The polyimide tape was prepared with holes that were aligned to the contact pads of the chip. A gold ball, such as used in ball wedge bonding, was pressed onto (and eventually partly through) the hole in the polyimide tape onto the connection pad on the silicon chip. By applying heat and ultrasonic actuation, a mechanical and electrical fixation between the tape and chip was achieved.

Special considerations for thin chips

The shear strain resulting from thermal mismatch is inversely proportional to the bump height. Since in thin chip applications the total thickness of the system is often minimized, stress is generally more severe. By using smaller chips, the stress from thermal mismatch can be reduced [6].

In [109] stencil printed tin-rich gold/tin eutectic solder (composed of 95% tin and 5% gold) was successfully used to connect thin chips (t_{Si} = 100 µm) to 50 µm thick polyimide tapes with 17 µm thick copper tracks. Bonding was carried out in a hydrogen atmosphere at 150°C for 3 minutes, achieving nearly void-free interconnects with a diameter of 200 µm and a thickness of 20 µm. The same research group achieved similar bonding results with sequentially electroplated gold and tin [110]. Bonding with gold/tin solder in an inert atmosphere was also shown to be almost void-free if the tin layer is protected from oxidation prior to soldering by a gold top layer [111], [86].

Regarding solder ball attachments to thin silicon, there have been some recent advances. In [29] for example, successful solder ball bumping onto 55 µm thick silicon wafers has been demonstrated. In preliminary trials of this thesis, eutectic gold/tin solder balls with a diameter of 80 µm were successfully attached to gold pads on 20 µm thick silicon chips using the SB2-jet process (see Section 1.2.2, page 27 for a process description).

Soldering to flex tapes

In [97] standard silicon chips with gold stud bumps (t_{Au} = 18 µm) were flip chipped onto PI flex tape (t_{PI} = 25 µm to 40 µm) with copper tracks (t_{Cu} = 8 µm to 18 µm), which were coated with an electroplated layer of tin (t_{Sn} = 0.5 µm to 1.0 µm). The substrate was heated to a maximum temperature of 180°C to avoid oxidation of the tin coating, while the chip was heated to a temperature between 300°C and 380°C. The substrate and chip were joined by thermocompression bonding with a bonding time of only 4 s. The ζ phase that formed between the copper tracks and the eutectic gold/tin served as a natural diffusion barrier, slowing down the phase growth between the copper and the eutectic and hence reducing void generation. A capillary flow underfill with very fine fillers was successfully introduced in the gap between chip and substrate. The resulting bonds exhibited very good mechanical stability with zero failures in all reliability tests. The chips with a size of 4.5 x 6 x 0.5 mm^3 contained 182 I/Os with an approximate pitch of 100 µm and a bump width below 75 µm.

1.2.3 Simulations based on the finite element method

General notes on simulations

In most simulations of solder joints based on the finite element method (FEM) the main objective is the influence of thermal stress on the reliability of the joints. Historically, tin/lead solder joints were often analyzed regarding their performance during cool-down after soldering or during thermal cycling (see [112] for an example of such an analysis using the ANSYS simulation software). The quality and reliability of simulation results depends on the accurate representation of the relevant physical characteristics of the simulated device in the model. Therefore, complex systems require large models, which in turn require excessive computational resources and/or simulation time. With the continuous advances in computer technology, ever more complex devices can be simulated. However, according to the 2007 ITRS Roadmap, one major bottleneck for complete SiP reliability simulations is seen in the excessive simulation time [113]. In general, finding a simple, but correct model is a key issue in successful FE analyses.

According to [53], many companies are beginning to perform mechanical finite-element modeling of flexible circuits to validate the design with respect to long-term reliability before committing the product to manufacture. Finite element analysis (FEA) modeling tools are widely available and can lead to substantial savings by eliminating iterative prototyping, provided the modeling parameters are wisely selected.

Gold/tin solder joint simulations are very scarce, especially in conjunction with thin chips. A few examples of such or related simulations are given in the following.

Simulations involving thin silicon chips

The stress resulting from flip chip soldering of a silicon chip ($t_{Si} \approx$ 150 µm) to a GaAs chip ($t_{GaAs} \approx$ 100 µm) was simulated in a 3D FEM analysis in [114]. Each interconnect had a height and diameter of about 35 µm and consisted of the solder material (either gold/tin or tin/silver) in between two gold stud bumps. Because the features size varied between 100 nm (UBM and passivation thickness) and 10 mm (size of the entire package) an FEM model was devised that consisted of a local model that was integrated into a global model. Apart from the stress resulting from the soldering process, the influence of thermal cycling was also analyzed. Due to the large CTE difference between Si and GaAs, stress in excess of 1 GPa was observed in the gold/tin interconnects and the underlying UBMs, leading to UBM delamination. In the viscoplastic tin/silver solder much smaller stress values (in the range of 70% to 80% of the gold/tin values) were observed, leading to fatigue failures instead.

In the framework of the European project *Hidden Dies* (see Section 1.2.1, page 18) the heat dissipation during operation and, to a limited extend, package deformations caused during cooling after resin hardening of a PCB with an integrated thin chip (t_{Si} = 50 µm) were simulated. An FE simulation was carried out to analyze the influence of thermal vias, which connected the chip to the back of the FR4 board, thereby serving as heat sinks for the embedded chip. To reduce computational time, an axisymmetric two-dimensional (2D) simulation was devised. In non-axisymmetric parts of the system (like near the chip edge) averaged material properties were required to model the system correctly. Parametric modeling was used, which allows varying parameters within certain bounds without remeshing. In another simulation, a 3D model was devised, using the simulation tool PATRAN in conjunction with a local-global approach, resulting in a model size of 330,000 nodes. This model was used to analyze package deformations and stress concentrations at vias due to the cooling process after resin hardening [115].

In [114] the same group presented a simulation model that comprised a stack of two thin chips and made use of 3 symmetry planes, thus reducing the model size by a factor of 8. It was found that the mechanical stress on the vias, caused by the cool-down, was insufficiently described if stress from the production process (here: via copper plating at 40°C) was not taken into account.

In [116] FEM simulations of the bow of silicon chips stressed by the thinning process were presented. The silicon thickness was varied between 4 µm and 1000 µm and an almost linear proportional dependence of the radius of curvature on the silicon thickness was found. Substantial bow reduced the stress in the damaged silicon surface. For example, assuming a 25 µm silicon chip with a stressed layer of 1 µm thickness (with an intrinsic stress of σ = 500 MPa) resulted in a radius of curvature of 100 mm and a stress reduction by 10% to 450 MPa.

In [117] silicon chips with a thickness of 48 µm were fabricated using the DbyT process and then characterized in a 3-point bending test. The force required for bending was measured and calculated by a 2D FEM simulation. Good agreement of about 5% between the simulated and measured displacement force was achieved.

In [118] an implantable flexible pressure sensor was presented that included a silicon chip (t_{Si} < 50 µm) mounted face-up onto a polyimide tape by thermocompression bonding. The electrical connections were achieved by electroplating and a second polyimide layer (type PI2611) was added to fully encapsulate the chip. Numerical analyses were carried out to calculate the maximum stress during bending of a 30 µm thick chip attached to 30 µm thick polyimide film. The silicon was assumed as isotropic, resulting in an error of only a few percent. The radius of curvature was varied between 5 mm and 50 mm in the simulation and, as expected, the von Mises stress was found to be inversely proportional to the radius of curvature. At a radius of 10 mm, the von Mises stress in the silicon was about 0.6 GPa, which was considerably lower than the tensile strength of silicon of 1.2 GPa. Silicon samples with a thickness of 25 µm, 40 µm, 80 µm, and 100 µm were bent until breaking occured and the radius of curvature just before breaking was later determined from captured optical images. A radius of curvature of 3.6 mm was found for the thinnest sample at the breaking point. These breaking experiments were simulated assuming a silicon tensile strength of 1 GPa. The radius of curvature was larger by 20% to 100% in the measurement than in the simulation. A possible explanation for these deviations was the residual stress from the thinning process, even though wet etching was the last step of the thinning process.

Simulations involving interconnects

In [119] the influence of voids on gold/tin interconnect reliability was simulated by FEA. It was found that the strain maximum appeared next to the voids. Because damage developed along phase boundaries, cracking only partially coincided with the strain maxima.

In [120] the influence of the type of used underfill on thermal stress in soldered flip chip assemblies for smart cards and the resulting mean cycles to failure as a function of creep strain were investigated by FEA. For hard underfill, a stress maximum was found at the chip edges between the silicon and underfill, while for soft underfill, the interconnects took most of the thermal stress. Different material models were used for the various materials, for example, the metal contact plate was modeled as elastic-plastic, while the lead/tin solder was modeled using time dependent (creep) and time independent plastic strain rates.

In [121] the stress distribution in solder interconnects was simulated in an FEA and correlated to experimental observations of crack lines inside the interconnects. The simulations were carried out for different bump geometries (truncated spheres and cones) and for varying distances to the neutral point (DNP). All bumps embedded in hard underfill showed very similar equivalent creep strain. For soft underfill, however, the corner bumps (corresponding to a large DNP) showed much larger equivalent creep strain than the center bumps. Also, the equivalent creep strain differences between the two bump geometries were more pronounced. Although the qualitative agreement between measured and simulated strains was good, the absolute values were lower by a factor of about 5 in the simulation. This difference between simulation and measurement was traced back to an anisotropic coefficient of thermal expansion of the underfill.

In finite element analyses of thermal stress in soldered lead/tin interconnects, the bump material was simulated using rate dependent plasticity, the copper tracks were considered elastic-plastic and the polymers (for example the underfill) were considered viscoelastic with temperature dependent elastic constants [54]. It was found that for flip chip connections without underfill, the stress in the interconnects was proportional to their distance to the neutral point and that interconnect failures typically started at the interconnects farthest away from the neutral point. When using underfill, the stress was more homogeneously distributed. By adding SiO_2 to the underfill, the CTE may be adjusted to be similar to the solder material, further homogenizing the stress. Because these silica particles often settle, a two-layered underfill model had to be used in the FEA.

1.3 Objectives of the work

Today, thin silicon chips are used in many applications and there is a growing trend toward thinner and more flexible silicon chips. Understanding and predicting deformations during different processing steps in such systems is important in order to improve the sample design and fabrication process. In this thesis, deformations during the reflow and soldering process are analyzed and the following topics are addressed in detail:

- use of the Dicing-by-Thinning process to fabricate chips with a silicon thickness between 3.5 µm and 45 µm,
- electroplating of the chips with gold tracks, gold-nickel UBMs, and gold/tin solder prior to the Dicing-by-Thinning process,
- measurement of temporary deformations during bump reflow,
- measurement of permanent deformations after bump reflow and flip chip soldering,
- development of suitable FEM models for predicting these deformations,
- analysis of the dependence of deformations on the chip design and chip thickness,
- measurement and simulation of local silicon deformations under and around individual bumps,
- assembly, mechanical and electrical characterization of flip chip attached ultrathin-silicon-to-flex-polyimide systems.

By using the well known Dicing-by-Thinning process ultrathin chips are fabricated. These chips are prepared with different geometries of gold tracks, under bump metallizations, and gold/tin solder bumps in order to analyze temporary deformations during the bump reflow and permanent deformations remaining after the subsequent soldering process. The metal structures also serve the purpose of electrical characterization after soldering. One of the major objectives of this thesis is to establish a simulation model that can predict these deformations, but which is considerably less complex than the fabricated samples, for example by using linear-elastic material models and many geometric approximations. In this scope, the dependence of deformations on the chip design and their relation to the chip thickness is of special interest. Local deformations, occurring under and around individual bumps, will also be analyzed. Finding a suitable measurement technique and simulation model for these rather small deformations is part of the work.

To show the viability of flip chip attachments of thin silicon chips to flexible substrates using gold/tin solder, complete systems were assembled and characterized electrically and mechanically. Therefore, fabricating robust thin test chips and establishing an assembly technique for these chips based on gold/tin solder flip chip interconnects was a major technological task.

In summary, the results in this thesis enhance the understanding of chip deformations during reflow and soldering and, consequently, can facilitate future design processes, allowing deformations to be minimized by choosing beneficial geometries and/or materials.

The structure of this thesis in described in the following. In Chapter 2, a phenomenological view on several topics relevant to this thesis is given. The gold/tin soldering process is described with a focus on gold/tin interdiffusion and the formation of intermetallic compounds in adjacent layers of pure gold and tin, like the ones used in this thesis. This is followed by a section on general bending and deformation concepts and a section on the influence of crystal damage on the flexibility of silicon chips. After describing the different sample designs, fabrication processes for thin silicon chips and polyimide tapes, and the gold/tin soldering process development in Chapter 3, a detailed description of the developed simulation model follows in Chapter 4. In Chapter 5, the major results of this thesis are presented. This also includes measurement techniques for deformations and thicknesses of deformed chips. A detailed section follows that relates measured deformations during and after the bump reflow or soldering process to simulation results. The adjustments of the simulation model are described and possible explanations are given for the observed deviations. Then, in Chapter 6, the viability of the proposed assembly process is analyzed. For this purpose, shear tests were performed on solder interconnects between silicon chips with standard thickness. Functional, soldered systems based on thin chips and flexible polyimide substrates were also fabricated and the bump yield and contact resistances were measured using Daisy chains and Kelvin contacts. Finally, mechanical tests prove the robustness of the soldered systems in forced bending situations. This thesis concludes with a summary of the major accomplishments and a section on possible future research areas in Chapter 7.

2 Fundamentals

2.1 Gold/tin soldering

The intermetallic reactions at the interface between pure gold and tin layers at room temperature and during reflow are fairly complex. It is the aim of this section to describe these dynamic processes in detail.

Properties of the Au-Sn system

The two solder constituents in this thesis are the elements gold and tin. For eutectic gold/tin solders, 80 wt.% gold and 20 wt.% tin are required. Volume-wise the required percentage of tin is much larger (almost 40 vol.% Sn), because its specific weight is about 2.6 times smaller than that of gold. More physical data of elemental gold and tin is given in Table 2.1.

Table 2.1: Properties of elemental gold and tin [61].

	Gold (Au)	Tin (Sn)
Position in PTE [1)]	11. sub-group	4. main group
Atomic number	79	50
Molar mass [g/mol]	196.97	118.69
Density [g/cm³]	19.32	7.31
Atomic radius [nm]	0.144	0.151

1) PTE = periodic table of elements

The properties of gold/tin alloys are best described by the gold/tin phase diagram, which represents the phases that exist in a gold/tin alloy and the equilibria between these phases as a function of temperature and gold-to-tin ratio. A phase is defined as an area in the phase diagram where the mechanical and physical properties do not change abruptly [61]. Generally, a phase is formed from the pure metals and/or their intermetallic compounds and usually has a varying fractional chemical valence. Often, the intermetallic compounds and pure metals are also referred to as phases. The thermodynamic equilibrium gold/tin phase diagram is shown in Figure 2.1.

A gold/tin solder interconnect may contain several phases. Therefore, the mechanical characteristics of a solder joint are defined by the kind, number, concentration, and amount of phases that are present in the joint [61].

The gray dots and numbers in the diagram represent distinct gold/tin concentrations (in at.%) while black numbers represent phase boundary temperatures. At the bottom and top of the diagram the tin concentration is given in atomic (at.%) and weight (wt.%) percentages, respectively. Due to its 5 intermetallic compounds and their interaction with each other a large number of phases exist. The IMCs are included in the phase diagram as vertical boundary lines between adjacent phases and are marked with Greek letters (β, ζ', δ, ε, and η; order by increasing tin content). The melting points of gold (1064°C) and tin (232°C) are quite different. They are included in the diagram as the values of the liquidus line at the pure metal concentrations. The notification in parentheses (e.g. (Au)) indicates that the material is not pure. Gold, for example, can absorb up to 7.1 at.% of tin without changing its crystal structure, or, in other words, the homogeneity range of gold ranges from 0 at.% to 7.1 at.% Sn. The region of liquid gold/tin of various concentrations is marked with the capital letter L. All regions bordering on the L-region along the liquidus lines are a mixture of solid and liquid material. All other regions below are purely solid and are separated from the solid-liquid regions by so-called solidus lines. During cooling of elemental tin, it changes its crystal structure at 13°C from tetragonal β-Sn to diamond-cubic α-Sn in an allotropic transformation [122], [123].

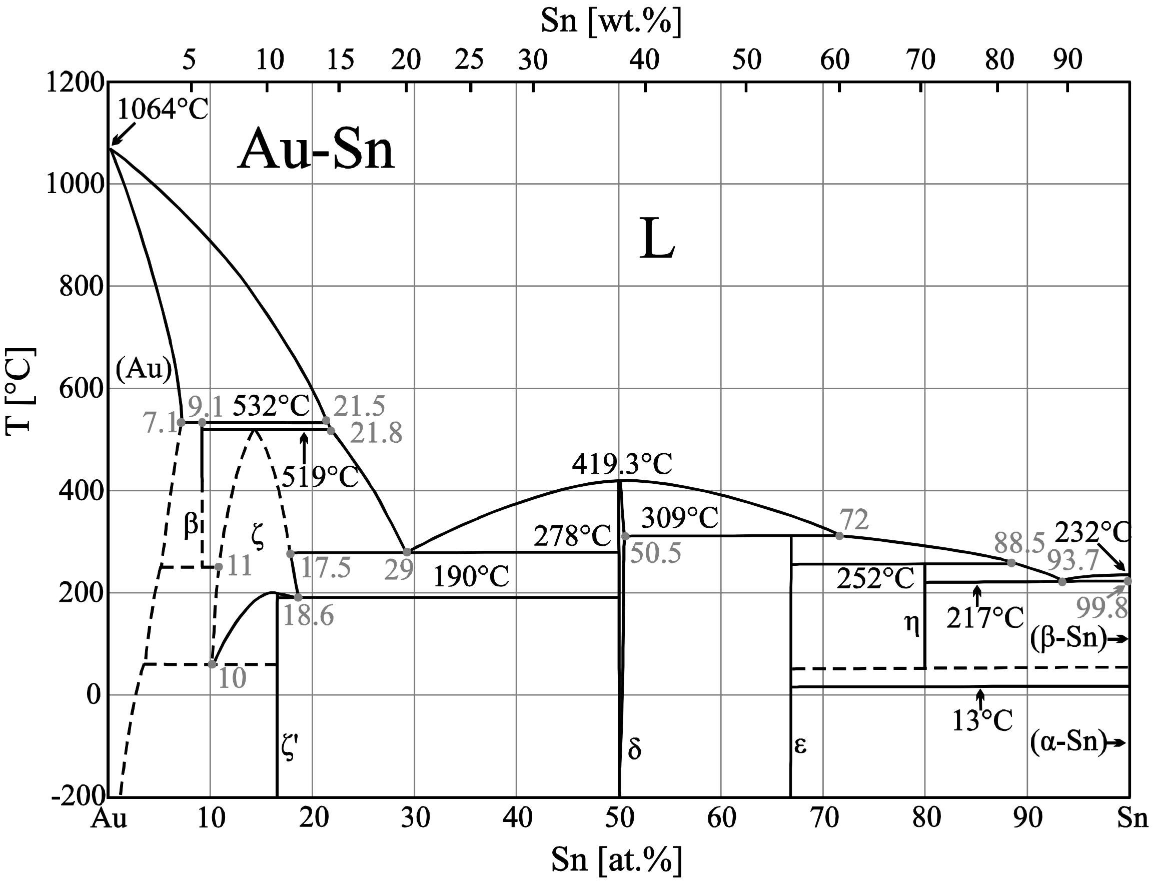

Figure 2.1: Phase diagram of the Au-Sn binary alloy system (modified from [122]). The gray numbers indicate Sn concentrations in atomic percent.

For several phases (marked with dashed lines in the gold/tin diagram), the properties and boundaries are still not well determined [123]. A newer diagram was calculated by Liu in 2003 [124]. Because experimental verification for this diagram is missing and because the differences to the diagram by Okamoto and Massalski in Figure 2.1 are only minor with respect to the described gold/tin soldering process, the diagram in Figure 2.1 is used. The gold-rich eutectic's melting point is given as 278°C by Okamoto and Massalski, while in [60] a differential scanning calorimetry analysis is presented, which indicates a melting temperature of 281°C. In recent publications and technical data sheets a value of 280°C is mostly given, which is therefore used throughout this thesis [98], [125], [126], [127].

The terminal solid solutions of gold and tin and the 5 intermetallic compounds have a number of different crystal structures and lattice parameters. For completeness, this data is given in Table 2.2.

Table 2.2: Crystal structure and lattice parameters a, b, and c of elemental gold, tin, and their intermetallic compounds [128].

Phase	Structure	a [nm]	b [nm]	c [nm]
Au	fcc (face-centered cubic)	0.40784	0.40784	0.40784
$Au_{10}Sn$ or β	hex (hexagonal)	0.2904	./.	0.9536
	hcp (hexagonal close-packed)	0.2908 1)	./.	0.4785 1)
Au_5Sn or ζ'	hex	0.292 2)	./.	0.477 2)
AuSn or δ	hex	0.43218	./.	0.55230
$AuSn_2$ or ε	orth (orthorhombic)	0.6867	0.7004	1.1784
$AuSn_4$ or η	orth	0.6446	0.6487	1.1599
β -Sn	bct (body centered tetragonal)	0.3182	0.3182	0.5832

1) At 12.7 at.% Sn.
2) At 16.7 at.% Sn.

The Au-Sn system has two eutectic points, at each of which the liquid L solidifies into a microcrystalline, lamellar texture of 2 distinct constituents upon cooling. The crystallization (i.e. the solidification) takes place simultaneously for both constituent phases. The formation of the lamellar microstructure is a consequence of the transition from complete liquid miscibility to limited solid miscibility of these 2 constituents [129]. The two gold/tin eutectics are compared in Table 2.3.

Table 2.3: Composition, constituent phases, and melting temperature for the tin-rich and gold-rich gold/tin eutectic.

Eutectic	Composition [wt.%]		Constituent phases	T_{melt} [°C]
	Gold	Tin		
Tin-rich	10	90	η + (β-Sn)	217
Gold-rich	80	20	ζ' + δ	280

The tin-rich eutectic has the advantage of a comparatively low melting point of 217°C and reduced cost due its low gold content of only 10 wt.%. However, it contains the brittle η-phase ($AuSn_4$) which often leads to reliability problems after soldering of gold-coated UBMs with tin-containing solders [59]. In [123] it was shown that under certain circumstances strong interconnects can be produced using non-eutectic tin-rich solders (specifically the 20Au80Sn and 5Au95Sn compositions (values in atomic percent)). The joints were shear tested after soldering according to Military Standard (MIL-STD) 883C, which resulted in breakage of the silicon. However, reliability testing (thermal cycling, high temperature storage, or elevated temperature + humidity storage) was not investigated. However, the gold-rich eutectic is generally described as more robust and stable in the literature, having many advantages (see page 22).

There are also 4 peritectic points and one congruent point in the gold/tin phase diagram. All transitions and additional data are listed in Table 2.4.

Table 2.4: Distinct phase transformations in the gold/tin system, according temperatures, concentrations c_{Sn}, and schematic reaction formulas.

Transformation type	Temperature [°C]	c_{Sn} [at.%]	Reaction
Eutectic	279	29	$L \leftrightarrow \zeta + \delta$
	217	93.7	$L \leftrightarrow (\beta\text{-Sn}) + \eta$
Peritectic	532	9.1	$L + (Au) \leftrightarrow \beta$
	519	13.7	$L + \beta \leftrightarrow \zeta$
	309	72	$L + \delta \leftrightarrow \varepsilon$
	252	88.5	$L + \varepsilon \leftrightarrow \eta$
Congruent	419	50	$L \leftrightarrow \delta$

Eutectic solder formation from pure tin and gold layers

In this section the metallurgical reactions during bump reflow or soldering of a gold-tin sandwich with combined eutectic composition is described. The eutectic is formed from the pure metal layers in a complex dynamic process.

Diffusion in the Au-Sn system at room temperature

In this thesis, a sandwich structure of gold and tin is plated onto a nickel diffusion barrier. Taking into account the target composition of 80 wt.% Au + 20 wt.% Sn and the densities of gold (19.32 g/cm³) and tin (7.31 g/cm) the deposited gold layer has to be about 52% thicker than the tin layer. This is fulfilled by first electroplating 5 µm of gold, followed by 3.25 µm of tin.

Hugsted et al. [130] found that in evaporated gold/tin interdiffusion couples the gold diffuses into the tin layer along the tin grain boundaries by the very fast interstitial mechanism and the AuSn phase is then formed at the tin grain boundaries. Because the tin grain boundaries are less densely packed (due to crystal defects and edge dislocations, which are often found where different crystal orientations meet), the speed of the interstitial diffusion of gold along the tin grain boundaries is further increased [61]. Once the grain boundaries are saturated with AuSn, the gold continues to diffuse into the tin grains by interstitial bulk diffusion, forming the $AuSn_4$ phase. The tin grain size in Hugsted's experiments was varied between 50 nm and 500 nm by changing the substrate temperature during evaporation. He found that in tin layers with smaller grains the formation of the AuSn phase is faster, probably due to the increased number of tin grain boundaries. The grain size in the electroplated tin layers in this thesis is between 500 nm and 1000 nm (see Figure 2.2) and hence slightly larger than in Hugsted's evaporated samples. Therefore, less tin grain boundaries are present, most likely resulting in a slower AuSn phase growth. The phase formation at room temperature and during bump reflow, however, should be similar. The average (per area) AuSn concentration increases with a square root of time dependence. Hence, the AuSn phase formation is believed to be limited by the gold diffusion through the newly formed AuSn phase at the tin grain boundaries. The $AuSn_4$ phase growth, however, shows a linear time dependence, indicating a reaction limited process [130].

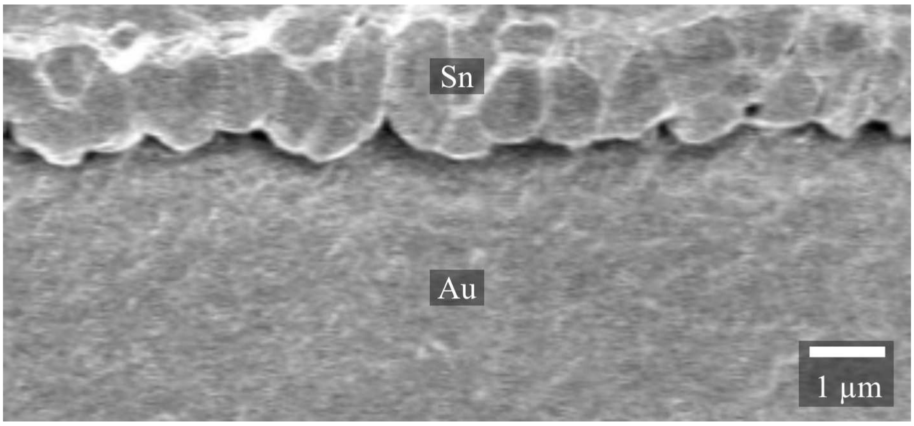

Figure 2.2: Scanning electron microscope (SEM) side view image of an electroplated Au/Sn bump. The tin grains are well visible.

The described AuSn and AuSn4 phase formation takes place at room temperature due to the large number of vacancies at the tin grain boundaries. Whether the AuSn phase forms immediately upon the diffusion of gold along the tin grain boundaries is a matter of debate. Nakahara et al. [131] argue that at the start of the diffusion process $AuSn_4$ is first formed at the tin grain boundaries, because the large vacancy density at the tin grain boundary allows the $AuSn_4$ phase to form with almost no disturbance of the crystal structure of the pure tin. On the other hand, the surplus of gold at the tin grain boundaries would generally favor the formation of AuSn. However, this would require the nucleation and growth of AuSn islands. Against these arguments Matijasevic [132] and Buene [133] suggest that the formation of AuSn precedes the formation of $AuSn_4$. The combined process of AuSn and $AuSn_4$ phase formation is reaction limited in the beginning, however, once the AuSn and $AuSn_4$ layers have reached a sufficient thickness, the reaction becomes diffusion limited, because the gold atoms have to pass the newly formed AuSn and $AuSn_4$ layers to reach unreacted tin for further phase formation. In [130] an overall gold concentration of $c_{Au} = 10$ at.% was found at which the process turns from reaction limited (with linear time dependent layer growth) to diffusion limited (with square root of time dependent layer growth). After sufficient AuSn and $AuSn_4$ thicknesses have been reached, an $AuSn_2$ phase forms between the AuSn and $AuSn_4$ layers, separating these layers [132]. The diffusion of tin from inside the tin grains through the AuSn phase at the tin grain boundaries is about 3 times slower than gold diffusion in the opposite direction. Hence, the tin diffusion only plays a secondary role [134].

Diffusion of tin into the original gold layer along the gold grain boundaries is also observed. However, because this process is based on a substitution diffusion mechanism, it is several orders of magnitude slower than the diffusion of gold into tin [130]. The assumption that gold diffuses faster in tin than tin in gold is supported by the observation of Kirkendall pores at the gold/tin interface inside the gold region. Kirkendall pores are generally formed in the presence of a concentration gradient (here between gold and tin). Also, the two diffusion species need to have largely differing interdiffusion coefficients. The species with the larger diffusion coefficient leaves vacancies behind. If these cannot disappear at a sink, they may agglomerate and become microscopically visible as pores [131].

On a macroscopic scale the continuing gold/tin interdiffusion at room temperature and the low energies required for phase formation lead to the growth of a multilayer structure of Sn / $AuSn_4$ / $AuSn_2$ / AuSn / Au in gold/tin diffusion couples. In [61] a bump consisting of a two layer structure of 40 µm Au + 10 µm Sn was stored at room temperature for 30 months. During this time, the tin layer thickness was reduced by about 40% due to the consumption of tin for the formation of the AuSn, $AuSn_2$ and $AuSn_4$ interlayers. Each of these intermetallic compounds had a thickness of approximately 2 µm.

For long storage times ($t \geq 30$ months), where local effects such as tin grain boundary diffusion und bulk diffusion into the tin grains only play a minor role, the layer thicknesses can be estimated by the parabolic growth law [61]:

$$t_{AuSn} \approx t_{AuSn_2} \approx t_{AuSn_4} \approx 2\ \mu m \cdot \sqrt{\frac{t}{30\ months}}, \qquad (2.1)$$

where t_{AuSn}, t_{AuSn2}, and t_{AuSn4} denote the layer thicknesses of AuSn, $AuSn_2$, and $AuSn_4$, respectively, and t denotes the time.

After a storage time of only 1 month at room temperature the three phases (AuSn, $AuSn_2$, and $AuSn_4$) have reached a combined thickness of approximately 1 µm. The IMC formation may be halted if samples are stored at -150°C [135]. In [61] the IMC thicknesses are minimized by reflowing the bumps immediately after gold/tin deposition by dipping the wafers in a liquid heated to 285°C. This liquid dipping process has the advantage of a very homogeneous temperature across the sample with no local overheating. Also, oxidation is greatly reduced, because the eutectic gold/tin solder oxidizes much slower in air than elemental tin [61].

Bump reflow and soldering profile

The process of eutectic gold/tin formation from separated gold and tin layers was analyzed in detail in [61] and is briefly described here. The soldering profile used in this thesis is similar to that used by [61] and is shown in Figure 2.3. The bump reflow process starts with 2 evacuation and nitrogen purging cycles to reduce the oxygen concentration inside the chamber to below 5 ppm (reflow oven manufacturer's specification for the described process sequence). This is necessary to avoid oxidation of the tin surface during reflow. There are two temperature hold segments at 50°C and 70°C for 60 s each to stabilize the oven and to reach a defined state for the ensuing ramp-up to 330°C at 2.5°C/s. Upon reaching a temperature of 70°C the nitrogen atmosphere in the chamber (chamber volume: 7 ℓ) is enriched by purging with 5.5 slm of 5% formic acid in nitrogen using a formic acid/nitrogen bubbler. Formic acid is known to be a suitable reducing agent that breaks up the native SnO_2 layer, which forms on the tin surface during exposure to air between the end of the electroplating process and the start of the reflow process. If not removed, the SnO_2 layer with a melting point of 1127°C and a thickness of a few nanometers [136], [101] does not dissolve in the molten tin leading to inhomogeneous reflow results (for details see Section 3.3).

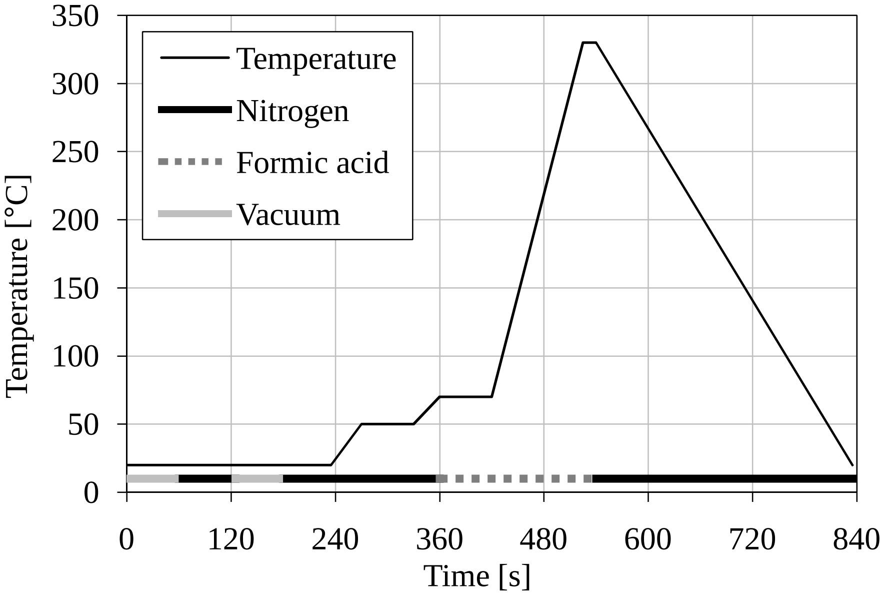

Figure 2.3: Profile for gold/tin bump reflow and soldering.

After ramp-up the heating plate is held at a temperature of 330°C for 15 s before being cooled down to room temperature at approximately -1°C/s by a nitrogen nozzle, which is pointed at the bottom side of the plate. The slow ramp-down rate allows for a better relaxation of mechanical stress in the solder.

Eutectic solder formation during reflow of gold-tin sandwich

The dynamic diffusion, IMC formation, and melting processes during the reflow of a gold-tin bumps composed of a stack of pure gold and tin are fairly complex. This is in part due to the large number of IMCs in the gold/tin system and their differing physical properties, like the temperature range of existence and the solid solubility ranges of atomic gold and tin in the IMCs. In Table 2.5 these data are given for all gold/tin IMCs, elemental gold and tin, and the ζ phase.

Table 2.5: Tin concentration and temperature range of existence for the different phases in the gold/tin system [61].

Phase	c_{Sn} [at.%]	Temperature range [°C]	
		from	to
(Au)	0 to 6.81	< 0	1064.43
β or $Au_{10}Sn$	9.1	< 0	532
ζ	10 to 18.5	120	519
ζ' or Au_5Sn	16.7	< 0	195
δ or AuSn	50 to 50.5	< 0	419
ε or $AuSn_2$	66.7	< 0	309
η or $AuSn_4$	80	< 0	252
(β-Sn)	99.8 to 100	13	231,97
(α-Sn)	99.994 to 100	< 0	13

The diffusion properties also vary greatly for the different IMCs and generally depend on the diffusion species and temperature. Diffusion prefactors D_0 and diffusion activation energies E_A for several combinations of hosts and diffusion species in the Au/Sn system are listed in Table 2.6. For better comparison the diffusion coefficients at the distinct temperatures of 20°C and 150°C were calculated with the following formula and included in Table 2.6:

$$D = D_0 \cdot e^{\frac{-E_A}{kT}} , \tag{2.2}$$

where D is the diffusion coefficient, k is the Boltzmann constant (= $8.617 \cdot 10^{-5}$ eV/K), and T the absolute temperature. When available, the temperature ranges for which the given values of D_0 and E_A are valid are listed in the two rightmost columns of Table 2.6. Nearly all values are valid for 150°C. Extrapolating them to 20°C may result in an error. However, considering that there are no phase transformations between 20°C and 150°C (except for the ζ-phase), the error is expected to be small compared to the overall variations. The diffusion coefficients are in accordance with the above description of room temperature interdiffusion of gold and tin, that is gold diffuses much faster in tin than tin in gold. Apart from the important role of the grain boundaries for the diffusion of gold in tin, there is also a strong dependence of the diffusion speed for gold on the crystal orientation of the tin host. Parallel to the basal plane of the tin crystal the diffusion coefficient is more than three orders of magnitude larger than in the perpendicular direction (which is the principal crystal axis). Between different IMCs the variations are also substantial, like for example the diffusion of gold and tin in $AuSn_2$ and $AuSn_4$. Here, diffusion in the $AuSn_2$ host is about 5 to 7 orders of magnitude faster, with the differences being larger at the lower temperature of 20°C. For the diffusion in eutectic gold/tin (that is in the Au_5Sn and AuSn phases) the data in Table 2.6 is incomplete. However, a qualitative estimate based on reflow experiments of evaporated Ni+Au+Sn samples is deducted in [137]. There, it is found that in AuSn the dominant diffusion species is gold, while in Au_5Sn tin is the primary diffusion species.

Table 2.6: Prefactors and activation energies for diffusion of gold and tin in their own host systems and intermetallic compounds [61], [138]; from these data the diffusion coefficients are calculated for 20°C and 150°C; in the two columns on the right the original temperature range, for which D_0 and E_A are valid, is given.

Host	Diffusion species	D_0	E_A	D @ 20°C	D @ 150°C	T [°C]	
		[cm²/s]	[eV]	[cm²/s]	[cm²/s]	from	to
β-Sn (II) 1)	Sn	10.7	1.09	$1.71 \cdot 10^{-18}$	$1.06 \cdot 10^{-12}$	N/A	N/A
β-Sn (⊥) 2)	Sn	7.7	1.11	$5.56 \cdot 10^{-19}$	$4.39 \cdot 10^{-13}$	N/A	N/A
β-Sn (II) 1)	Au	0.0058	0.52	$6.26 \cdot 10^{-12}$	$3.62 \cdot 10^{-9}$	125	232
β-Sn (⊥) 2)	Au	0.16	0.86	$2.37 \cdot 10^{-16}$	$8.76 \cdot 10^{-12}$	125	232
$AuSn_4$	Au, Sn	0.00509	1.00	$2.90 \cdot 10^{-20}$	$5.95 \cdot 10^{-15}$	68	212
$AuSn_2$	Au, Sn	0.00634	0.59	$4.25 \cdot 10^{-13}$	$5.78 \cdot 10^{-10}$	68	212
AuSn	Au	N/A 3)	0.59	N/A	N/A	-170	150
ζ	Au, Sn	0.0399	0.63	$5.46 \cdot 10^{-13}$ 4)	$1.21 \cdot 10^{-9}$	100	250
Au	Sn	0.0027	1.65	$9.52 \cdot 10^{-32}$	$5.54 \cdot 10^{-23}$	689	1003
Au	Au	0.09	1.81	$5.52 \cdot 10^{-33}$	$2.28 \cdot 10^{-23}$	N/A	N/A

1) Diffusion parallel to the basal plane of the crystal lattice unit cell
2) Diffusion perpendicular to the basal plane of the crystal lattice unit cell
3) N/A = Not available
4) Correctness of value uncertain

The dynamics of the eutectic solder formation from individual gold and tin layers, described in the following, is mainly based on [61] and [98], where the bump reflow of 10 µm of electroplated tin on top of 40 µm of electroplated gold was analyzed. The described process dynamics hold true for bump diameters equal to or larger than 100 µm, which is comparable to bump diameters in this thesis. The gold layer is thicker than necessary for eutectic Au/Sn bumps and hence is not completely consumed by the tin. After the bump reflow a gold base remains which is separated from the eutectic Au/Sn by an Au_5Sn layer. In this thesis, however, the electroplated gold and tin thicknesses were chosen such that they result in the formation of a eutectic 80Au20Sn layer during bump reflow (t_{Au} = 5 µm and t_{Sn} = 3.25 µm). A nickel layer was electroplated underneath the gold layer to serve as a barrier against tin diffusion along the gold tracks.

Upon heating the sample, the thickness of the IMCs increases rapidly. This is due to increased diffusion through the already existing IMCs. The diffusion coefficient for diffusion of gold and tin through the $AuSn_4$ layer for example increases by a factor of about 10^5 between 20°C and 150°C. Since $AuSn_4$ is the IMC with the smallest diffusion coefficient of the three main IMCs and its thickness is comparable to that of the other IMCs (i.e. the AuSn and $AuSn_2$ layers), it predominantly defines the IMC layer growth in the diffusion limited case, that is after sufficient IMC thicknesses have been formed [61], [98].

At 232°C, the (β-Sn) phase melts, forming a liquid spherical cap on the solid IMC interlayers. The IMC layers start to dissolve in the melted tin phase by solid-liquid diffusion. Depending on the heating rate, the $AuSn_4$ layer may not yet be fully dissolved when its melting point of 252°C is reached. Using the reflow profile in Figure 2.3, the temperature increases further and at 280°C a liquid phase of eutectic gold/tin forms underneath the IMCs (towards the gold base), resulting in the coexistence of the melted tin (with some dissolved gold) above and the melted eutectic 80Au20Sn phase below the IMCs. The IMCs continue to dissolve in the bordering molten materials. The IMC dissolution is fastest in the bump middle where the amount of liquid tin is largest (due to the spherical shape of the melted tin). As a consequence, the two liquid phases first come into contact at or near the bump middle. The mixing of these two liquids with largely differing tin-to-gold ratios results in an exothermal reaction, mainly heating up the middle section of the bump. The Sn-rich liquid, now in contact with the gold layer, strives to become 80Au20Sn leading to the dissolution of the gold base. Moreover, the gold dissolution process is sped up in the bump middle due to the increased temperature. Meanwhile, floating solid IMC remnants continue to dissolve in the melt. Because the gold base dissolution continues, the gold concentration in the melt and with it the melting point increases steadily. If the sample is held at a constant reflow temperature, the melting point will increase until it is equal to the current temperature. At that point, the melt starts to solidify in a process called isothermal solidification [61], [98].

For the gold/tin reflow in this thesis the situation is slightly different, because the size of the gold reservoir is matched to the amount of electroplated tin. A nickel diffusion barrier is used to prevent consumption of the underlying gold tracks. Once the liquid Au/Sn reaches the nickel diffusion barrier, the gold dissolution continues sideways until the defined gold volume is dissolved resulting in the eutectic gold/tin solder cap.

According to [61] and [98], the formation of a cone-shaped interface between the Au/Sn and gold layer is observed, if the bump reflow process is aborted just before the two coexisting melts come into contact in the center of the bump. However, this problem can be eliminated by storing the as-plated samples (10 µm of Sn on top of 40 µm of Au) in an oven at 200°C for 1 to 4 hours prior to reflow. Due to this high temperature storage the tin layer completely transforms into AuSn, $AuSn_2$ and $AuSn_4$, which remain mostly solid during the ensuing bump reflow. Hence, the formation of a low-melting spherical tin-rich solder cap during bump reflow is avoided. Accordingly, the exothermal mixing of two liquid phases is avoided as well, resulting in a more homogeneous bump reflow, visible by a more planar interface between Au/Sn and gold. Apart from temporary increased temperatures in the bump middle during bump reflow, the mixing of the two coexisting liquids is expected to have no major influence on the reflow process in this thesis. Therefore, the high temperature storage before bump reflow, as suggested by [61], is omitted.

In [59] the IMC formation between eutectic Au/Sn melt and an electroless nickel diffusion barrier were analyzed in detail and no excessive formation of Au/Sn/Ni intermetallics during soldering was observed. The interaction of the nickel UBM and the gold/tin solder depends on the specific circumstances. In [139], for example, gold/tin evaporated onto sputtered nickel resulted in an Au/Sn/Ni IMC with a thickness of approximately 1.7 µm after 30 s at 330°C. At lower temperatures, the phase growth is much slower. In [140], for example, the IMC layer thickness between electroplated gold/tin and nickel is 1 µm after deposition and only slightly increases to 2 µm after 400 h at 200°C.

2.2 Deformations in flexible systems

General notes on deformations in silicon wafer processing

When stressed layers are grown or deposited onto thin wafers, the wafers bow accordingly. The direction of the bow is determined by the type of stress (compressive or tensile). In isotropic samples (round isotropic substrate in conjunction with isotropic stress in the added layer) the deformed shape resembles a dish. In non-isotropic samples (for example due to the anisotropic properties of monocrystalline silicon, or due to the presence of structures without radial symmetry), the deformed shape may be better described as cylinder-like. Depending on the actual circumstances, the deformed sample can take an intermediate form between the spherical and cylindrical shape. Moreover, for large deformations, two stable states may exist [12].

Layers added during standard CMOS wafer processing are often stressed. The stress can result from a CTE difference in combination with temperature changes, which either leads to compressive stress (example: cooling after thermal oxidation of silicon) or tensile stress (example: cooling after physical vapor deposition of aluminum) in the added layer. The stress may also be structurally motivated, for example as a result of different doping levels in adjacent layers. In CMOS wafer processing a silicon nitride (Si_3N_4) passivation layer is generally added at the end of the process by plasma enhanced chemical vapor deposition (PECVD). By adjusting the plasma deposition parameters, a stress-free passivation layer can be created or, if necessary, a passivation layer with tensile stress can be fabricated to compensate compressive stress in the underlying layers. Wafer bow can also be compensated by a backside metallization. Depending on the deposition method and the type of metal or alloy, tensile or compressive layers can be formed. If bow compensation is not an issue, bowing caused by the backside metallization can be minimized by combining a compressive and a tensile layer with adjusted thicknesses [12].

For regular unthinned wafers ($t_{Si} = 625\ \mu m$), the bow from standard CMOS processes is typically in the order of a few microns to several tens of microns. Since, as a first approximation, the wafer bow b is inversely proportional to the square of the wafer thickness (b ~ $1/t_{Si}^2$), thinned wafers show substantially larger bow. Wafers with a thickness of 200 µm may, for example, show bowing in the order of a few hundred microns [12].

Relaxation during subsequent tempering processes can change the stress level (example: tempering of plasma grown low-stress silicon oxide at 1100°C more than doubles the compressive stress after cool-down). But also storage at room temperature may change stress levels and types significantly. Depending on the mode of deposition (sputtering or evaporation), the metal layers are not in thermal equilibrium after deposition. By physical processes, such as room temperature diffusion, the properties of the layers may change, which can be observed as a change in wafer bow [12].

The layer stress after deposition strongly depends on the deposition parameters. In sputtered layers, for example, a lower pressure in the deposition chamber leads to more compressive stress [21], and in electroplated layers the intrinsic stress is generally a function of the plating current density [141].

Analytical calculation of bending angles

The analytical calculation of bending states in plates and shells is fairly complex and many books have can been written on the subject (for example [142]). For very simple geometrical conditions and homogeneous material properties, analytical formulas often exist that allow deformations to be calculated with only a small error. The small error results from the approximations that were made to derive the formula. One such example is Stoney's formula (see formula (5.8) in Section (5.3.1)) that can be used to calculate the stress in a thin layer deposited onto an underlying much thicker layer from the measured radius of curvature, the layer thicknesses, and the material properties. For more complex structures (for example multi-layer structures or structures with layers of similar thickness) analytical models involve lengthy calculations.

One such model is described in [143], where deformations of thin coated plates are calculated analytically using the free energy concept. The model is expandable, allowing several deposited stressed layers to be incorporated. The calculations distinguish between deformations directly after (or even during) the deposition of each layer and deformations caused by stress after deposition of all layers (for example resulting from temperature changes). Thermal stress and intrinsic stress are both included in the model. Because the model is based on the assumption of homogeneous properties within each layer, errors occur when considering samples with anisotropic properties. However, a distribution of mechanical properties along the normal direction, as often found in deposited layers, can be taken into account by splitting up the inhomogeneous layer into two or more homogeneous layers. Information on the distribution within the layer can be obtained by observing deformations during the deposition process or by analyzing the stress distribution in the normal direction after deposition, for example using infrared spectroscopy. Because the formulas become increasingly complex with each added layer, a commercial software called *Elastica 3.0* was developed by Advanced Surface Mechanics GmbH [144] that incorporates the described analytical model. The calculations are purely analytical and are faster than equivalent finite element analyses (FEA) by a factor of about 1000. However, geometrical variations other than multi-layer composites cannot be handled by the software and require other methods like FEA.

The interaction of layer thicknesses in intrinsic bending

But even in simple geometries, like a stack of two homogeneous layers, analytical models quickly become inaccurate if assumptions made in the derivation of the model about the layer thicknesses are no longer fulfilled. The following thought experiment illustrates this problem. We assume a silicon wafer with homogeneous isotropic material properties and a standard thickness of, say, 625 µm, and coated with an unstressed gold layer with a thickness of 2 µm. Upon cooling, the gold layer contracts relative to the silicon, resulting in tensile stress in the gold layer and compressive stress in the silicon wafer. As a first approximation the stress can be calculated from the temperature decrease, the difference of the coefficients of thermal expansion, the Young's modulus, and the Poisson's ratio. As a consequence, the wafer both contracts laterally and stretches upwards with its edge, forming a dish. Both effects (lateral contraction and dish formation) are fairly small, but, nonetheless, influence the originally calculated stress in the gold layer. By continuously reducing the thickness of the silicon, both effects steadily increase. The reason for the bow increase is the reduction of the geometrical moment of inertia of the silicon (which is proportional to t_{Si}^3), while the increased lateral compression is caused by the decrease of the cross sectional area of the silicon. For very thin silicon (i.e. for silicon with less than about half of the metal layer thickness), however, the bowing is reduced again, while the silicon contraction continues to increase. The bow reduction is a consequence of the reduced force (or moment) that the thinner silicon can exert. And the increased lateral contraction of the silicon is caused by the increasing gold cross section relative to the silicon cross section. This situation becomes clearer, if the silicon thickness is reduced to close to zero. The gold layer now almost behaves as if no silicon is present at all, that is, it remains flat and contracts according to the temperature difference and its coefficient of thermal expansion. The silicon on the other hand contracts exactly at the same rate as the gold, because it is attached to the gold layer and cannot exert any substantial force on the gold layer. These effects, and especially the transition from the thick silicon to the thin silicon regime, are difficult to capture in analytical calculations. If several layers come into play or a truly three-dimensional geometry is considered, the analytical calculations become even more complex. Here, finite element analysis may be a suitable tool to analyze and predict deformations.

Maximum flexibility

The term *flexible* is ambiguous, because it can denote either a sample which can be bent quite easily (applying only a small force) or a sample which can be bent to a very small radius (like a polymer foil, which can be rolled forming a narrow tube, or even folded without breaking). With respect to flexibility in thinned silicon, however, both definitions are generally fulfilled.

If a thin chip is, for example, placed onto two support beams with distance d_{sb} and pressed down by the force F, applied to a third beam in the middle between the two support beams, the resulting bow b is calculated by

$$b = \frac{1}{48} \cdot \frac{F \cdot d_{sb}^3}{E \cdot J}, \tag{2.3}$$

where E is the Young's modulus of silicon, and J the geometrical moment of inertia ($J = 1/12 \cdot wt^3$, w = width, t = thickness) [12]. Obviously, the thin chip can be deformed more easily if its thickness t (and hence its geometrical moment of inertia J) is reduced.

The moment required for bending a homogeneous plate can be calculated by

$$M = \frac{E \cdot t^3}{12(1 - \nu^2)R}, \tag{2.4}$$

where E is the Young's modulus, t the plate thickness, ν the Poisson's ratio, and R the radius of curvature [145]. The maximum stress σ_{max} is found at the plate surface and can be calculated by

$$\sigma_{max} = \frac{3}{2} \cdot \frac{M}{wt^2}, \tag{2.5}$$

where w is the plate width [10]. From formulas (2.4) and (2.5) it is found that the maximum stress is proportional to the plate thickness, if all other parameters, including the radius of curvature, remain the same. Because the breaking risk is directly associated with the stress level in the surface layer, thinner samples are inherently less prone to breaking at a defined bending radius. Therefore, thinned silicon samples can also be considered more flexible with respect to the minimum radius of curvature that can be achieved.

2.3 Effects of crystal damage on flexibility

Silicon is a very brittle material that does not deform plastically when stressed, but rather deforms elastically and then suddenly breaks if the force becomes too large [146]. It is the purpose of this section to illustrate correlations between the presence of crystal damage at the silicon surface and the tendency of damaged chips to crack under mechanical loading or intrinsic stress. The process of crack evolution and growth is fairly complex and not included in this section. A detailed analysis is given in [147].

In standard silicon wafer fabrication the wafers are sliced from a single crystalline ingot using a mechanical multi-wire saw. The surface is heavily damaged by this process and a three-step treatment follows to obtain a good quality surface for subsequent CMOS processing. Traditionally, the surface is lapped, removing most of the damage from the wire saw process, but also adding a new layer of damage, which is, however, thinner and more homogeneous than the original damaged post-sawing layer. Then, the lapped layer is removed in wet etching, resulting essentially in a defect-free but uneven surface. Finally, CMP polishing is carried out on one side of the wafer, resulting in a very even, defect-free surface, suitable for CMOS processing. Although lapping, grinding, and wire slicing are quite different processes, the nature of the induced crystal damage is very similar in these processes. The removal process can be divided into two groups: ductile and non-ductile. Because silicon is a brittle material, the ductile mode can only be achieved if the cutting force is small enough to avoid fracture. With only a small force, the volume that can be removed is also very small, implying that small abrasive particles are required and that the resulting removal rate is fairly small [148].

When working with ultrathin silicon wafers or chips, the breaking risk is increased for two reasons: defects at the silicon back surface [19] or at the chip edge [8] reduce mechanical strength and may result in breakage when the samples are intentionally or unintentionally bent (for example during handling). The chipping length at the back of silicon chips is a good indication of crystal defects caused by mechanical dicing. In Figure 2.4 the strength for mechanically diced silicon chips is given as a function of backside chipping length. Increasing the chipping length from 10 µm to 50 µm decreases the die strength by almost 75%.

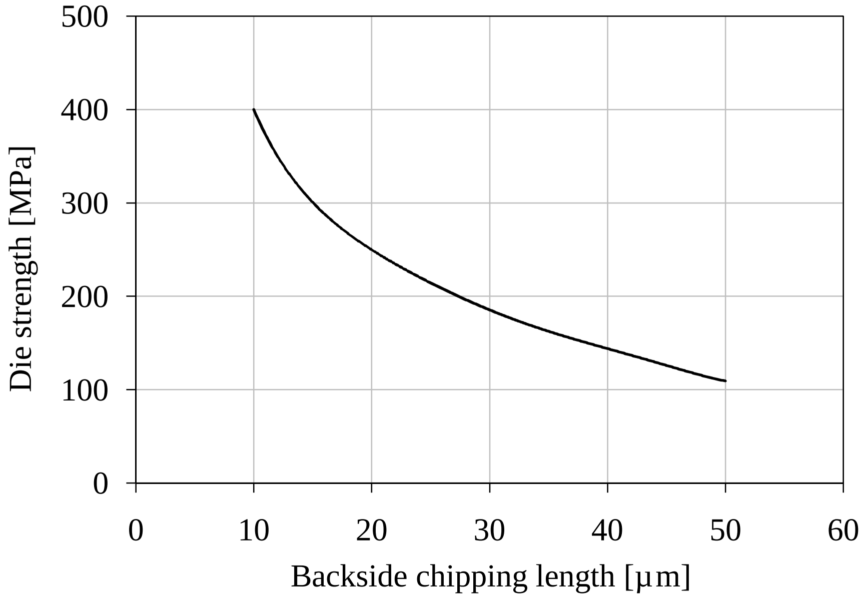

Figure 2.4: Die strength as a function of backside chipping length in chips that were first diced and then ground (modified from [8]).

Additional thinning steps are required to remove the backside and edge damage, like wet etching, dry etching, or CMP [8]. The damage depth and hence the required stock removal for effective defect elimination not only depends on the thinning technique, but also on the processing parameters. Typical defect depths are between 2 µm for fine grinding and 20 µm for rough grinding. Other sources give similar values for grinding depths, for example 6 µm in [27] and 5 µm to 15 µm in [10]. However, when analyzing sample strength as a function of removed layer thickness in wet etching, the optimum is often found to be larger, for example 30 µm in [26] or 45 µm in [149].

Hadamovsky devised a model for subsurface damage in silicon wafers after mechanical treatment, such as grinding or lapping. The model assumes four zones with different structural properties. The topmost zone at the silicon surface is fairly thin ($t \leq 2$ µm) and is called the polycrystalline zone. It is made up of small pieces of monocrystalline silicon, packed very densely. The packing density is however smaller than in undamaged monocrystalline silicon, resulting in compressive strain in the polycrystalline zone. Underneath, extending to a depth of about 15 µm, lies the crack zone with a large density of cracks. Between a depth of 15 µm to 20 µm the cracks vanish in the so-called transition zone. Finally, between 20 µm and 100 µm the stress zone is located, which experiences tensile stress resulting from the compressive stress in the disturbed layers above [7], [148].

Instead by a reduced packing density of the polycrystalline layer, the stress in the damaged layer may be explained by the oxidation of the silicon in the damaged layer, which requires a larger volume and therefore results in compressive stress [48].

The stress in the disturbed layer may vary greatly, for example between 40 MPa after fine grinding and 400 MPa after rough grinding [116]. Because this stressed layer is removed along with crystal defects in the last step of the thinning sequence, this step is often referred to as the stress relief process. Although the stress reduction increases with the amount of silicon removed in this process, the bow may first decrease and then increase again [26]. The reason for this observation is that, with continuing stress relief, the total thickness is reduced, making the wafer less rigid and resulting in bending caused by stress in the active areas on the front.

Regarding the ability of thin chips to be deformed without breaking, the surface quality at the back and the edges of the chip play a major role [10]. The defect density in the bulk silicon may also have an influence on the maximum non-destructive deformation. Therefore, float zone wafers, which typically have a lower defect density than Czochralski grown wafers, are preferably used [9].

The theoretical tensile strength for silicon varies with crystal orientation. In Table 2.7 the theoretical (defect-free) tensile strength is given for different crystal direction. Because all naturally occurring crystals have defects (vacancies, cracks, pores, and most importantly surface defects), which lead to stress concentration, the large theoretical values in Table 2.7 are never reached in reality [150].

Table 2.7: Theoretical tensile strength of defect-free silicon, assuming simultaneous breaking of all atomic bonds in the respective plane [150].

Crystal plane	σ_{theor} [GPa]
{100}	58
{110}	46
{111}	33

Although other sources calculated different values for the maximum theoretical strength of silicon (for example $\sigma_{tensile}$ = 23 GPa and σ_{shear} = 6.5 GPa in [141]), they are commonly much larger than typically achieved tensile strengths in fabricated samples, which are usually in the order of 350 MPa. In [9] a very detailed overview on maximum achieved tensile strengths for different silicon samples and on typical defect types in silicon is given.

Another aspect, influencing flexibility in thin silicon samples, is the geometric shape of the wafer edge, which changes during mechanical thinning from the initial round shape to a razor-like shape. These very sharp edges can easily break off during handling, representing starting points for cracks [12].

3 Sample preparation

This chapter first gives an overview on the design features of the samples analyzed in this thesis. The silicon chip and polyimide tape designs required to fabricate flexible ultrathin-silicon-to-polyimide systems are described in detailed, including the Daisy chain and Kelvin contact structures, their working principles, and typical layer thicknesses. Then the fabrication steps for ultrathin chips (including the Dicing-by-Thinning process) and for polyimide tapes are described. Finally, the bump reflow and soldering process development is detailed.

3.1 Sample design

Ultrathin and standard thickness silicon test chips with Au-Sn bumps of different geometries were fabricated. For flex tape fabrication, wafers with a thickness of 525 µm were used as carrier substrates for the processing of spin-on polyimide (type PI2611 by HD Microsystems). Three different silicon chip designs and one polyimide tape design were implemented. In Table 3.1 these designs are listed with information on the bulk material, sample dimensions, metal structures and the purposes of these structures. Because Design C is the most complex, it is considered in detail in the following sections.

Table 3.1: Silicon test chip and polyimide flex tape designs and their bulk materials, dimensions, structures, and usage.

Design	Bulk material	Sample dimensions	Metal structures	Used for
A	Silicon	l = 5 mm w = 5 mm t_{Si} = 2 to 45 µm	Au/Sn lines on Au-Ni-UBM with line widths and spacings of 25 µm, 50 µm, 75 µm, 100 µm, 150 µm, and 200 µm and line lengths of 4.8 mm	In-process observation of chip bow during reflow; analysis of local silicon deformations; FEM model verification
B	Silicon	l = 5 mm w = 5 mm t_{Si} = 525 µm	Cylindrical Au-Sn bumps on Au-Ni UBM	Optimization of bump reflow and soldering process; shear tests
C	Silicon	l = 10 mm w = 10 mm t_{Si} = 2 to 40 µm	30 Daisy chains and 18 Kelvin contacts with a total of 1014 Au-Sn Bumps on Au-Ni UBM; gold tracks as electrical connection between bumps and measurement pads	Reflow of thin chips and polyimide tapes; analysis of local silicon deformations; Soldering process between flex tapes and flex chips; electrical measurements (yield and quality of interconnects)
	Polyimide	l = 15 mm w = 15 mm t_{PI} = 5 µm		

Test chip and substrate designs

The complete test chip and polyimide flex tape layout of Design C is shown in Figure 3.1.

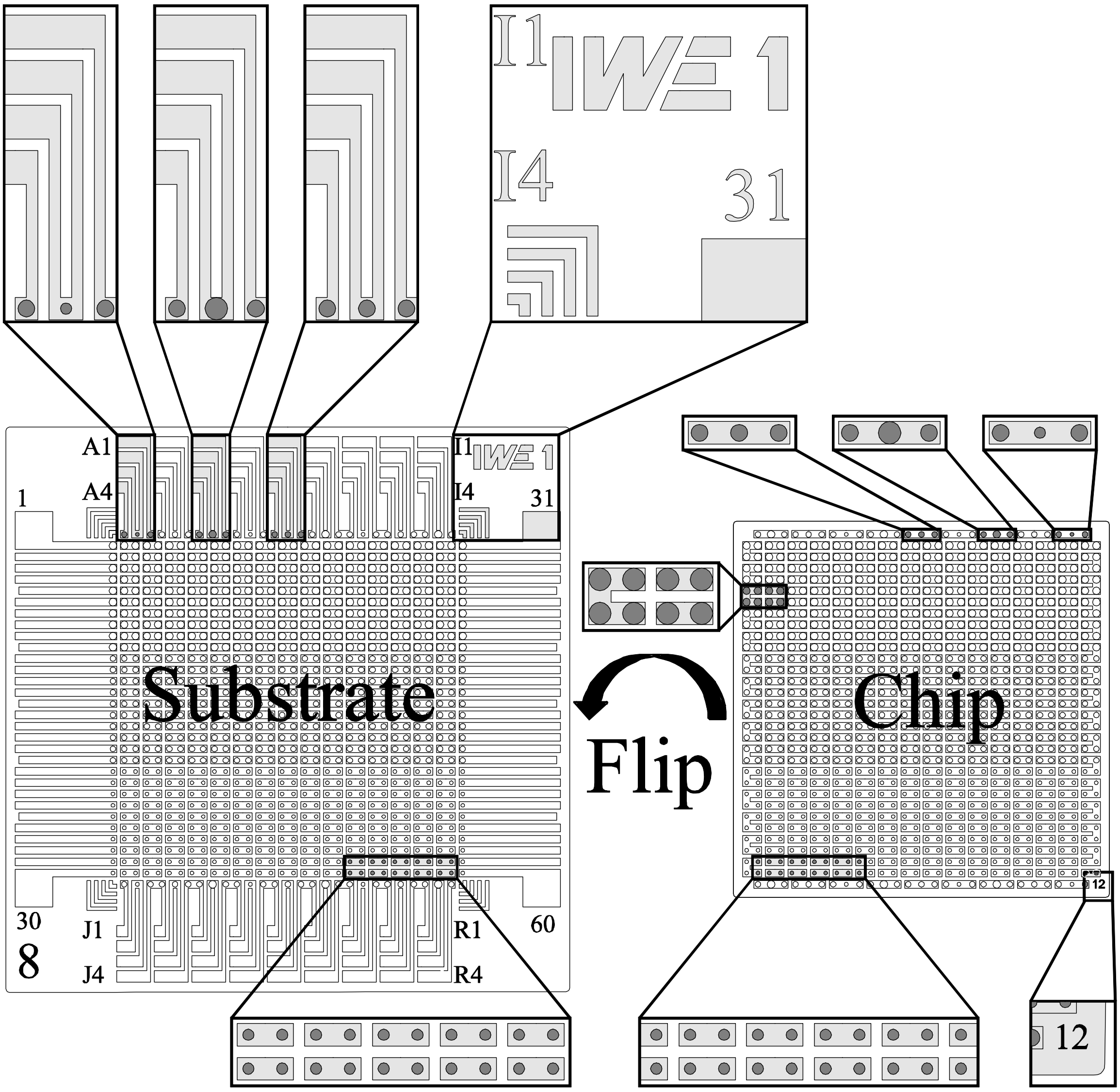

Figure 3.1: Complete substrate and chip layout of Design C; several areas are enlarged 3-fold: top middle: IWE1 logo, pad numbers (I1, I4, and 31), and optical chip alignment structures; top left: 3 Kelvin contacts, each comprising 4 contact pads and wires connected to 3 bumps; top right: chip counterparts of 3 Kelvin contacts, each consisting of one wire and 3 bumps; middle: U-shaped wire that connects two adjacent Daisy chains; bottom: parts of Daisy chains no. 29 and 30, each part comprising 5 gold tracks and 10 bumps; bottom right: chip corner with individual chip number and round chip corner from plasma etched dicing trenches.

Daisy chains

A Daisy chain is a structure that is used to measure the connectivity of a large number of interconnects in a single resistance measurement. It consists of a sequence of wires on the substrate and chip that were connected by bumps such that a long continuous current path through all bumps is formed. In this design the Daisy chains are running from left to right and have a sideways pitch of 300 µm. The polyimide substrate part of each Daisy chain comprises:

- 15 short gold tracks (length = 500 µm, width = 200 µm, spacing = 100 µm),
- 2 long electrical measurement tracks at each end (length = 1000 µm, width = 200 µm),
- 32 circular Au/Sn interconnects on Au-Ni UBMs.

The silicon chip part of each Daisy chains comprises:

- 14 short gold tracks (length = 500 µm, width = 200 µm, spacing = 100 µm),
- 2 halves of U-shaped wires to connect all 30 Daisy chains to form a single 900 bump Daisy chain that can be measured through Pads 1 and 30 of the substrate,
- 32 bumps with a pitch of 300 µm.

The individual Daisy chains are referred to by numbers ranging from 1 for the topmost chain to 30 for the bottommost chain. The Daisy chains can be electrically measured by contacting the long connection wires on the left and right side of the polyimide substrate. Every fifth of these connection wires is slightly shorter to facilitate counting. The 30 Daisy chains are split in 3 groups, each containing 10 Daisy chains with bumps of the same diameter (100 µm, 150 µm, and 200 µm, respectively). A cross section schematic of a Daisy chain is shown in Figure 3.2.

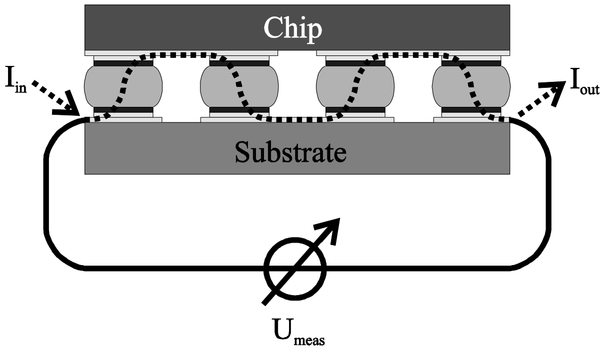

Figure 3.2: Schematic Daisy chain with 4 bumps. A current I_{in}/I_{out} (dashed line) is forced through the series of alternating gold/tin interconnects and gold tracks from left to right. Meanwhile, the voltage drop U_{meas} is measured and used as an indicator for an open circuit.

Kelvin contacts

A Kelvin contact is a structure consisting of three interconnects that is used to perform electrical 4-point-measurements on individual interconnects. The 18 Kelvin contacts are located near the top and bottom edges of the chip. Each Kelvin contact comprises 4 measuring lines and 3 interconnects. A schematic of a Kelvin contact is shown in Figure 3.3. The contact resistance of the interconnect in the middle is measured by applying a constant current flowing up through the left interconnect and back down to the substrate through the middle interconnect (dashed line in Figure 3.3). Meanwhile, the voltage drop through the middle interconnect is measured by help of the right interconnect, which serves as a connecting line to the top of the middle solder joint. Because the voltage measurement itself requires only a very small current, the voltage drop in the voltage measuring lines is negligible and, consequently, the reading on the voltage meter represents the voltage across the middle interconnect. For all Kelvin contacts the left and right interconnects have a diameter of 150 µm while the middle interconnect has various diameters of 100 µm, 150 µm and 200 µm. The interconnect pitch within a Kelvin structure is 350 µm.

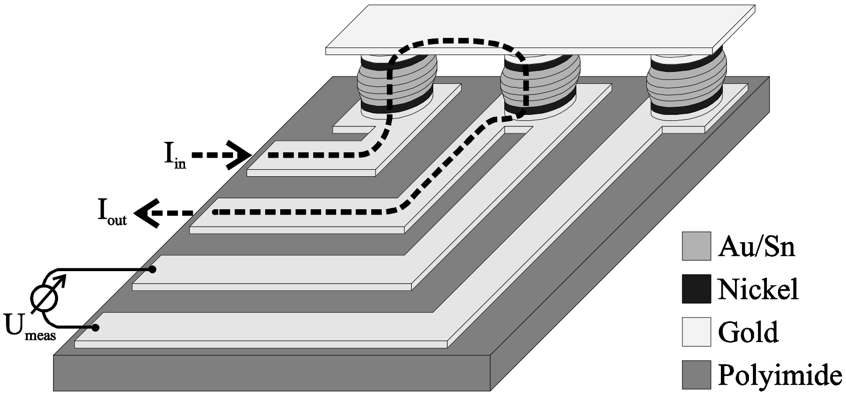

Figure 3.3: Schematic of Kelvin contact for electrical 4-point-measurements of a single interconnect. The silicon of the flip chip is omitted to offer a free view of the interconnects. The current path and the voltage measurement path cross in the middle interconnect only, allowing the middle interconnect's resistance to be determined.

Altogether the Daisy chains and Kelvin contacts of one soldered system comprise 1014 interconnects, each of which is obtained by soldering of 2 gold/tin bumps. A schematic cross section of an individual solder interconnect with approximate dimensions is shown in Figure 3.4.

.

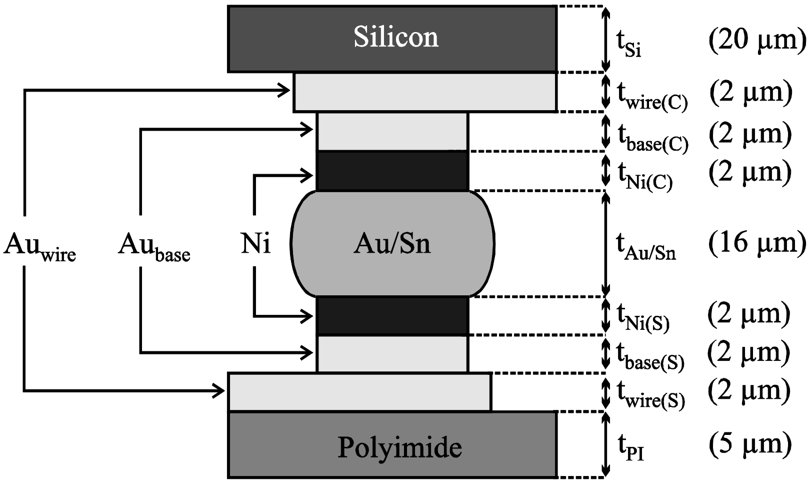

Figure 3.4: Schematic cross section of gold/tin solder interconnect between a thin silicon chip and polyimide tape; symbols and approximate thickness values for all layers are included (the indices C and S denote the chip and substrate, respectively); the current path starts at the left end of the bottom gold wire, then goes through the interconnect to the top and continues to the right end of the top gold wire.

3.2 Fabrication of ultrathin silicon chips and polyimide tapes

This section is divided into 3 parts. First, the wafer level fabrication processes for gold tracks, gold-nickel UBMs, gold-tin solder, and silicon trenches are described. Then the Dicing-by-Thinning process is described that is used to thin and singulate the chips of a single wafer. Finally, the fabrication processes for polyimide tapes used as substrates for flip chip soldering of flex chips are described.

3.2.1 Fabrication of tracks, bumps, and silicon trenches

In Figure 3.5 the fabrication steps required to realize chips with Daisy chains, Kelvin contacts, and silicon trenches are shown schematically.

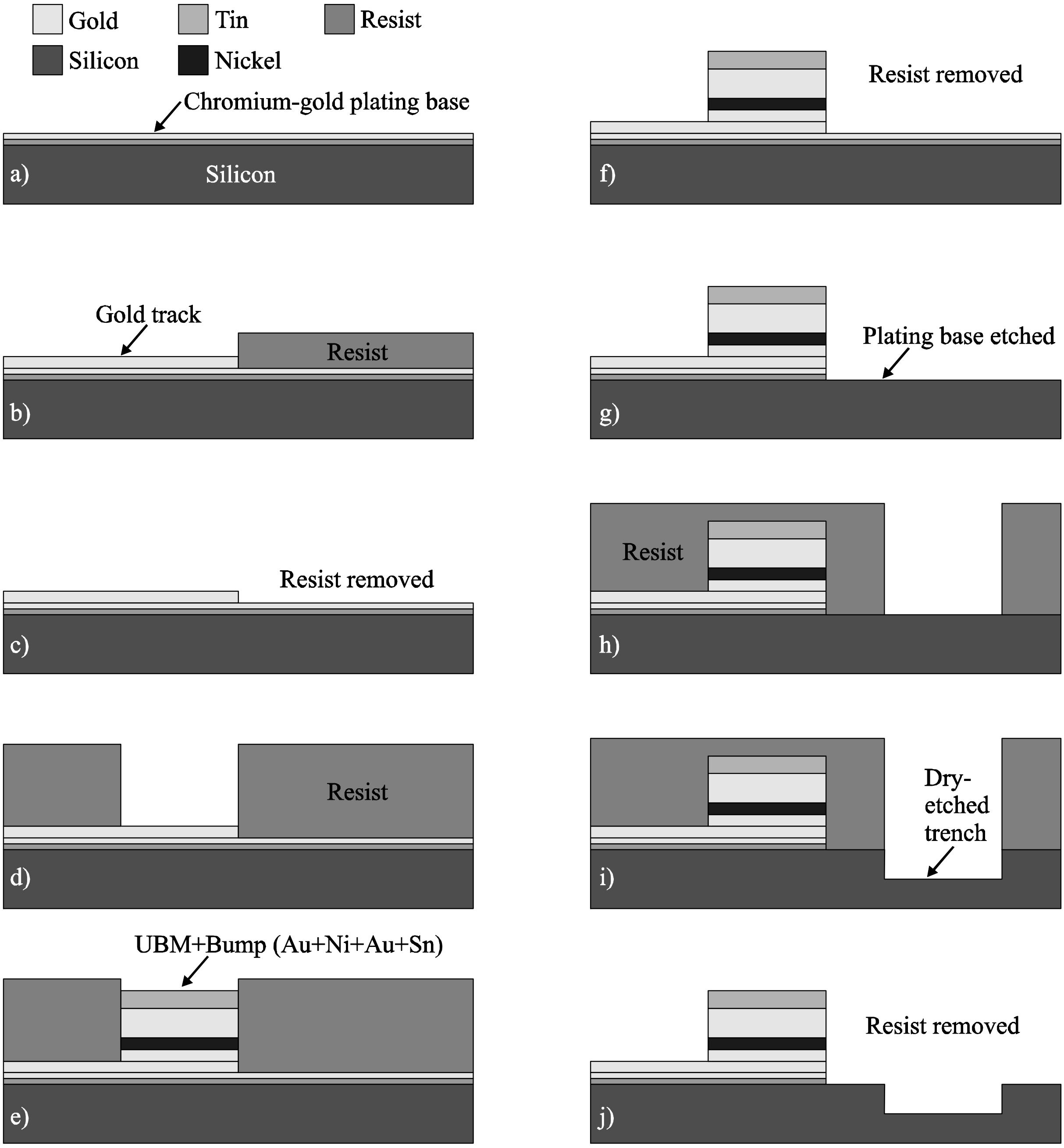

Figure 3.5: Schematic cross-sectional overview of fabrication process for wafers with gold tracks, gold-tin solder bumps (on a gold-nickel UBM), and trenches etched into the silicon. The wafers can be used to produce thin flexible chips using the Dicing-by-Thinning process described in Section 3.2.2; a) Starting wafer with a 50 nm Cr + 200 nm Au plating base; b) electroplated gold tracks defined by patterned resist; c) resist removed; d) patterned resist for UBM and solder plating; e) sequential electroplating of 2 µm Au + 2 µm Ni + 5 µm Au + 3.25 µm Sn; f) resist removed; g) wet etching of plating base; h) patterned resist for silicon trench etching; i) dry etched silicon trenches (for chip singulation); j) resist removed.

All resist exposures described here were carried out using laser-printed prototyping masks that are about 10 times cheaper than regular quartz masks. These masks allowed for fast prototyping at the expense of a slightly rugged edge definition with a typical deviation from the design lines of about 5 µm. The above description of the fabrication process is reduced to the important steps. Most deionized (DI) water rinsing, baking steps, and in-process control (IPC) measurements were omitted. The designating letters in the following detailed process description are the same as in Figure 3.5.

a) Starting material: standard 4” p-type (100) silicon wafers (t_{Si} = 525 µm, resistivity ρ = 1.5 Ωcm to 4 Ωcm, Czochralski grown). The wafers were coated with evaporated chromium (t_{Cr} = 50 nm) and gold (t_{Au} = 200 nm). The gold layer served as a plating base and the chromium layer served as an adhesion layer for the gold layer. Chromium was preferred over titanium as an adhesion layer, because it could be etched selectively against the tin layer of the solder.

b) Using a Karl Süss RC8 spin coater, the wafer was coated with 15 µm of the positive resist AZ4562. After exposure on a Karl Süss MA6 mask aligner and development in AZ400K the wafers were electroplated with 2 µm of gold using a pulse plater with an alkaline sulphite bath (type Enthone BDT 200). This step defined the gold tracks necessary for Daisy chains and Kelvin contacts.

c) The resist was removed.

d) The wafer was again coated with AZ4562 (t = 15 µm), exposed, and then developed. This layer defined the UBM and bump layout.

e) A 4-step electroplating sequence followed that defined the UBM (2 µm of gold + 2 µm of nickel) and solder (5 µm of gold + 3.25 µmof tin). These two layers were later reflowed to form a gold/tin eutectic. For the gold plating the same pulse plater as above was used. For nickel an Enthone Nickelsulfamat EL electrolyte was used and for tin plating a Dr. Schlötter SAT20 electrolyte (based on methane sulphonic acid) was used.

f) The resist was again removed.

g) The plating base was etched in two steps: the top gold layer was etched using an a iodine (J_2) + potassium iodine (KJ) etching solution. The electroplated gold was also etched by this solution. However, with a thickness ratio of 10 : 1 (electroplated gold : gold plating base) the effect is minor. The underlying chromium adhesion layer was then etched in a potassium hydroxide (KOH) + potassium ferricyanide ($K_3[Fe(CN_6)]$) etching solution. Both etching solutions showed good etching selectivity against the tin top layer of the bump.

h) The wafer was again coated with 15 µm of AZ4562, exposed, and then developed. This layer defined the trenches in the silicon surrounding each chip area. After development, the wafer was placed onto a hotplate and heated to 130°C. After a temperature hold segment of 1 minute, the wafer was cooled down by switching off the hotplate. This temperature treatment increased the resist's mechanical stability, improving its resistance against the ensuing dry etching process.

i) Dry-etching of silicon trenches using a reactive ion etcher (type RIE80 of Oxford Plasma Technology). Various dry-etching recipes were tested, resulting in different trench profiles and surface qualities. Three of these recipes are described in Table 3.2 and corresponding images are given in Figure 3.6. Each etching recipe offered different advantages. The optimal recipe would render a smooth vertical sidewall with small inhomogeneity at fast etching rates. The main quality criteria was the surface roughness and the trench profile, which were best fulfilled by Recipes a) and b). Because the trenches later came into contact with wet etching chemicals, the rough surface for Recipe a) did not pose a problem. Therefore, Recipe a) was mostly used with these additional dry etching parameters:

- pressure: 100 mTorr (≈ 13.3 Pa) and
- etching time: 20 min to 100 min (depending on intended chip thickness).

j) The resist was again removed.

Table 3.2: RIE etching parameters and resulting etching rate, trench profile, surface quality, and etching depth homogeneity (per wafer) [18].

Recipe	Power [W]	SF_6 / O_2 [sccm]	Etching rate [µm/min]	Trench profile	Surface quality	Homogeneity [%]
a)	100	7.8 / 2.5	0.4	Vertical	Rough	2.5
b)	100	4.7 / 0.0	0.5	Round	Smooth	N.A.
c)	150	7.8 / 5.0	1.0	Round	Rough	5.0

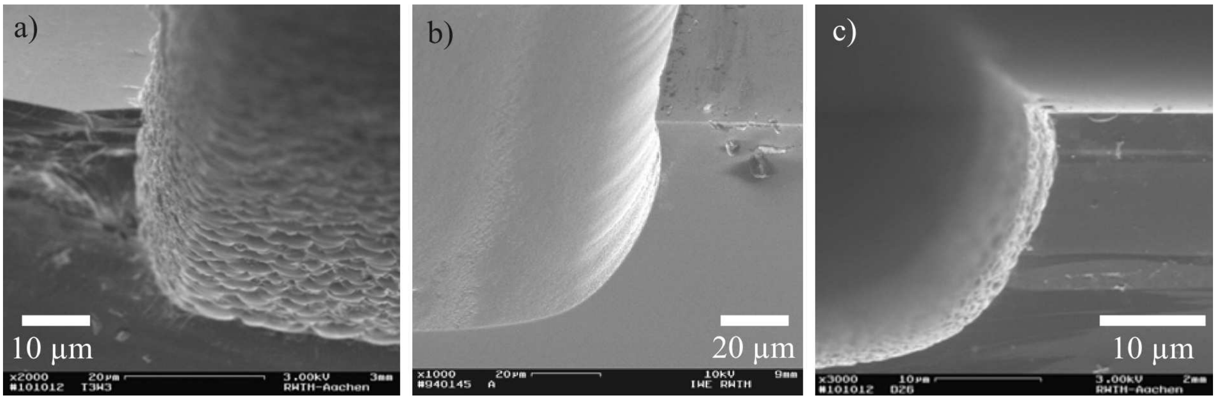

Figure 3.6: SEM images of trench profiles and surface qualities for 3 different dry etching recipes given in Table 3.2; a) vertical profile and rough surface; b) round profile and smooth surface; c) round profile and rough surface [18].

3.2.2 Dicing-by-Thinning

The Dicing-by-Thinning process was introduced by Fraunhofer IZM [30] and was used in this thesis to fabricate ultrathin silicon chips. The schematic process flow is shown in Figure 3.7.

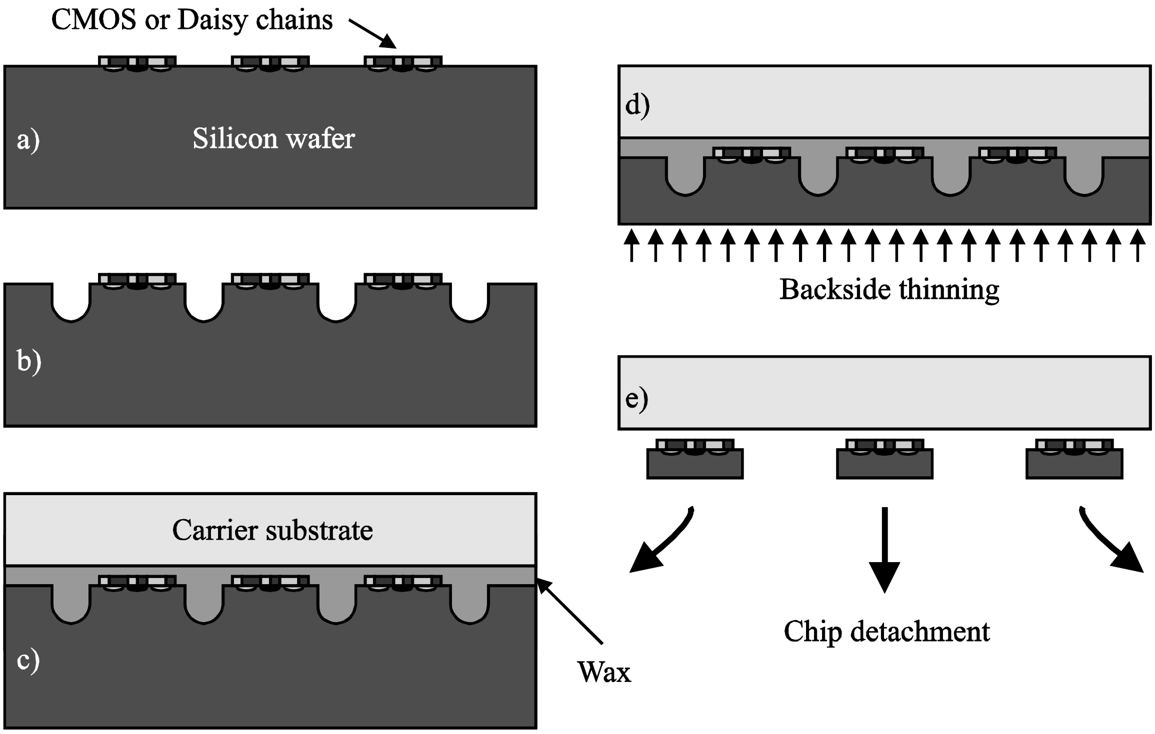

Figure 3.7: Schematic Dicing-by-Thinning process flow; a) silicon wafer with CMOS or Daisy chain areas; b) plasma-etched trenches around CMOS or Daisy chain areas; c) gluing of wafer to carrier substrate with wax (type Logitech 0CON-200, melting point ≈ 69°C); d) backside thinning by sequential lapping, chemical mechanical polishing, and wet etching; e) chip detachment by wax dissolution in limonene or heated isopropanol [18].

The advantage of this process was that the chip edges were not defined by mechanical dicing, but by plasma etching (see Figure 3.7 b)). This allowed for almost any chip outline and resulted in very little crystal damage at the chip edges. Because crystal damage was the most critical parameter, limiting the maximum bow that could be applied before breaking the chip, the dry-etched edges were preferred over mechanical dicing lines. Plasma etching of an ultrathin wafer posed several problems. The resist processing on an ultrathin wafer was only possible when using a carrier substrate for handling. This however changed the dynamics of the lithographic process, requiring a completely new process development. Also, a glue was required for attachment of the wafer to a carrier that could withstand hardbake temperatures of 130°C and did not gas out in the vacuum plasma chamber. In ultrathin wafer applications out-gassing of glue poses a major problem, because the thin wafers cannot mechanically withstand the local pressure caused by out-gassing. As a result the silicon bulges or breaks. The DbyT process circumvented all these difficulties by first performing the plasma-etching of silicon trenches on the front side of a standard silicon wafer and then thinning the wafer from the backside. Upon reaching the trenches the wafer separated into individual chips (giving this process its name, because the thinning process is the final step of the dicing process). By dissolving the wax glue, these singulated chips were detached from the carrier substrate. The wax used in this thesis (0CON-200 by Logitech) had an approximate melting temperature of 69°C and could be dissolved at room temperature in limonene (Ecoclear by Logitech) or at elevated temperatures in isopropanol. Because the thinning process was stopped when the trenches were reached, the trench depth defined the resulting chip thickness.

The maximum bow that can be applied before chip breaking occurs not only depends on the chip edge quality, but also on the surface quality of the wafer backside [66]. Therefore, the thinning process was very critical and is described in detail here. The thinning process consisted of a sequence of lapping steps on a Logitech LP50 lapping system, followed by a CMP step, which was also done on the LP50. Finally, a wet etching step was carried out using an etching solution composed of hydrofluoric acid, nitric acid, and acetic acid (also called HNA solution). A list of all thinning steps is given in Table 3.3.

Table 3.3: Process steps for thinning a silicon wafer from 525 µm to 20 µm.

Step	Process	Silicon thickness [µm]		Removal rate [µm/min]	Purpose
		before	after		
1	Lapping	525	100	5	Fast stock removal with substantial surface damage
2	Lapping	100	40	2	Medium stock removal with reduced surface damage
3	Lapping	40	30	1	Slow stock removal with minor surface damage
4	CMP	30	25	≈ 0.1	Planarization and surface damage reduction
5	HNA	25	20	≈ 2	Surface damage removal

Processing details for the three lapping steps

The LP50 lapping system mainly consisted of

- an automatic slurry feeder (set to approximately 1 drop of slurry per second),
- a rotating, grooved, cast-iron lapping plate,
- a lapping jig that held the glass carrier (to which the wafer was attached) by vacuum, and
- a monitor jig that continuously checked and corrected the rotating plate's planarity in-process.

The lapping jig was a round, passive, mechanical device that was placed onto the lapping plate. Two ball bearing wheels prevented the jig from moving around with the lapping plate, however, without preventing rotation of the jig about its own center axis. The tangential velocity of the lapping plate increased with radius, resulting in a larger tangential force on the jig near the lapping plate edge than near the lapping plate's rotation axis. This force difference lead to a momentum on the jig, resulting in the rotation of the jig about its own center axis. In summary, both the lapping plate and jig are rotating about their respective center axis, leading to a continuously changing direction of the local relative motion between wafer and lapping plate and hence to a more homogeneous material removal. Although generally increasing with speed, the removal rate may be reduced at very large speeds due to hydroplaning. This was mostly a problem in plain (i.e. non-grooved) lapping plates. The grooved plate, that was used throughout this thesis, reduced the likelihood of hydroplaning significantly, even at the maximum rotation speed of 70 RPM (revolutions per minute) used in the first and second lapping step. The removal rate also depended on the force used to press down the wafer. This force could be changed between 0 N and 40 N by a spring mechanism.

The lapping slurry used in this thesis was an aqueous solution containing 10% aluminum oxide (Al_2O_3) powder with a particle diameter of 15 µm. It allowed for sufficient removal rates of 5 µm/min that are comparable to removal rates for an aqueous solution of 9 µm SiC powder, but resulted in less crystal damage. Although finer powders generally resulted in less mechanical damage, they exhibited one problem: the distance between the silicon surface and the cast-iron lapping plate was relatively small. Therefore, thickness variations of the lapping plate or break-away silicon particles could cause permanent damage to the sample. Both 3 µm and 9 µm Al_2O_3 powders were tested, but both often lead to complete chip destructions, especially in chips with a final thickness below 10 µm.

The lapping jig was equipped with a dial indicator with a 2 µm division that could be used to check the total stock removal and removal rate in-process. A micrometer gauge was additionally used to measure the total thickness of the glass carrier + wax + wafer. By comparison with the total thickness at the start of the lapping process, the amount of removed material was calculated. In order to obtain a homogeneous silicon thickness across the wafer during lapping, several conditions had to be met:

- homogeneous wax thickness: this was best achieved by using a very thin wax layer (t_{wax} = 1 µm to 3 µm). To accomplish this, the wax was heated to about 120°C (which was about 50°C above its melting point, making the wax less viscous) and by applying a large force of approximately 500 N between the wafer and glass carrier during wax attachment (more efficiently squeezing out excess wax),
- small wedge error between the wafer and the lapping plate: this was best achieved by planarizing the lapping jig (i.e. the outside ring and the chuck holding the glass carrier) and using a planarized glass carrier, and
- good planarity of the lapping plate: the flatness monitor continuously checked the bow of the lapping plate. By automatically moving the flatness monitor and the lapping jig

slightly towards or away from the rotation axis of the lapping plate, the bow was corrected. Typically, a flatness of ± 0.5 µm is achieved.

Processing details for the CMP step

The chemical mechanical polishing was also carried out on the LP50 system. The cast-iron grooved lapping plate was replaced by a plain polyurethane plate and a colloidal solution (Syton SF1 by Logitech) of SiO_2 particles (diameter 30 nm) with an approximate pH value of 10.5 was used instead of the Al_2O_3 slurry. The flatness monitor was not required for the CMP process and the position of the lapping jig was changed continuously by a sweeping action of a few centimeters. The force pressing down the wafer was set to 60 N (by attaching an additional weight to the lapping jig). Because the polyurethane plate was not grooved, hydroplaning could occur, requiring the rotation speed to be reduced. Typical maximum rotation speeds before hydroplaning occurred were between 30 RPM and 40 RPM. Instead of the hard polyurethane plate an expanded rubber plate (type Chemcloth Polishing Cloth by Logitech) could be used, which generally did not show hydroplaning effects and resulted in a near perfect optical grade surface. However, this came at the expense of good planarity. The soft foam rubber squeezed into the opened silicon trenches, increasing the pressure and removal rate near the chip edge. This effect resulted in an increased stock removal of up to 2 microns on the chip edge.

In the first few minutes of the CMP process the removal rate was fairly large (about 0.5 µm/min), because the crystal damage and rough surface remaining after lapping facilitated the removal of the silicon material. The removal rate decreased and reached a stationary value of 2.5 µm/h after about 40 minutes. Because of this exceedingly slow removal rate after the first 5 microns were removed, the CMP was not suitable for further thinning. The long-term removal rate could be increased substantially by polishing a small number of chips instead of a whole wafer. This was, however, contradictory to the DbyT concept which is a wafer based process. Therefore, the main purpose of the CMP step was not to remove substantial amounts of silicon material, but rather to reduce the surface roughness left by the preceding lapping process, so that the ensuing wet etching was carried out on a more planar starting surface. After the CMP process the wafer was cleaned of the CMP chemicals by using large amounts of water on the polishing plate and keeping the jig on the rotating plate for 2 more minutes.

Processing details for the wet etching step

The wet etching was performed using a combination of hydrofluoric acid, nitric acid, and acetic acid (called HNA etching solution) for approximately 3 minutes to remove most of the subsurface damage and to round off the possibly sharp chip edges. Wet etching was chosen for stress relief, because it generally leads to less defects than dry etching [19]. Because the HNA etching solution also etched glass, the glass carrier's backside was protected by a commercial blue tape. In Figure 3.8 the backside of a silicon wafer with a partly opened trench intersection and its evolution during wet etching is shown. The increasing roughness was a good indication for subsurface defect removal [18].

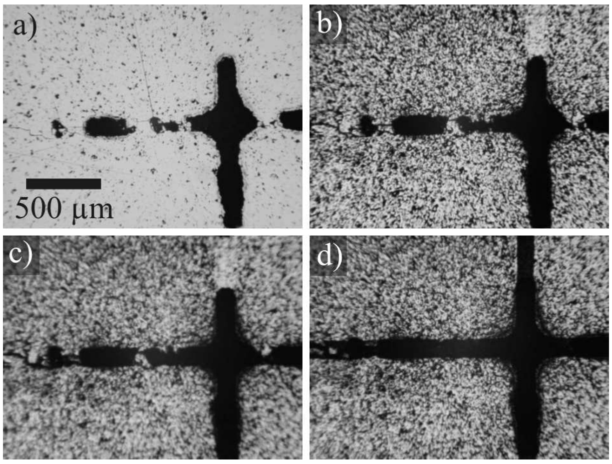

Figure 3.8: HNA wet etching of CMP polished silicon surface; a) silicon surface directly after CMP, trenches partly opened; b) after 1 minute of HNA etching, trenches further opened, surface roughening at crystal defects; c) after 2 minutes of HNA etching; d) after 3 minutes of HNA etching, trenches completely opened [18].

The HNA etching process can be described as three consecutive reactions. First, the nitric acid oxidized the silicon surface. Then, the silicon dioxide reacted with the hydrofluoric acid to form a salt, which then dissolved in the acetic acid. The etching rate largely depended on the availability of all three etching components at the silicon surface and the removal of dissolved reaction products. Agitation of the etching solution by use of a rotational shaker resulted in improved etching rate homogeneity. Also, the heat generated by the exothermal etching reaction was better dissipated. Otherwise, a local temperature increase sped up the local etching rate, further increasing the temperature, resulting in a very inhomogeneous etching rate across the wafer. Also, good heat dissipation minimized the risk of local wax melting, which would have resulted in the loss of chips.

HNA etching had the big advantage that the solution also came into contact with the sidewalls of the dry-etched silicon trenches, reducing crystal damage at the chip edge. Although the HNA etching reduced the planarity of the silicon surface significantly, the chips became more flexible, because crystal defects were efficiently removed. After the HNA etching step a final CMP might have been added to replanarize the silicon surface. However, this came at the risk of reintroducing crystal damage from breakaway silicon particles. Therefore, the HNA etching step was preferably the last step of the thinning process.

After HNA etching, the chips had to be removed from the glass carrier. The best method was to cover the bottom of a container with a piece of cloth, add isopropanol and then place the glass carrier into the container and heat the isopropanol to about 80°C. The wax melted and slowly dissolved in the isopropanol. The loose chips were washed down from the glass carrier by carefully using an isopropanol jet. By retrieving the piece of cloth with the chips from the container and drying in air, the risk of damage to the chips was minimized. The chips now had to be handled with great care to avoid new crystal damage on the edge or backside.

Good results were achieved by handling with a vacuum nozzle end effector with only a slight vacuum of about 50 mbar below ambient.

3.2.3 Fabrication of polyimide tapes

The fabrication process for polyimide substrates described here was very similar to the fabrication process of thin chips in Section 3.2.1. The major differences were:

- the starting wafers had a sacrificial layer of Al + Ti instead of a Cr + Au plating base,
- spin-on polyimide was added and patterned before electroplating of tracks and bumps,
- no silicon trenches were etched. Instead the polyimide tapes were detached from the carrier wafer by etching in 2% hydrochloric acid. The bumps were protected by a resist.

The fabrication process is schematically shown in Figure 3.9.

3.3 Bump reflow and soldering process development

Most of the reflow and soldering trials were carried out in a closed chamber rapid thermal processing (RTP) oven with vacuum and formic acid capabilities (type SRO-704 by ATV Technologie GmbH). Using the reflow and soldering profile in Figure 2.3, the manufacturer guaranteed a maximum oxygen concentration of 5 ppm. The oven was equipped with 8 IR quartz lamps with a total power of 6 kW for heating the bottom side of a thin aluminum plate onto which the chips were placed. A nitrogen nozzle underneath the heating plate was used for cooling. The oven could follow the set profile very closely, except for the last phase of the cool-down (below 100°C), where the temperature decreased exponentially instead of linearly and had a maximum deviation of 10°C above the set value. The ultra thin silicon chips often had a weight below 10 mg and were blown away during chamber evacuation or nitrogen cooling. Therefore, the samples were fixated by a needle tip during reflow using a force of 20 mN. For soldering, a flat weight piece was placed onto the silicon chip, also exerting a force of 20 mN. To compensate for heat absorbed by the weight piece, the time at the peak soldering temperature was increased to 5 minutes. In order to observe the chip bow in-process, a flip chip bonder (type Fineplacer Lambda by Finetech) equipped with a heating plate and forming gas module was used. However, the forming gas unit had to be removed to allow observation of the samples during reflow.

Tin is known to oxidize in air at room temperature forming an SnO/SnO_2 layer of a few nanometer thickness [136]. Upon heating the tin in air, the oxidation accelerates substantially [101]. By using 5% formic acid in nitrogen, the oxide thickness could be reduced significantly. In Figure 3.10 two samples are shown that were reflowed in a formic acid atmosphere and in ambient air, respectively. Although reflowing gold-tin contacts in air resulted in poor electric performance due to tin oxidation, it was assumed that the oxide layer has only a minor influence on deformations during reflow of thin chips with gold-tin lines.

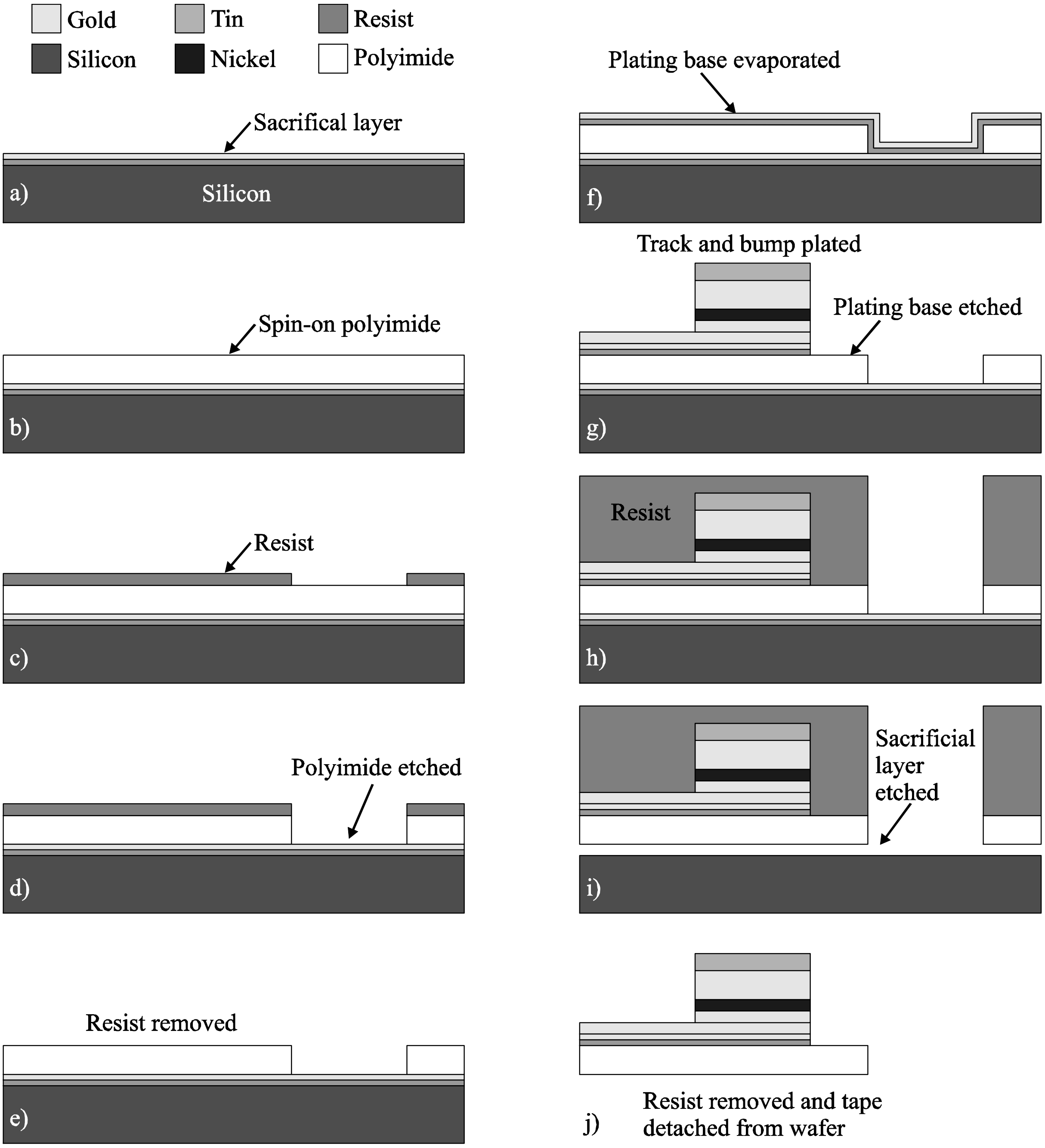

Figure 3.9: Schematic cross-sectional overview of fabrication process for polyimide tapes used as substrates for flip chip soldering of flex chips; a) starting wafer with 1000 nm Al + 30 nm Ti sacrificial layer; b) 5 µm of spin-on polyimide added; c) patterned resist for polyimide patterning; d) polyimide patterned; e) resist removed; f) Cr + Au plating base evaporated; g) electroplating of gold tracks, Au-Ni UBM, and Au-Sn solder and plating base etched (see Figure 3.5 b) to g) for more details on the process sequence); h) protective resist patterned; i) sacrificial layer etched; j) resist removed and tape detached from wafer.

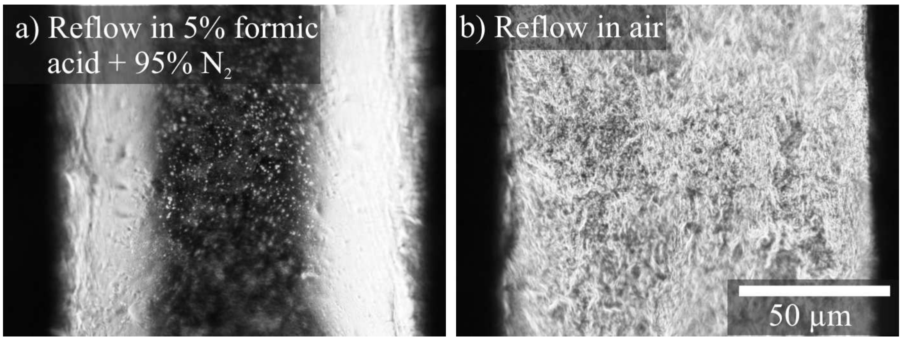

Figure 3.10: Top view of reflowed Au-Sn lines with a width of 100 µm; a) reflow in reducing atmosphere (5% formic acid); b) reflow in ambient air with strong indication of surface oxidation.

When directly soldering as-plated samples, a large number of voids were present in the gold/tin eutectic. By reflowing the bumps before the soldering process (see Figure 3.11), the voids were substantially reduced. Using a reflow temperature of 300°C, almost void-free interconnects could be fabricated during soldering. In Figure 3.12 this effect is demonstrated with 4 interconnects that were reflowed at temperatures of 270°C, 280°C, 290°C, and 300°C, respectively, prior to soldering. Possible explanations for the observed voids in samples that were directly soldered were out-gassing of organic remnants from the plating bath electrolyte or air (or nitrogen) trapped between the rough as-plated tin surfaces. Although a bump reflow temperature of 300°C was sufficient for nearly void-free solder joints, the bump reflow temperature was set to 330°C for convenience in nearly all bump reflows in this thesis.

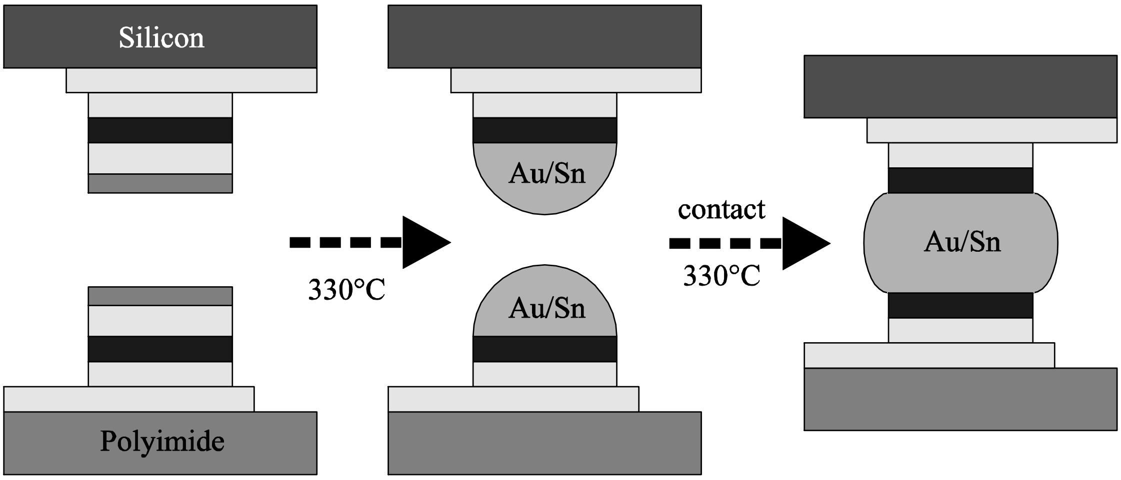

Figure 3.11: Solder process modification to reduce voids after soldering. The as-plated bumps were reflowed prior to soldering.

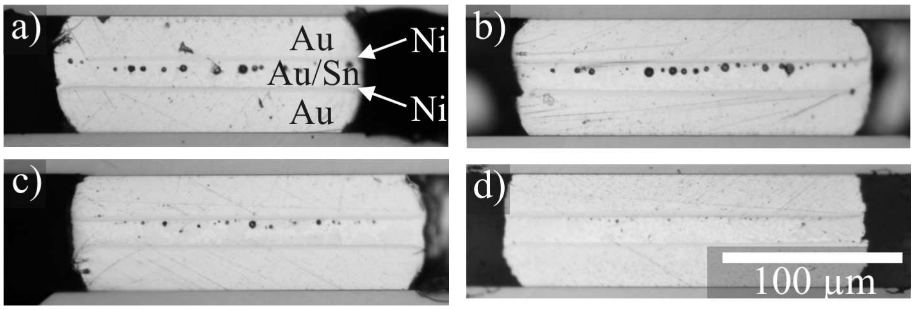

Figure 3.12: Cross sections of soldered Au/Sn interconnects between two 525 µm thick silicon chips. The solder was separated from the gold bases by two nickel diffusion barriers. Before soldering at 330°C the bumps were reflowed at different temperatures; a) bump reflow at 270°C; b) bump reflow at 280°C; c) bump reflow at 290°C, voids slightly reduced; d) bump reflow at 300°C, almost void-free interconnect.

An explanation for the observed reduction of voids is given in Figure 3.13, where bumps are shown that were reflowed at different temperatures and for different time spans at the peak temperature. In the sample reflowed at 280°C for 15 s a rugged topography can be seen. Obviously, the tin had started reacting with the underlying gold, but without forming a spherical cap. When two such surfaces were placed onto each other and soldered, gas would likely be trapped, resulting in voids. Increasing the time at 280°C to 60 s improved the eutectic formation, but still did not yield a spherical solder cap. At 300°C, a nice round solder cap was formed. Further increasing the reflow temperature to 330°C lead to similar round solder caps. Round caps like these only touched at a single point during soldering. This way, gas trapping was avoided. Also, the high reflow temperature possibly allowed remnants from the plating bath electrolytes to gas out before the soldering step.

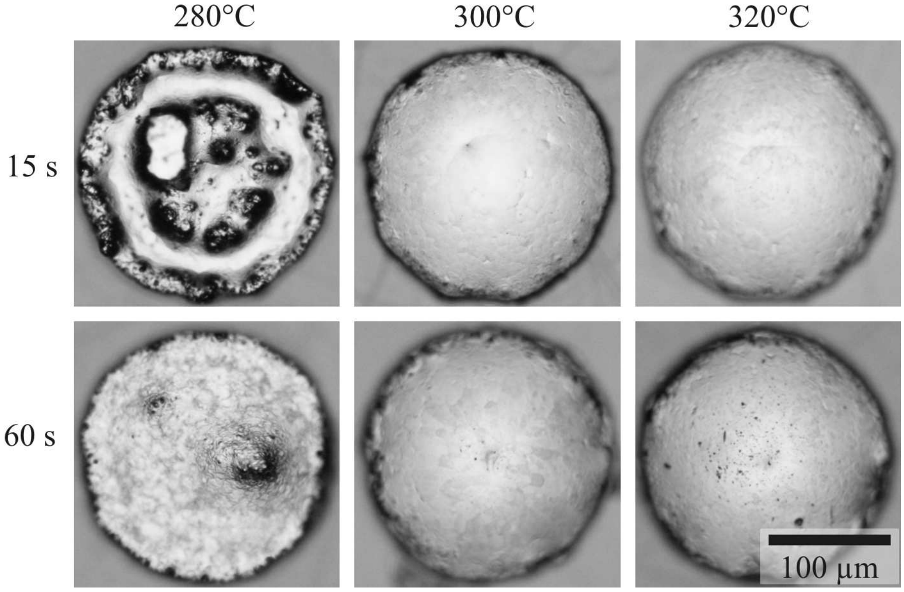

Figure 3.13: Top view of gold-tin bumps on a gold-nickel UBM reflowed at different temperatures for different time spans. At 300°C and above, round solder caps are obtained.

Based on these results, a reflow of the as-plated samples was deemed necessary to obtain void-free interconnects. However, the ultra thin silicon chips and polyimide substrates were bent after reflow. To align the samples for soldering, they were manually flattened and fixated to microscope slides using 2% adipic acid in isopropanol. Upon evaporation of the isopropanol at room temperature, the adipic acid formed solid crystals that could keep the chips and tapes flat on the microscope slides. Adipic acid also served as a reducing agent, possibly eliminating tin oxides [151]. By using microscope slides as carriers for the polyimide tape, it was possible to see through the carrier and tape at the silicon chip and manually align the chip and polyimide tape. In Figure 3.14 a schematic cross section of the Daisy chains fabricated this way and a corresponding top view image of an actual Daisy chain are shown. A cross section corresponding to A-A' in Figure 3.14 is shown in Figure 3.15. Reflowing the bumps to form round solder caps prior to soldering resulted in almost void-free interconnects with functional electrical contacts between ultra thin silicon chips and polyimide tapes.

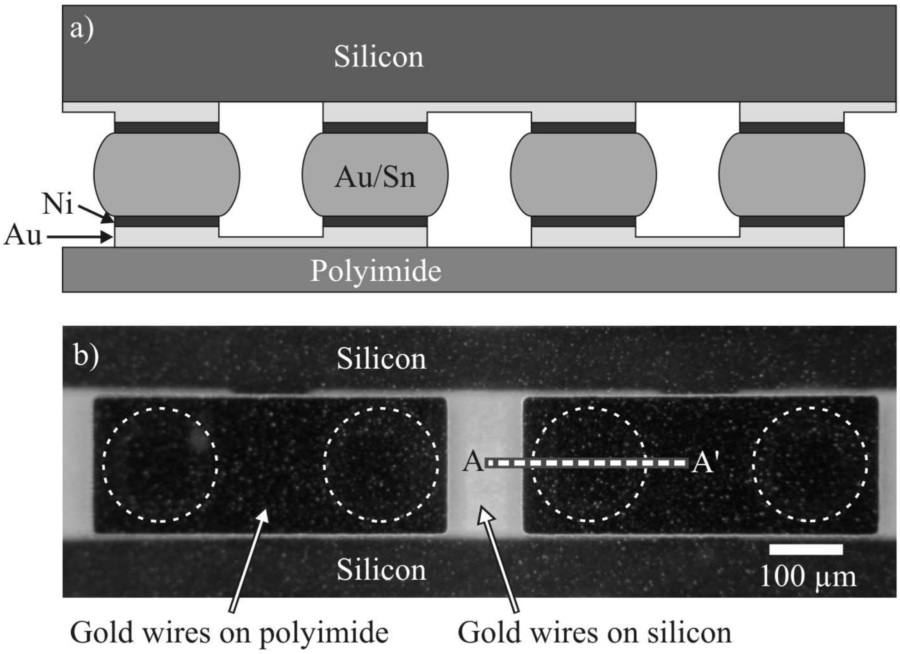

Figure 3.14: a) Schematic side view of Daisy chain between a polyimide tape and thin silicon chip; b) corresponding top view micrograph of such a Daisy chain through the transparent polyimide. The Daisy chain is oriented left-to-right. The chromium backside of two gold wires attached to the polyimide tape are visible as black rectangles. In between these rectangles the bright surface of a gold wire on the silicon chip is visible. The positions of the 4 interconnects sandwiched between the wires are indicated by dashed circular lines. A micrograph corresponding to the cross section A-A' is shown in Figure 3.15.

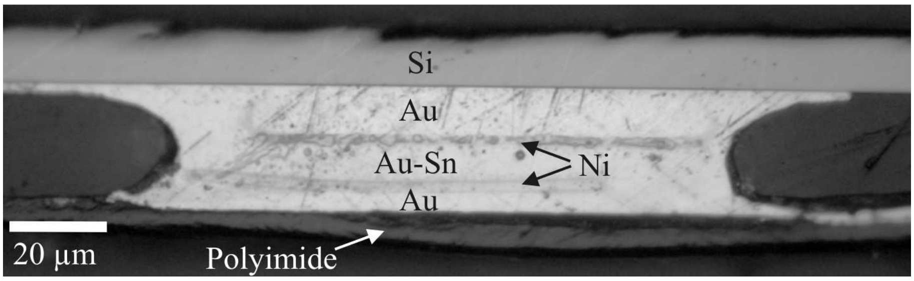

Figure 3.15: Cross section of a gold/tin interconnect with two nickel diffusion barriers and gold bases and wires. The interconnect has a diameter of 150 µm and the silicon chip has a thickness of approximately 20 µm.

In order to obtain in-process data on the bending of ultra thin chips during reflow, samples were reflowed in air on a Fineplacer flip chip bonder without using a fixation. The heating mechanism on an open hotplate and the missing chip fixation resulted in reduced thermal coupling between the heating plate and the sample. Therefore, the time at peak temperature was increased to 40 s. In Figure 3.16 the bending angle is plotted as a function of dwell time at 330°C. The major deformations occurred in the first 20 s at 330°C, suggesting that the 40 s time span was sufficient in reflow experiments in ambient air without chip fixation using the flip chip bonder. In another experiment a silicon chip was held at 330°C for a total of 150 s. Here, apart from random fluctuations, no significant change of the bending angle was observed after the first 40 s at 330°C.

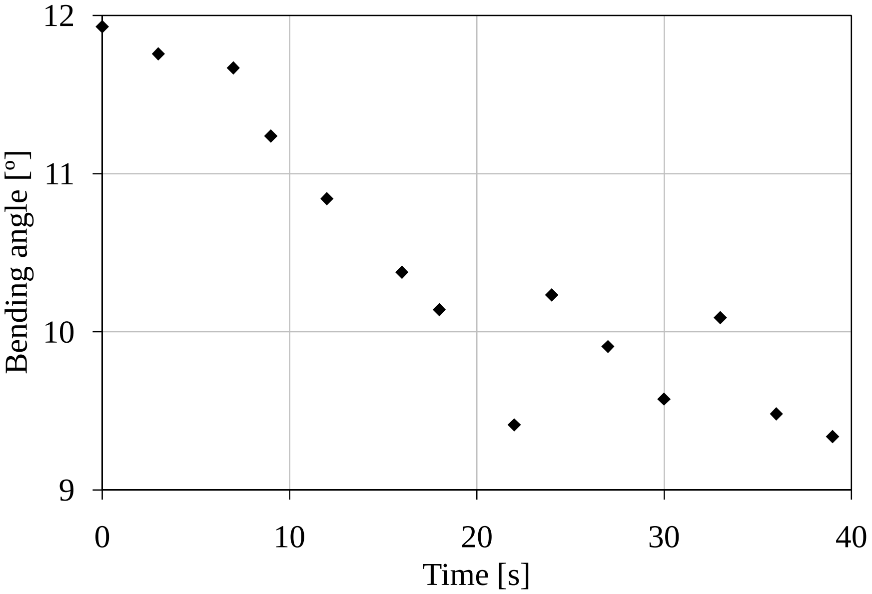

Figure 3.16: Optical in-process bow measurements for a chip with gold/tin lines as a function of hold-time at 330°C. The bending angle is strongly reduced within the first 20 s at 330°C and then slightly between 20 s and 40 s.

4 Modeling

First, the geometry of the system and the mechanical properties of the used materials were input into the FEM simulation program. In a process called *meshing*, each geometric unit was subdivided into a sufficient number of elements such that each element could be considered to have homogeneous properties. Due to the large number of bumps and wires on the chip and substrate and their complex individual shapes with many round surfaces and edges and because of the very small thickness of the chip and substrate compared to their widths and depths, the number of required simulation elements was very large. In order to achieve practical simulation times, the model was reduced to a small section of the actual system. For example, for the reflow simulation of bumps on a thin silicon chip, this section represented a small unit of the spatially periodic chip layout, that could be used as a building block for almost the entire chip. An external influence, called a *load*, was applied and the reaction of the system was calculated by the simulation software (through the solution of a large set of equations by the so-called *solver*). In the case of solder process simulations, the load was generally represented by a temperature change, leading to different thermal expansions (or contractions) for geometric units composed of different materials. If attached to each other, stress built up between these connected units and the sample deformed. This situation was similar to a bimetal that bends under thermal load. Because of the many model simplifications, the simulation results that are extracted from the model needed to be further analyzed and verified by comparison with measurements on actual samples. Because different models were required for the different steps of the soldering process, the individual simulation results were superimposed to calculate the resulting deformation of the system after soldering. Therefore, linearity of the simulation model was required.

4.1 Material properties for finite element simulations

In general, FEM simulations require all used materials to be defined via their material properties. The required set of parameters depends on the simulation type and may for example include parameters describing electrical performance, mass density, or magnetic or mechanical properties. In this thesis, the geometric deformations caused by thermal stress from the soldering process were analyzed. Therefore, the mechanical parameters of all present materials were required for the simulations. These were:

- Young’s modulus E and Poisson’s ratio ν (required for all predominantly isotropic materials: gold, nickel, eutectic gold/tin solder, and polyimide),
- elastic constants c_{11}, c_{12}, and c_{44} (required for silicon), and
- coefficient of thermal expansion α (required for all materials).

For all isotropic materials the FEM software ANSYS calculated the shear modulus G from the given Young’s modulus E and Poisson’s ratio ν using

$$G = \frac{E}{2 \cdot (1 + \nu)} . \tag{4.1}$$

Therefore, the shear modulus did not have to be entered separately.

Very thin layers, like the thermal silicon dioxide (t_{SiO2} = 300 nm) and the chromium-gold plating bases ($t_{Cr\text{-}Au}$ = 50 nm + 200 nm) were not expected to have a relevant impact on deformations and were therefore omitted in the model. Although the material properties of the individual lamellae of the eutectic gold/tin were well known (see [152] for details), the solder was modeled using homogeneous solder material properties. The complete set of material properties is shown in Table 4.1. The references for each parameter are discussed in the following.

Table 4.1: Material properties for mechanical FEM simulations of deformations caused during the reflow or soldering process.

Material	α [ppm/°C]	Young's modulus [GPa]	Poisson's ratio
Gold	14.5	78	0.42
Nickel	14.3	195	0.312
80Au20Sn	16.2	72	0.405
PI2611	3.0	8.5	0.35
Silicon (anisotropic)	3.12	Elastic constants for (110) silicon: $c'_{11}=c'_{22}$= 194.48 GPa; c'_{33}= 165.78 GPa $c'_{12}= c'_{21}$= 35.24 GPa $c'_{13}=c'_{31}=c'_{23}=c'_{32}$=63.94 GPa $c'_{44}=c'_{55}$= 79.62 GPa; c'_{66}=50.92 GPa	

Material properties of gold

Young's modulus for gold

In monocrystalline gold the Young's modulus shows a strong dependence on crystal orientation. For example the Young's moduli in the <100>, <110>, and <111> directions are: $E_{Au,<100>}$ = 43 GPa, $E_{Au,<110>}$ = 81 GPa, and $E_{Au,<111>}$ = 117 GPa [153]. For the Young's modulus of gold a value of 78 GPa is generally given in the literature [154], [155], [156] with some small deviation upwards or downwards depending on the manufacturing process (cast, wrought, annealed, or not annealed, etc…). However, in thin film gold the values show a much larger spread ranging from 39 GPa to 116 GPa [157], [158], [159], [160], [161], [162]. The effects influencing the actual Young’s modulus in thin films are the crystallographic texture, possible strain gradients, the mean grain size [160], and contamination or alloying with trace elements [163]. Generally, the Young’s modulus for gold is larger for smaller mean grain sizes due to denser packing and confinement effects such as strain gradients [164].

For electroplated gold the available mechanical data is very scarce. Electroplated gold may be present in varying grades of crystallinity with different predominant crystal orientations resulting in different Young’s moduli. In [157] electroplated gold layers showed a Young’s modulus of only 35 GPa to 42 GPa for plating current densities of 2 mA/cm² and 4 mA/cm² and layer thicknesses of 3.4 µm and 4.0 µm, respectively, while in [165] much larger values were found depending on the layer thickness (150 GPa for t_{Au} = 0.3 µm, 94 GPa for t_{Au} = 1.19 µm, and 79 GPa for t_{Au} = 5.12 µm). Because of the large spread of literature values for electroplated gold, the Young's modulus for electroplated gold should preferably be determined experimentally for the specific plating conditions. Due to the complexity of such measurements, the standard literature value of 78 GPa for the Young’s modulus of gold was used.

Poisson’s ratio for gold

In the literature a value for the Poisson’s ratio of gold of 0.42 is usually given and was adopted in this thesis [154], [155].

Coefficient of thermal expansion for gold

Generally, the coefficient of thermal expansion depends on the temperature range that is being considered. Although the temperature used during soldering or bump reflow in this thesis ranged from 20°C to 330°C, the maximum relevant temperature for FEM simulations is 280°C. This is because most of the deformation that was caused during the heating step from 280°C to 330°C, would vanish when the sample was cooled to 280°C.

A CTE value for gold at room temperature of 14.2 ppm/°C is generally given in the literature (for example in [155]). However, the CTE of gold rises with temperature. In [155] a formula for the temperature dependence of the CTE of gold valid for a temperature range of 0°C to 950°C is given. According to this formula, the CTE of gold rises from 14.20 ppm/°C at room temperature (RT) to 14.26 ppm/°C at 280°C, resulting in an average CTE of 14.23 ppm/°C. Because of the large temperature range for which the formula is deemed valid, it may not offer the best values for the temperature range considered in this thesis. In [166] actual length measurements of an expanding gold sample at various temperatures are given. For 18°C an elongation versus 0°C reference length of $\varepsilon_1 = \Delta L/L = 0.248 \cdot 10^{-3}$ is given, while for 279.6°C a value of $\varepsilon_2 = \Delta L/L = 4.063 \cdot 10^{-3}$ is given. From these values a CTE for gold between T_1 = 18°C and T_2 = 279.6°C can be calculated:

$$\alpha_{Au,\,18\ldots279.6°C} = \frac{\epsilon_2 - \epsilon_1}{T_2 - T_1} = 14.58\,\frac{ppm}{°C}\,. \tag{4.2}$$

Because this calculation is based on elongation values for gold at temperatures very close to the actual relevant temperatures in the FEM simulations, the resulting value of α_{Au} = 14.58 ppm/°C was considered most suitable. However, as a compromise this value was slightly reduced towards the value calculated from the formula in [155] and a final value of α_{Au} = 14.5 ppm/°C was used to build the FEM material model.

Material properties of nickel

Young's modulus for nickel

The values for the Young's modulus of nickel found in the literature vary between 158 GPa and 247 GPa [167]. The elastic modulus of electroplated nickel layers produced using the same electroplating module as in this thesis were previously analyzed in [164]. Different measurement techniques were used resulting in slightly differing values, as shown in Table 4.2. In this thesis, the three results were averaged and the resulting value of E_{Ni} = 195 GPa was used.

Table 4.2: Young's modulus for electroplated nickel layers determined using different measurement techniques [164].

Measurement technique	Young's modulus [GPa]
Stress-strain diagram	192
Laser-acoustic measurement on vibrating beams	207
Nanoindenter	185

Poisson's ratio for nickel

The Poisson's ratio for nickel is generally given as 0.31 [168]. In [167], however, a more precise value for soft unmagnetized nickel of 0.312 is given.

Coefficient of thermal expansion for nickel

In [166] elongations of a nickel sample at 20.1°C and 279.2°C were determined. Using formula (4.2), a CTE for nickel of α_{Ni}' = 14.33 ppm/°C was calculated from these values. In [169] the CTE of nickel layers, electroplated using a nickel sulfamate bath, was determined. Depending on the solution composition and temperature range slightly different values were found, as shown in Table 4.3.

Table 4.3: Coefficient of thermal expansion for electroplated nickel layers from two different nickel sulfamate plating baths [169]. Values are given for temperature ranges relevant to gold/tin reflow and soldering in this thesis.

Temperature range [°C]	CTE [ppm/°C]	
	Plating bath A	Plating bath B
25 to 271	14.3	14.2
25 to 295	14.4	14.3

Taking the average of all 4 given values in Table 4.3 yields α_{Ni} = 14.3 ppm/°C. Because this value resulted from measurements on nickel samples fabricated with similar nickel plating baths as in this thesis, α_{Ni} = 14.3 ppm/°C was assumed to be the best value for FEM material models in this thesis.

Material properties of eutectic gold/tin

Young's modulus for eutectic gold/tin

For eutectic gold/tin, data on mechanical properties is very scarce. In [170] bulk samples of eutectic gold/tin were analyzed using a dilatometer. The room temperature value was found to be 70.6 GPa. In [152] a nanoindentation tool was used to measure the Young's modulus not only of eutectic gold/tin, but also for the individual phases making up the lamellae of gold/tin. A value of 74 GPa was found for eutectic gold/tin. Because the dilatometer measurement generally has a smaller margin of error, a weighted average value of 72 GPa was used in this thesis.

Poisson's ratio for eutectic gold/tin

Regarding the Poisson's ratio for eutectic gold/tin, only one independent source could be found. From a diagram given in [170] a value of 0.405 was extracted.

Coefficient of thermal expansion for eutectic gold/tin

In [170] different values for the Young's modulus of 80Au20Sn are given for heating and cooling and for temperatures below and above the $\zeta \rightarrow \zeta'$ transition temperature of 190°C.

Table 4.4: Coefficients of thermal expansion for eutectic Au/Sn during heating and cooling at temperatures below and above 190°C.

Reflow or soldering phase	CTE [ppm/°C]	
	T < 190°C	T > 190°C
Heating	16.2	17.0
Cooling	16.8	15.6

Because silicon deformations are caused by CTE differences during ramp-down, an average value of 16.2 ppm/°C based on the cooling phase values was used.

Material properties of silicon

Stiffness matrix for silicon

Because silicon is an anisotropic material, the Young's modulus and Poisson's ratio depend on the direction relative to the crystal lattice. For the three standard directions the values are given in Table 4.5.

Table 4.5: Young's modulus and Poisson's ratio for monocrystalline silicon for different directions in the crystal [171], [172].

Crystal orientation	Young's modulus [GPa]	Poisson's ratio
<100>	130.2	0.064
<110>	168.9	0.361
<111>	187.5	0.262

If a predominant direction is considered, the values for the according direction can be taken from Table 4.5. In this thesis, however, all directions were relevant, because three-dimensional deformations occurred during reflow and soldering. Therefore, the silicon material was better described by its 6x6 stiffness matrix. The stiffness matrix for silicon completely describes the mechanical behavior of silicon in all 3 dimensions by representing the correlation between the 6 strain values and the 6 stress values according to

$$\begin{pmatrix} \sigma_x \\ \sigma_y \\ \sigma_z \\ \tau_{xy} \\ \tau_{yz} \\ \tau_{zx} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & c_{13} & c_{14} & c_{15} & c_{16} \\ c_{21} & c_{22} & c_{23} & c_{24} & c_{25} & c_{26} \\ c_{31} & c_{32} & c_{33} & c_{34} & c_{35} & c_{36} \\ c_{41} & c_{42} & c_{43} & c_{44} & c_{45} & c_{46} \\ c_{51} & c_{52} & c_{53} & c_{54} & c_{55} & c_{56} \\ c_{61} & c_{62} & c_{63} & c_{64} & c_{65} & c_{66} \end{pmatrix} \cdot \begin{pmatrix} \varepsilon_x \\ \varepsilon_y \\ \varepsilon_z \\ \gamma_{xy} \\ \gamma_{yz} \\ \gamma_{zx} \end{pmatrix}, \tag{4.3}$$

where c_{11}, c_{12}, …, c_{65}, c_{66} are the 36 stiffness coefficients, ε_x, ε_y, ε_z are the direction strains, γ_{xy}, γ_{yz}, γ_{zx} are the shear strains, σ_x, σ_y, σ_z are the direction stresses and τ_{xy}, τ_{yz}, τ_{zx} are the shear stresses according to Voigt's notation [9]. The x, y, and z directions are oriented parallel to the [100], [010], and [001] directions of the silicon crystal lattice. In silicon many of the stiffness coefficients are zero and due to silicon's cubic symmetry only three independent stiffness coefficients exist, greatly simplifying formula (4.3):

$$\begin{pmatrix}\sigma_x\\\sigma_y\\\sigma_z\\\tau_{xy}\\\tau_{yz}\\\tau_{zx}\end{pmatrix}=\begin{pmatrix}c_{11}&c_{12}&c_{12}&0&0&0\\c_{12}&c_{11}&c_{12}&0&0&0\\c_{12}&c_{12}&c_{11}&0&0&0\\0&0&0&c_{44}&0&0\\0&0&0&0&c_{44}&0\\0&0&0&0&0&c_{44}\end{pmatrix}\cdot\begin{pmatrix}\varepsilon_x\\\varepsilon_y\\\varepsilon_z\\\gamma_{xy}\\\gamma_{yz}\\\gamma_{zx}\end{pmatrix}. \tag{4.4}$$

The values for c_{11}, c_{12}, and c_{44} are generally referred to as the elastic constants of silicon and can be found in the literature based on a coordinate system oriented along the <100> directions of the silicon crystal. The crystal lattice in silicon wafers used in this thesis was oriented such that the wafer flat was parallel to the <110> direction. Accordingly, with the chip design being parallel to the wafer flat, the chip edges were also oriented in the <110> direction. Therefore, the stiffness matrix in formula (4.4) had to be transformed by a 45° rotation. The transformation formulas were taken from [9]:

$$c'_{11}=\left(l_1{}^4+l_2{}^4\right)c_{11}+2l_1{}^2l_2{}^2(c_{12}+2c_{44})\,, \tag{4.5}$$

$$c'_{12}=2l_1{}^2l_2{}^2(c_{11}-2c_{44})+\left(l_1{}^4+l_2{}^4\right)c_{44}\,, \tag{4.6}$$

$$c'_{13}=c_{12}\,, \tag{4.7}$$

$$c'_{33}=c_{11}\,, \tag{4.8}$$

$$c'_{44}=c_{44}\,, \tag{4.9}$$

$$c'_{66}=2l_1{}^2l_2{}^2(c_{11}-c_{12})+\left(l_1{}^2-l_2{}^2\right)c_{44}\,, \tag{4.10}$$

where c_{11}, c_{12}, and c_{44} denote the original elastic constants, c'_{11}, c'_{12}, c'_{13}, c'_{33}, c'_{44}, c'_{66} denote the transformed elastic constants for <110> oriented silicon, and l_1 and l_2 are the direction cosines, defined as the cosine of the angle between the rotated x-axis and the original x- and y-axes. For a 45° rotation a value of cos (45°)= $1/\sqrt{2}$ is obtained for both l_1 and l_2. Using the new stiffness coefficients, formula (4.4) transforms into:

$$\begin{pmatrix}\sigma_x\\\sigma_y\\\sigma_z\\\tau_{xy}\\\tau_{yz}\\\tau_{zx}\end{pmatrix}=\begin{pmatrix}c'_{11}&c'_{12}&c'_{13}&0&0&0\\c'_{12}&c'_{11}&c'_{13}&0&0&0\\c'_{13}&c'_{13}&c'_{33}&0&0&0\\0&0&0&c'_{44}&0&0\\0&0&0&0&c'_{44}&0\\0&0&0&0&0&c'_{66}\end{pmatrix}\cdot\begin{pmatrix}\varepsilon_x\\\varepsilon_y\\\varepsilon_z\\\gamma_{xy}\\\gamma_{yz}\\\gamma_{zx}\end{pmatrix}. \tag{4.11}$$

The elastic constant for the <100> and <110> orientation of silicon are listed in Table 4.6.

Table 4.6: Elastic constants for silicon for the <100> and <110> orientations of the coordinate system (values for <100> silicon were taken from [173], while values for <110> silicon were calculated using formulas (4.5) to (4.10)).

Elastic constants for <100> silicon	
Symbol	Value [Gpa]
c_{11}	165.78
c_{12}	79.62
c_{44}	63.94

Elastic constants for <110> silicon	
Symbol	Value [Gpa]
c'_{11}	194.48
c'_{12}	35.24
c'_{13}	63.94
c'_{33}	165.78
c'_{44}	79.62
c'_{66}	50.92

Coefficient of thermal expansion for silicon

The thermal expansion of silicon strongly depends on the temperature range that is being considered. For room temperatures values around 2.65 ppm/°C can be found in the literature [174], [175]. In [176] the CTE of silicon was determined for a large temperature range and compared to results from 7 earlier publications. An approximation function for temperatures between 100 K and 1400 K was calculated that describes the CTE of silicon within 5% to 6%:

$$\begin{aligned}\frac{\alpha_{Si}(T)}{\frac{ppm}{K}} = &-15.2459 + 3.43026\,\ln\left(\frac{T}{K}\right) - 5.394\cdot 10^{-3}\left(\frac{T}{K}\right)\\ &+ 1.286\cdot 10^{-6}\left(\frac{T}{K}\right)^{2} - 88.6853\left(\frac{T}{K}\right)^{-1},\end{aligned} \tag{4.12}$$

where α_{Si} is the coefficient of thermal expansion of silicon and T is the temperature in units of Kelvin. Using formula (4.12), an average CTE for silicon of $\alpha_{Si} = 3.12$ ppm/°C was calculated for the temperature range from 20°C to 280°C.

Material properties of polyimide PI2611

PI2611 is distributed by HD Microsystems. The manufacturer only provided the Young's modulus and coefficient of thermal expansion for room temperature [177]. The Poisson's ratio was taken from [178]. These values (E_{PI2611} = 8.5 GPa, α_{PI} = 3.0 ppm/°C, and ν_{PI} = 0.35) may change with temperature. In [179] for example, the mechanical properties of polyimide from another supplier (type Upilex-S by Ube Industries, Ltd.) were analyzed as a function of temperature. The Young's modulus was found to drop from 8.3 GPa at room temperature to 5 GPa at 277°C, while the Poisson's ratio increased from 0.35 to 0.38 (values extracted from diagrams in [179]). However, due to the lack of similar data for the temperature dependence of the mechanical properties of PI2611, the room temperature values were used for the whole reflow and soldering temperature range in this thesis.

Possible errors for material parameters

Mechanical properties generally change with temperature. For example, according to [155], the Young’s modulus of nickel drops from 206.85 GPa to 179.27 GPa when the temperature is increased from room temperature to 200°C. Also, phase transitions or annealing may occur at elevated temperatures, significantly changing the mechanical properties. For example eutectic gold/tin undergoes a phase transition at 190°C, which results in a lower Young’s modulus below 190°C [170]. Moreover, intrinsic stress, as observed after plating, changes significantly during heating. Kebabi et al. [168] found an increase of intrinsic stress from 76 MPa in as plated nickel layers to 374 MPa after storage at 250°C for 1 hour. These nickel layers were fabricated using the same plating bath as in this thesis. Even though some of the thermal stress at elevated temperatures is compensated by the deformations of flexible systems, it was expected that relaxation processes still took place reducing thermal stress. However, the extent of these dynamic changes is unknown. Because of the multitude of factors influencing the actual mechanical properties of the flex systems, the simulation results needed to be correlated to measurements of deformations of actually reflowed samples.

The properties of electroplated layers not only depend on the plating parameters, but also on the material, crystallinity, and crystal orientation of the substrate. For example, if nickel is plated at low current densities (≈ 2 mA/cm²) onto electroplated copper having a disoriented surface, the resulting nickel layer has isotropic properties with small amounts of (111) and (110) oriented crystals [164]. At larger current densities (≈ 20 mA/cm²) the resulting nickel layer is (100) oriented. However, if a gold layer that was deposited by physical vapor deposition (PVD) is used as a plating base, the gold layer may show a predominant (111) surface, leading to a (110) oriented nickel layer at low plating currents and a (100) oriented nickel layer at larger currents. This analysis was made on samples that were fabricated using the same electroplating equipment as in this thesis and are described in detail in [164]. The grain size also influences mechanical properties. Generally, smaller grains show less ductility, because smaller grains exhibit a larger grain boundary density and these grain boundaries are obstacles for slipping of atomic layers. At electroplating current densities used in this thesis, the grain size increases with increasing current density (in accordance with the inhibition mechanism). At much larger current densities, the grain size may, however, decrease again [164].

One factor that can influence material properties is the size of the object being considered [115]. Small solder joints, for example, may have different properties than the bulk material [59]. In [114] the properties of a small gold/tin solder joint (size: 30 µm to 40 µm) were analyzed using a nanoindentation tool on the cross section of such a solder joint.

Even contamination with trace elements in the nickel sulfamate electrolyte can change the intrinsic stress in the deposited nickel at the rate of 0.1 MPa per ppm contamination [163].

In [180] it was found that BPDA-PDA type polyimide (which is similar to PI2611 used in this thesis) shows anisotropic properties. Specifically, the in-plane and out-of-plane CTEs of spin-on polyimide are very different. In 5 µm thick polyimide films the lateral CTE was determined as 5.7 ppm/°C, while the vertical CTE was almost 20 times larger (108 ppm/°C). Moreover, the lateral CTE was found to increase with polyimide thickness, while the vertical CTE was found to decrease. Two possible explanations are given, stating that either the regular orientation of molecules, caused by the spin-on process, was to blame, or that the biaxial confinement of the spin-on layer during solvent evaporation resulted in biaxial tensile stress, which lead to anisotropic material properties.

4.2 Model simplifications

When defining a simulation model, the similarity between the model geometry and the actual sample geometry should be as good as possible, without increasing the complexity to a point where the simulation time increases to impractical values or the model cannot be solved by the used simulation computer at all. These conflicting aims are best tackled by reducing the model complexity significantly and comparing the extracted simulation results to measurement results of actual samples. If the differences are small, the model is a useful tool in order to understand the causes of observed sample properties and to predict the influence of design changes on sample properties in prototype fabrication. The model may also be used to optimize existing products. The simplifications for the FEM model used in this thesis concerned the material properties or were of geometric nature.

Major geometric simplifications:

- model size reduced to a square-shaped fraction of the original sample size by replacing the complete sample design with a small unit of the spatially periodic design that can be used as a building block for the entire chip (in Design C (see Figure 3.1), for example, the chip perimeter had almost the same track and bond density (per chip area) as the center of the chip, resulting only in a small error),
- in Design C the polyimide substrate was 50% wider compared to the silicon chip, which was taken into account by increasing the polyimide thickness by 50% instead,
- chips and substrates with homogeneous metal geometry (differences of the periphery not taken into account) and homogeneous silicon thickness distribution,
- no stress or thickness distribution for electroplated layers across the wafer or within individual samples,
- cuboid-shaped electroplated gold tracks with vertical side walls and without round corners or a rugged edge shape (caused by usage of prototyping polymer masks for resist exposure),
- no thermal misalignment (which usually was caused by different thermal expansion of the chip and substrate at the soldering temperature of 280°C),

- electroplated layers considered as a homogeneous material without grains and grain boundaries,
- solder considered as a homogeneous material without voids or lamellar microstructure of alternating AuSn and Au_5Sn layers,
- no oxidation of tin or gold/tin,
- no surface roughness for electroplated layers or wet etched silicon,
- thermal silicon dioxide neglected ($t_{SiO2} \approx 300$ nm),
- no over etching (especially of nickel) during plating base etching taken into account,
- Cr-Au plating base neglected ($t_{Cr} = 50$ nm, $t_{Au} = 200$ nm),
- gravitational impact on solder shape of reflowed gold/tin ignored (for long gold/tin lines a truncated cylinder and for round gold/tin bumps a truncated sphere was used),
- impact of weight piece during soldering represented by a homogeneous 30% overhang of the solder for all bump diameters (that is the solder's maximum diameter within an individual interconnect was equal to $Ø_{pad} + t_{solder} \cdot 1.3$, where $Ø_{pad}$ is the pad diameter and t_{solder} the solder thickness). Due to the surface tension of the liquid solder the shape of the solder surface generally resembles a truncated sphere. However, with an external force being applied to hold down the chip during soldering, the liquid solder slightly bulged out to the sides. This effect lead to an overhang of approximately 30% relative to the solder thickness. Because the exact overhang has only minor influence on the simulated deformation, the 30% overhang was used for all simulations.
- the thicknesses of electroplated layers were assumed to be equal to the set thickness.

Major simplifications for the material properties:

- use of linear material properties (examples: Young's modulus assumed to be independent of stress; coefficient of thermal expansion averaged over the considered temperature range). Where necessary, average values were calculated for the considered temperature range.
- material properties were mostly taken from the literature (the exact material properties often depended on the mode of deposition and the processing parameters),
- directional dependence of Young's modulus in spin-on polyimide ignored,
- eutectic solder with 80 wt.% gold and 20 wt.% tin assumed (deviations may occur due to inhomogeneous electroplating),
- intermetallic compounds forming at the interface between nickel and gold/tin were neglected. In [57] it was shown that a Ni_3Sn_2 layer develops at this interface and grows with a square root of time dependence. After 225 s at 175°C this layer reached a thickness of 0.8 µm, which was a good lower estimate for the thickness of this layer in this thesis, where the reflow temperature was above 175°C for typically 240 s.

While the impact of the different simplifications on the simulation results may have varied, the overall impact was assumed to be small. However, a comparison of the simulation results and measurements was necessary to prove the correctness of the model.

Apart from the mentioned simplifications regarding material properties and model geometries, the dynamics of the reflow and soldering process were neglected. These include for example the heat distribution, which was assumed to be homogeneous.

By implementing 2 symmetry planes (i.e. the xz plane and the yz plane), the model size was further reduced by a factor of 4, without increasing the error. For Design C only a single bump remained that was used to build the model geometry, and for the Design A (see Table 3.1) two gold/tin lines remained.

The silicon chip comprises 3 main areas with different bump diameters. Each of these areas deforms differently during the reflow and soldering process. Therefore, a different model geometry was used for each of the 3 main areas.

4.3 Deduction of model geometry from sample design

With a width-to-thickness ratio of more than 150, the chips required a very large number of elements to build the simulation mesh and obtain good quantitative results. The UBM, a two-layer structure with a total thickness of only 4 µm, and the sharp solder edge, which resulted from the cylindrical or spherical approximation of the solder shape for the cool-down phase, further increased the element count. For bumps with small widths, the required number of elements was beyond the capabilities of the academic license simulation software used, which allowed for a maximum of 500,000 nodes, corresponding to approximately 100,000 elements. Also, the simulation time increased substantially with increasing node count. Therefore, the model was reduced to a small square section of the original sample.

In Figure 4.1, this is shown for the ramp-up model of a thin chip with 200 µm wide gold/tin lines on a gold-nickel UBM. Only the center area (making up 1/9 of the total chip area) was simulated. The size of this area was adapted for each chip design to comprise 4 lines and spacings. Hence, the simulated model's expansion parallel to the chip surface was between 1600 µm (for lines 200 µm wide) and 200 µm(for lines 25 µm wide). Using the two symmetry planes A and B, the size of the FEM model was further reduced by 75%, without any loss of accuracy.

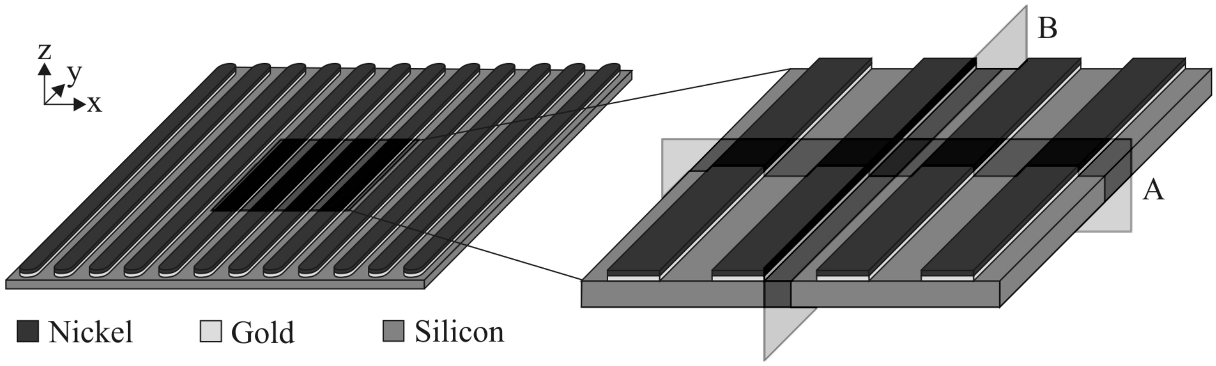

Figure 4.1: Deduction of simulation geometry from the sample design of a thin chip with 200 µm wide gold/tin lines on a gold-nickel UBM. The left schematic shows the complete geometry for ramp-up (without solder) and the right schematic shows an enlargement of the center area comprising 1/9 of the original sample. The symmetry planes A and B were used in the simulation to further reduce the model size by a factor of four.

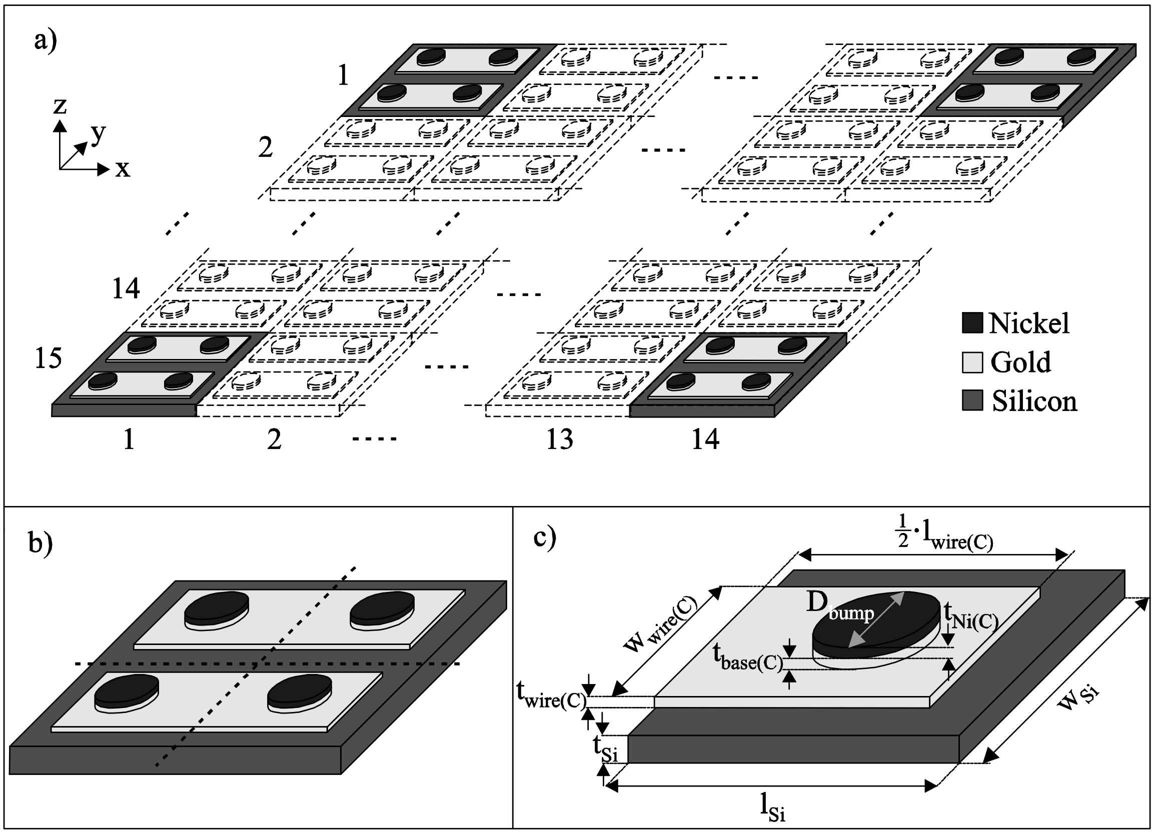

Figure 4.2: Deduction of reflow simulation model for silicon chips prepared for Daisy chains. The schematics show the ramp-up geometry (without solder); a) Daisy chain area without U-shaped tracks on the sides; b) unit of 4 UBMs and 2 tracks on a piece of the silicon chip. Two symmetry planes are indicated by dashed lines; c) actual simulation model deduced from b) by making use of the two symmetry planes. All major dimensions are also defined in this schematic.

Similarly, in Figure 4.2, the model size reduction for a chip prepared for Daisy chains is shown. The main Daisy chain area (disregarding the U-shaped tracks on the side) was a homogeneous matrix of 14 by 15 square units. Each unit comprised 2 gold tracks and 4 round Au-Ni-Au/Sn bumps. One such unit was simulated, reducing the number of required elements by a factor of 210. Once again two symmetry planes were used to further reduce the model size by 75%. The size of the models for the reflow of polyimide tapes prepared for Daisy chains and for completely soldered polyimide-silicon flex systems were reduced in the same way. The effect of the model size reduction on simulation results was analyzed for thin chips with lines of gold/tin. For each line width, the results from two models with different size reductions were compared. The geometries of the two-line model and the four-line model are shown in Figure 4.3.

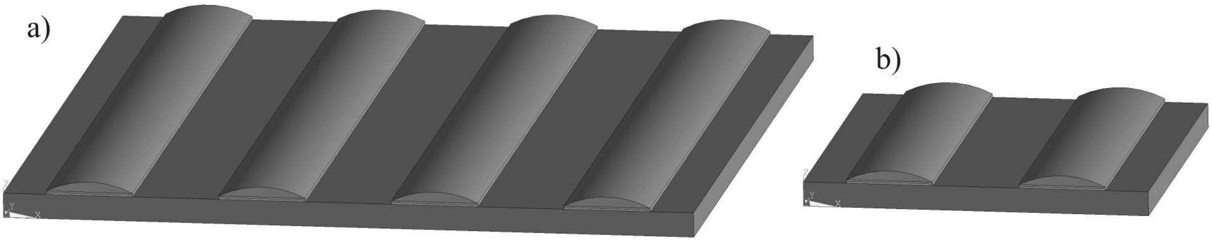

Figure 4.3: Simulation model for the ramp-down of 28 µm thick chips with 100 µm wide gold/tin lines; a) model based on 4 lines; b) model based on 2 lines.

The simulated local deformations have a maximum deviation below 1% between the 2-line and 4-line model for all line widths. For the radius of curvature the deviations between the two models were generally around 1%. Only for the case of 25 µm and 50 µm line widths the differences were larger (about 6% and 2%, respectively). For line widths larger or equal to 75 µm (corresponding to a minimum simulated silicon area of 300 µm x 300 µm) the model size reduction only played a minor role. Detailed results for this simulation study are given in Table 4.7.

Table 4.7: Radius of curvature extracted from the ramp-down simulation of thin chips with a silicon thickness of 28 µm and different line widths. The simulations were carried out with the standard model based on 2 lines and a larger model with 4 lines. The relative deviation between the simulation results is also given.

Line width [µm]	Radius of curvature [mm]		Deviation [%]
	Model with 4 lines	Model with 2 lines	
25	15.642	16.563	5.89
50	14.862	15.144	1.90
75	14.569	14.737	1.15
100	14.480	14.608	0.88
150	14.618	14.757	0.95
200	15.272	15.474	1.32

4.4 Geometries

The complete geometry was made up of a hierarchy of geometric elements of different dimensions. The volumes represented the most abstract geometric form and were defined by their bordering areas. The areas in turn were defined by their bordering lines, which were defined by the two key points at their ends (for straight lines). For round lines, the radius of curvature and a third keypoint, defining the plane of the arc, were additionally required. To define this complex geometry, two approaches existed:

- the top-down approach:

  only the geometry of the volumes was defined by the user. All other objects (areas, lines, and keypoints) were automatically generated by the program and numbers were allocated to each object. The numbering algorithm was very complex and difficult to predict. Also, in the top-down approach double objects were created in a dimension where objects of the next higher dimension touched (for example, two neighboring volumes had 2 areas at their interface). These double objects later had to be merged into a single object, resulting in the allocation of new object numbers.

- the bottom-up approach:

  all objects were defined by the user, starting with the keypoints, followed by the lines, then the areas, and finally the volumes. This procedure was very time consuming, but had the great advantage that all objects were numbered according to the user's preferences and that no object merging was necessary, resulting in a more stable model with better control of the sequential simulation stages. Because a large number of simulations with different geometrical and/or material parameters were programmed using macro code, the model had to be as stable as possible to avoid simulation halts or software crashes. Therefore, the bottom-up approach was chosen for all simulation models.

The FEM model geometries for Design A are shown in Figure 4.4 with a small section of each geometry enlarged to the right. Different layer sequences and/or geometries were required for the 3 steps of the reflow process.

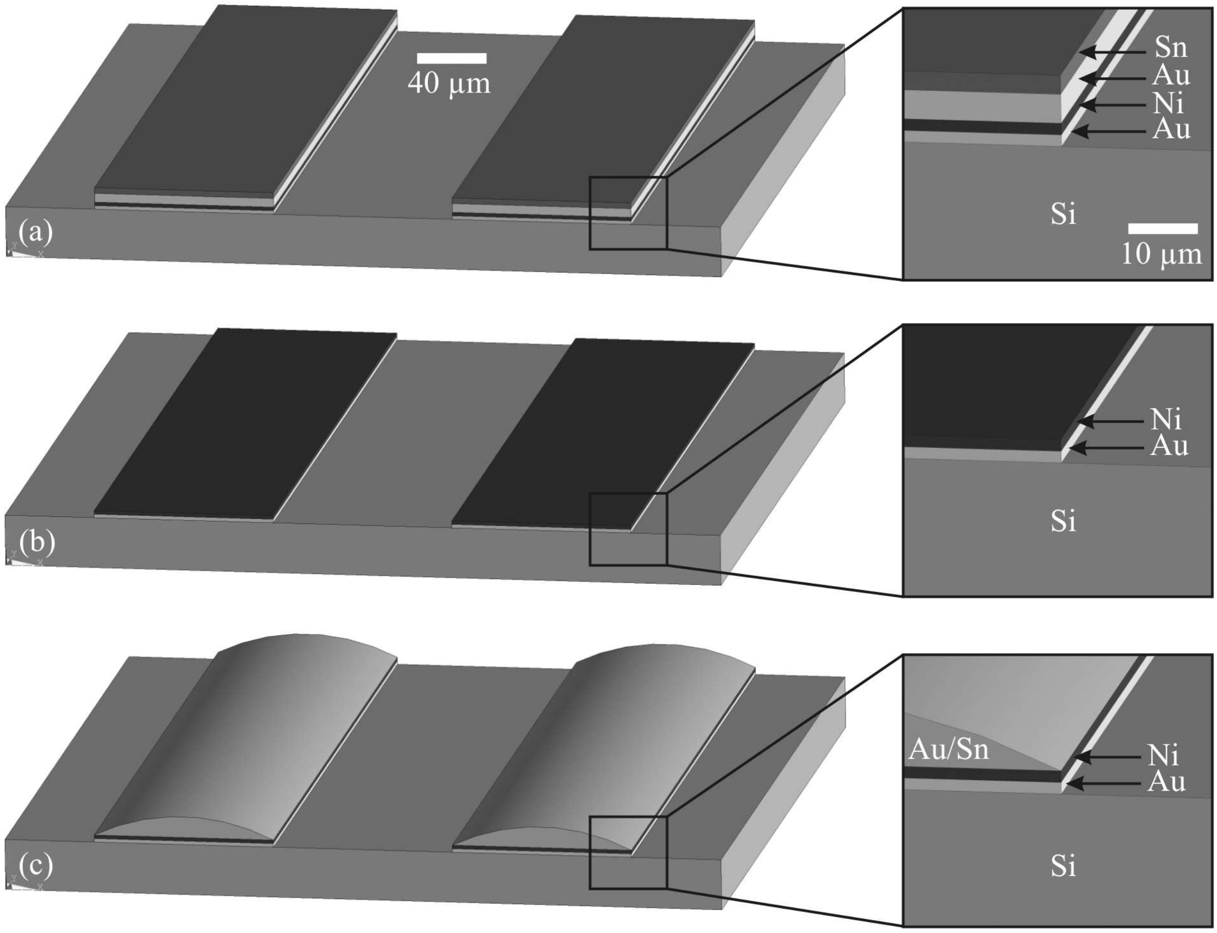

Figure 4.4: FEM model geometries for different stages of the reflow process of thin chips with a silicon thickness of 20 µm and 100 µm wide gold/tin lines on gold-nickel UBMs. The images were exported from the original ANSYS models; a) geometry of as-plated sample for ramp-up simulation, valid for temperatures below the melting point of tin; b) geometry for temperatures above the melting point of gold/tin; c) geometry for ramp-down simulation with the gold/tin having the shape of a truncated cylinder.

In Figure 4.5 similar images for Design C are shown. The first 3 images represent the different stages of the reflow process of a silicon chip, while the 4th image represents the soldering process.

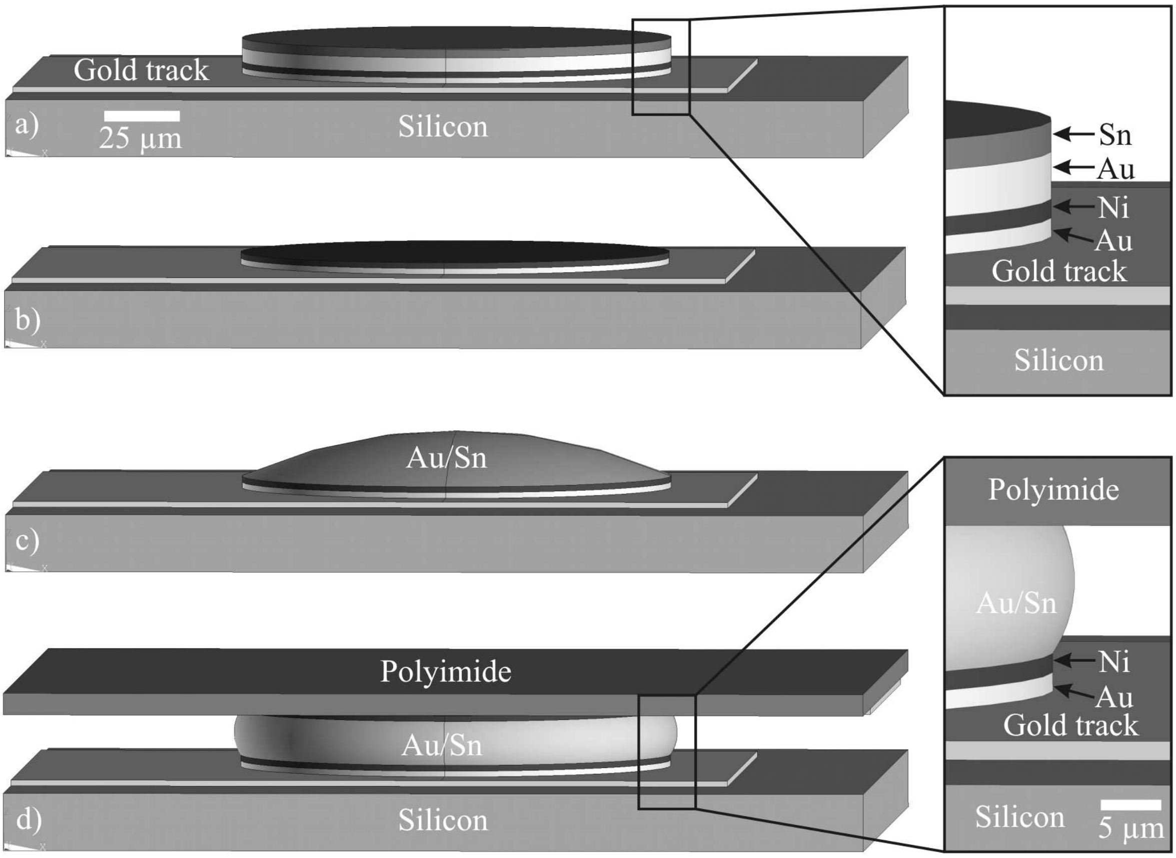

Figure 4.5: FEM model geometries for different stages of the reflow or soldering process of samples with Design C (for a bump diameter of 150 µm and a silicon thickness of 20 µm); a) geometry of as-plated silicon chip with gold track and Au-Ni-Au-Sn bump layer sequence for simulation of the ramp-up stage of the reflow process for temperatures below the melting point of tin. To the right, a close-up is shown that includes labels for all layers; b) geometry for temperatures above the melting point of gold/tin; c) geometry for ramp-down simulation with the gold/tin having the shape of a truncated sphere; d) geometry for simulation of completely soldered system; the gold track and gold-nickel UBM on the polyimide tape are only partly visible. To the right, a close-up is shown, again with labels for all layers.

4.5 Meshing

The finite element method is based on the approximation of the model geometry by individual simulation entities called *elements*. All elements together are referred to as the mesh of the model and the generation of these elements is called *meshing*. For all volume meshes, the structural solid 10-node element SOLID187 was used, because of its tetrahedral shape (necessary for automatic meshing), the anisotropic material property capabilities (necessary to model silicon), and quadratic displacement behavior (which delivered better results for irregular meshes). In order to extract data at a homogeneous distribution of locations on an area (greatly simplifying further data analysis), this area needed to be meshed in a mapped style. For this, the 2D 8-node element PLANE183 was used, which also had quadratic displacement capabilities. The general geometries of SOLID187 and PLANE183 elements are shown in Figure 4.6.

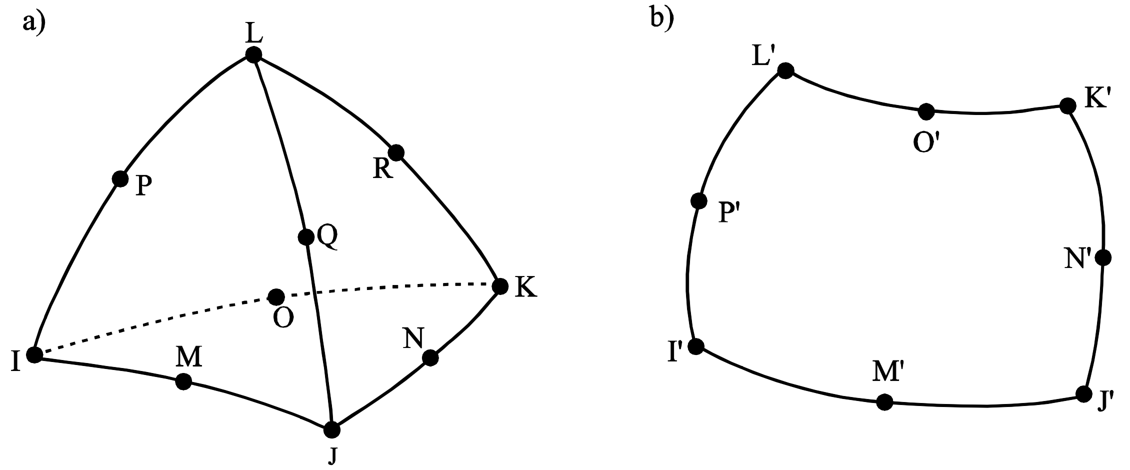

Figure 4.6: Geometries of two ANSYS meshing elements; a) the element SOLID187 is a 3D 10-node structural solid with a tetrahedral shape. The 10 nodes (marked as black dots) are labeled with letters (corner nodes: I, J, K, and L; midside nodes of the base plane: M, N, and O; all other midside nodes: P, Q, and R). Because every line comprises 3 nodes, they can assume a bow. This is generally referred to as quadratic displacement; b) the element PLANE183 is a 2D 8-node element with quadratic displacement capabilities. It is suitable for mapped meshing of areas. In the schematic, the corner nodes and midside nodes are labeled with letters I' to L' and M' to P', respectively.

In Figure 4.7 the ramp-down geometry of a silicon chip prepared for Daisy chains is shown. The model was meshed with tetrahedral Solid187 elements.

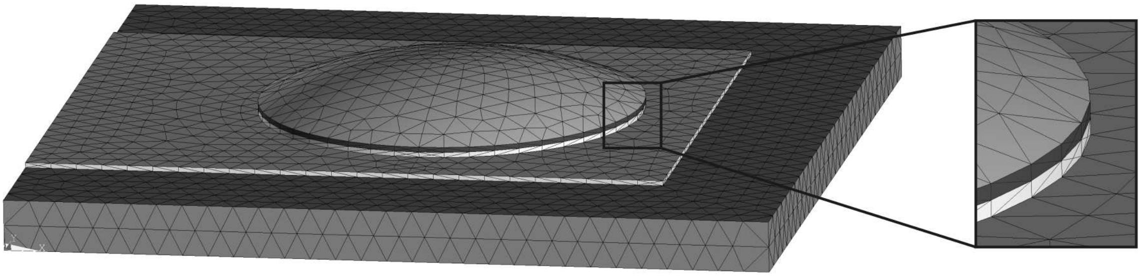

Figure 4.7: Meshed simulation model for the ramp-down of a silicon chip with a thickness of 20 µm. The chip was prepared for Daisy chains and included a gold track with a thickness of 2 µm, a circular gold-nickel UBM (t_{Au} = t_{Ni} = 2 µm) with a diameter of 150 µm and a eutectic gold/tin solder cap.

4.6 Solution

In all simulations, the sparse direct solver was used, which involves a direct elimination process starting with the decomposition (factorization) of the solution matrix [K] into lower and upper triangular matrices, [K] = [L]·[U]. Then forward and back substitutions using [L] and [U] are made to compute the solution vector {u} [181].

The loads were applied as a single incident without dynamic effects such as inertia or damping. The analysis type was therefore called static (as opposed to transient dynamic analysis, which may include loads that change over time). Although a static analysis was carried out, the solution process could be split up in smaller load steps, to achieve more accurate results. The initial number of substeps was set to 5 and, depending on the accuracy of the results, the auto time stepping function could again solve the model with a larger number of substeps. The simulations were carried out without nonlinear material properties, which could have been used to model changing material properties (for example in the case of large deformations, plasticity, creep, or stress stiffening). In the case of very large deformations the shape of the model might have changed significantly. If necessary, this could have been taken into account via the nonlinear geometry command (NLGEOM, on), which would have allowed for the repetitive adaptation of the equation matrix to the new model shape. In this thesis, however, the model size was reduced such that the deformations within the model were small enough, not to require an adapted solution matrix.

4.7 Results extraction

The local and global deformations were calculated from displacement data at certain points of each simulated model. In order to read out displacement data at a certain point, this location had to be occupied by a node. This was achieved by defining so-called hard points at the read-out positions before meshing. The mesh was then created such that the predefined hard points were included. The coordinates of 3 points located in the same plane were required to calculate the radius of curvature using the formula in Appendix B.

The results extraction could be divided into four groups:

- global deformations (bow) in thin chips with long gold/tin lines: 15 hard points were defined on the bottom silicon surface such that the bow along the two gold/tin lines and along 3 more lines (on the left, right, and in between the gold/tin lines) could be determined. In Figure 4.8 a schematic with the 15 hard points is shown. The radius of curvature R_2 underneath the left gold/tin line, for example, was calculated from the displacement data of locations A2, B2, and C2. The bow underneath the gold/tin lines was always slightly larger than halfway between the gold/tin lines. To obtain a representative value for the radius of curvature R_{avg}, the 5 values (R_1, R_2,…, R_5) were averaged such that the bow underneath and between the gold/tin lines were weighted equally. The following formula was used:

$$\mathrm{R_{avg}} = \frac{1}{8} \cdot (\mathrm{R_1} + 2 \cdot (R_2 + R_3 + R_4) + R_5)\,, \tag{4.13}$$

- local deformations in thin chips with long gold/tin lines: the hard points labeled A and B in Figure 4.8 were used to calculate values for the local deformation D underneath A2, A4, B2, and B4. For example, to calculate the local deformation D_{A2} underneath A2 the following formula, based on the z coordinates z_{A1}, z_{A2}, and z_{A3} underneath A1, A2, and A3, respectively, was used:

$$\mathrm{D_{A2}} = \frac{1}{2} \cdot (\mathrm{z_{A1}} + \mathrm{z_{A3}}) - \mathrm{z_{A2}}\,, \tag{4.14}$$

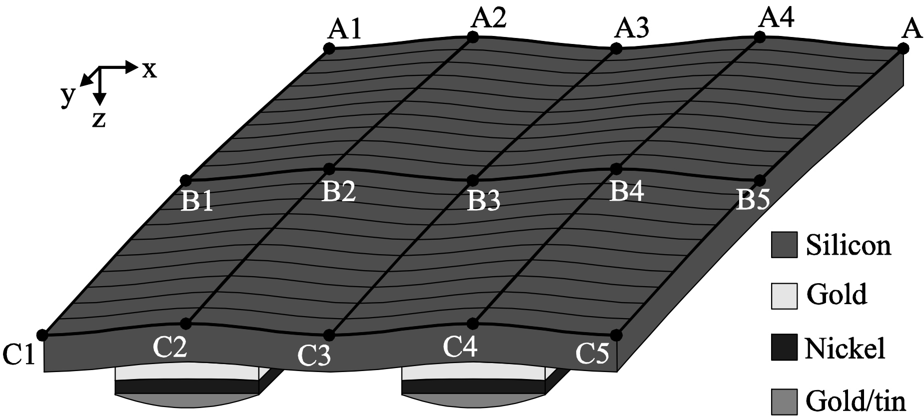

Figure 4.8: Definition of readout locations for simulations of chips with long lines of gold/tin.

- global deformations in thin chips prepared for Daisy chains: 9 evenly distributed hard points were added to the silicon surface of each simulation model and used to obtain 3 radii of curvature (for bending around the y-axis). Similarly as above, a weighted average was calculated from these 3 values. For polyimide substrates prepared for Daisy chains, the same method based on 9 hard points in the polyimide surface was used. For silicon-polyimide flex systems only hard points in the silicon surface were used to extract displacement data from the simulation model,
- local 3D profile of silicon surface in chips prepared for Daisy chains: the silicon surface area was meshed with a regular pattern of 30 by 30 square elements of type PLANE183. Each element had 4 corner nodes and 4 midside nodes, leaving a nodeless spot in the center of each element (see Figure 4.6 b) for a schematic representation of an individual element). Displacement data of all 2821 nodes in the silicon surface area was extracted from the model and used to interpolate the displacement at the center of each of the 900 planar elements. The combined data was then used to generate 3D graphs and analyze local deformations.

4.8 Linearity of the simulation models

Most simulations of the reflow process in this thesis consisted of 3 different steps with separate results, which were then superimposed. In order for this superposition to be correct, the model had to be linear, meaning that the simulation result (for example the bending angle) had to be proportional to the input variable (for example the thermal stress). For a wide variety of simulations, the simulation results were compared to results based on thermal stress reduced by 50%. The bending angle or local deformations were expected to be reduced by 50% as well. Indeed, the observed reduction was typically in the range of 49.9 to 50.0%. In the ramp-up simulation of local 3D profiles of the silicon surface, the average deviation was found to be below 1 nm and the maximum deviation was 1.6 nm, which was much smaller than the ramp-up displacement values, typically in the range of a few microns. Therefore, the superposition of individual simulation results only lead to a small acceptable error.

5 Deformations

This chapter describes all aspects of deformations in flex silicon and polyimide tapes during reflow and soldering of gold/tin bumps and interconnects. Several examples of samples in their deformed or undeformed state are shown in Figure 5.1.

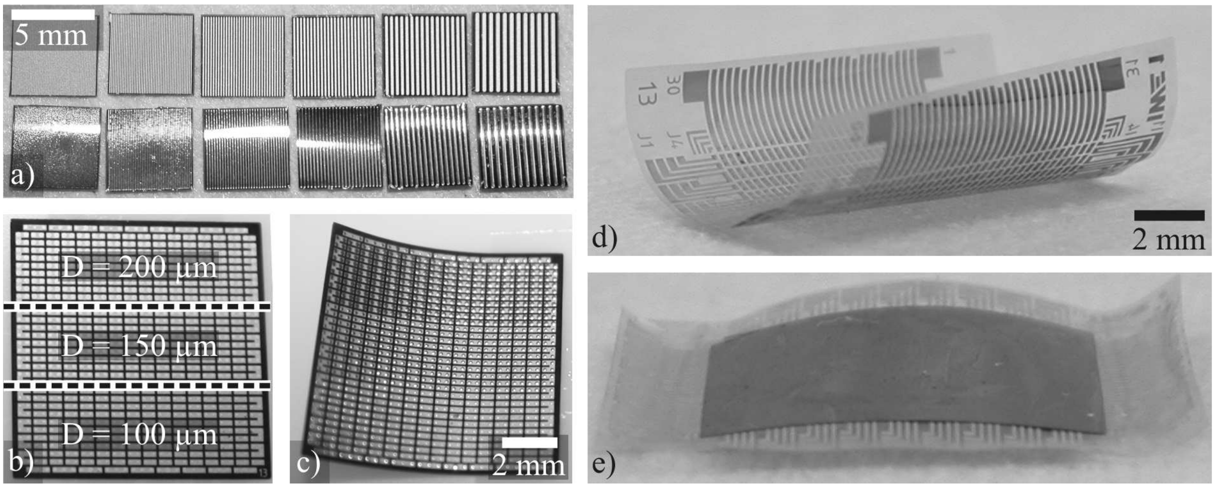

Figure 5.1: Reflow or soldering induced deformations in different samples; a) silicon chips with long lines of gold/tin (running from top to bottom) with track widths and spacings of 25 µm, 50 µm, 75 µm, 100 µm, 150 µm, and 200 µm. The top row shows as-plated samples, which were not bent, and the bottom row equivalent samples after reflow with their characteristic bow; b) as-plated silicon chip prepared for Daisy chains and Kelvin contacts. The areas with different bump diameters are separated by dashed lines and labeled with their respective diameters; c) a similar chip after reflow. The chip bow and its dependence on bump diameter is clearly visible; d) polyimide tape prepared for Daisy chains and Kelvin contacts after reflow. The bow was much larger than in the reflowed silicon sample; e) soldered silicon-polyimide system with functional Daisy chains and Kelvin contacts.

In order to facilitate the understanding of the section on measurement techniques a brief description of observed deformation types and possible causes for these deformations is given first. Because the measurement techniques had to be developed in parallel to the sample fabrication, reflow, and soldering processes, some initial experimental results are included in the measurement techniques section. Several sections follow that describe and compare measured and/or simulated deformations in

- thinned 4" wafers with different electroplated metal layers,
- reflowed thin chips with gold/tin lines (Design A),
- reflowed thin chips with gold/tin bumps for Daisy chains and Kelvin contacts (Design C),
- soldered silicon-polyimide flex systems (Design C),
- reflowed thin chips with round gold/tin bumps (Design B).

In each of these sections the simulation model is adapted or refined to improve agreement with the measurement results.

5.1 General notes on deformations

5.1.1 Temporary deformations during reflow

Apart from the measurable deformations after the reflow process, many different temporary deformations occurred during the reflow process. These are schematically shown in Figure 5.2. The coefficient of thermal expansion of silicon was typically 4 to 7 times smaller than for the electroplated metals and eutectic gold/tin solder. Therefore, temperature changes lead to deformations. Whether the deformation after reflow was convex (as shown in Figure 5.2 e)) or concave primarily depended on the silicon thickness, as will be shown later. For extremely thin silicon samples, the concave deformation at the solidification point of 280°C was so large that it could not be fully inverted during ramp-down, resulting in a concave deformation. The magnitude and direction of the local deformations in Figure 5.2 were valid for samples where the silicon was stiff compared to the metal layers (that is where the silicon was thicker or at least not significantly thinner than all metal layers together). An experimental confirmation of these expected deformations is presented in Section 5.4.2.

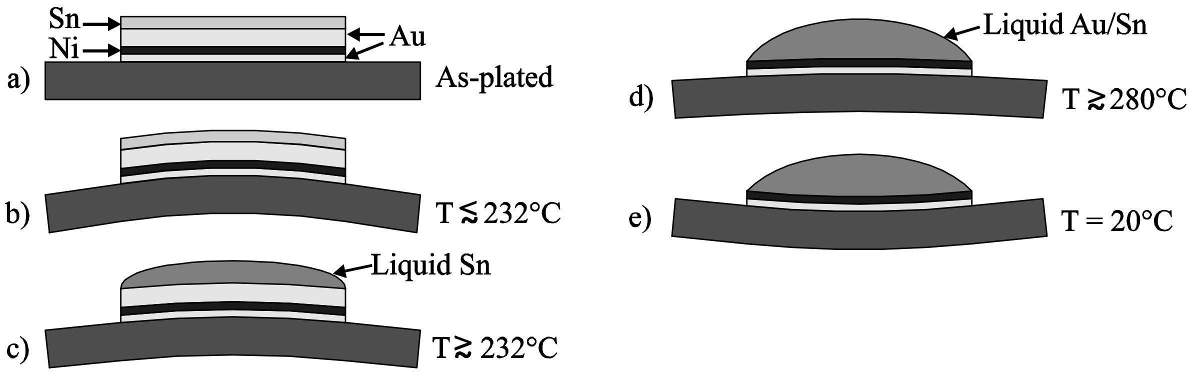

Figure 5.2: Schematic cross section of expected temporary and final local deformations in ultrathin silicon chips with gold/tin bumps and gold+nickel UBM during reflow; a) as-plated layers; b) concave deformation just before reaching the melting point of tin; c) slightly reduced concave deformation upon reaching the melting point of tin; d) further reduced concave deformation after formation of eutectic 80Au20Sn; e) convex deformation after cooling to room temperature.

5.1.2 Definition of local and global deformations

In this thesis, the term *local deformation* refers to the height difference D on the backside of the silicon chip between the position directly underneath a bump and the position in the middle between this bump and an adjacent bump. A graphical definition of D is given in Figure 5.3. Each of the local deformations not only resulted in a local curvature of the silicon surface, but also in a small bending angle. Globally, these bending angles added to a macroscopically visible chip bow as can be seen in the schematic in Figure 5.4. The bending angle not only defines the amount of bending, but also its direction. If the chip's back has a convex shape, the bending angle is defined as negative. The term *chip bow*, however, can refer to convex or concave deformations. So large bow can either mean a very large positive bending angle or a very small negative bending angle. For an actual image of global deformations see Figure 5.16 on page 112.

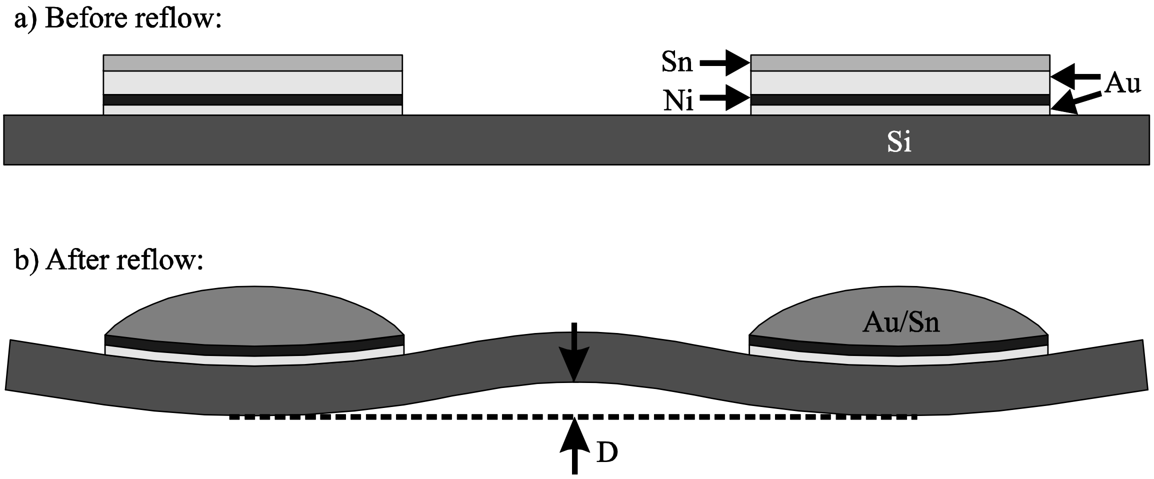

Figure 5.3: Schematic cross section of 2 gold/tin bumps + gold-nickel UBM on thin silicon a) before and b) after reflow. In the bottom schematic the local silicon deformation D is defined. For clarity of presentation, the local silicon deformation caused by the reflow is not to scale.

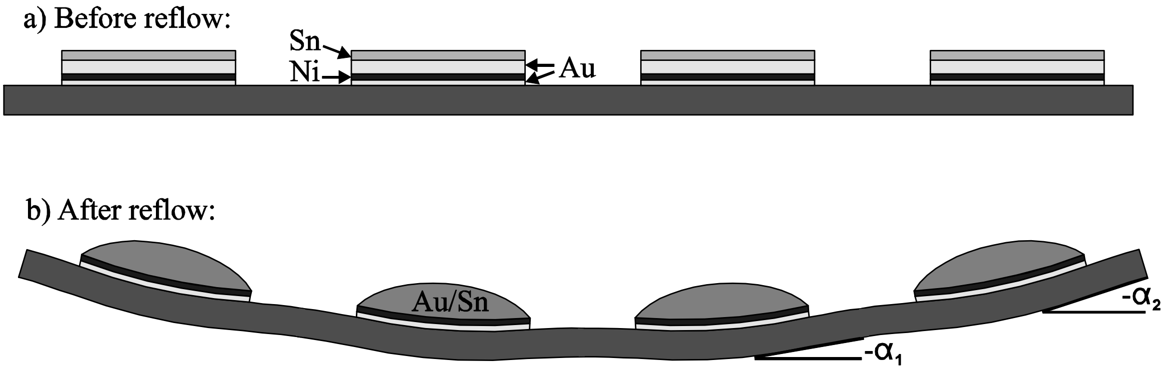

Figure 5.4: Schematic cross section of 4 gold/tin bumps + gold/nickel UBM on thin silicon a) before and b) after reflow. In the schematic after reflow the global silicon deformation after reflow is well visible. With each bump the cumulative bending angle α increases (|α_2| > |α_1|, etc...). For clarity of presentation, the silicon warpage caused by the reflow is not to scale.

5.2 Measurement techniques

5.2.1 Measurement of chip thickness

The silicon thickness had a large influence on deformations during reflow and soldering. Therefore, knowledge of the silicon thickness was of profound importance to understand the type and magnitude of deformations. The silicon chips in this thesis were generally lightweight and flexible and could be easily damaged during measurements. Therefore, special attention had to be given to chip thickness measurement techniques.

5.2.1.1 Chip thickness measurement using a micrometer gauge

Micrometer gauges were used to measure chip thicknesses with a precision of about 2 µm. However, due to the direct contact of the micrometer gauge and the chip, damage was probably induced in the silicon surface, increasing the risk of breakage during bending. For very thin chips, the measurement itself sometimes resulted in immediate chip breakages. The micrometer gauge had a mechanism, which limited the maximum momentum applied to the adjusting screw and hence reduced, but did not eliminate, the risk of surface damage. Another problem with micrometer gauges is that any structures on the silicon, like gold/tin bumps, may have been deformed or damaged during the measurement. This was, of course, not intended and would have lead to incorrect readings. Altogether, micrometer gauges were used to measure chip thicknesses, if surface damage could be tolerated.

5.2.1.2 Chip thickness measurement using profiler scans

Using a Tencor profiler P-10 allowed to precisely measure step heights down to about 10 nm. While the profiler needle moved sideways over the surface its distance to the wafer chuck was measured and recorded. In order to measure the thickness of thin silicon chips, the needle had to move across the chip edge. During the measurement, the chip had to be secured by vacuum for two reasons:

- the needle exerted a small, but sufficient force on the chip to move it sideways,
- thin chips were often bent. Therefore, when the needle passed the chip edge, the measured step height was often much larger than the actual silicon thickness.

The fixation of the chip by vacuum, however, lead to deformations in turn, falsifying the measurement. This can be seen in the profiler scan in Figure 5.5. Therefore, profiler scans were in general not suitable to measure the thickness of thin chips.

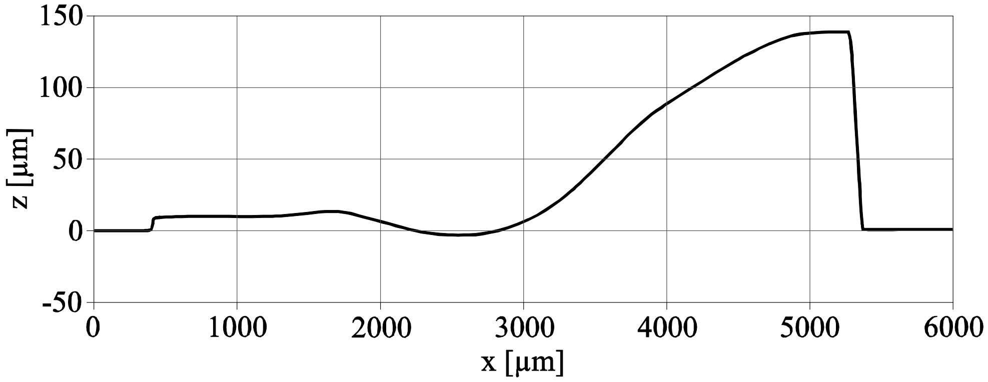

Figure 5.5: Tencor profiler scan of a thin silicon chip. The indentation in the middle was caused by the vacuum fixation. On the left side the chip was pulled flat onto the wafer chuck and a silicon thickness of about 10 µm was determined from the scan. On the right side, however, the chip was bend upwards giving a false chip thickness reading of 140 µm.

5.2.1.3 *Chip thickness measurement using an optical microscope*

Optical measurements offered the advantage of being non-contact techniques, not inflicting any damage on the sample. The chip was set upright onto the microscope table and, using the ocular scale inside the eyepiece, the silicon thickness was measured. Using a stepper motor actuated xy table with digital position readout could improve the accuracy of the measurement to about ± 0.5 µm. To prevent the chip from falling over, a mechanical support had to used. This measurement technique turned out to be the most reliable and was almost always used to measure silicon thicknesses in thin chips in this thesis. Because a clear view of the chip edge was required to carry out optical thickness measurements, the most reproducible results were obtained, if the round edge of a deformed chip was observed. In order to account for thickness variations across a chip and to reduce measurement errors, the thickness was measured on the left, middle, and right of the two round edges of each deformed chip and then averaged per chip edge using

$$t_{Si} = \frac{t_{left} + 2 \cdot t_{middle} + t_{right}}{4}, \tag{5.1}$$

where t_{Si} is the resulting average silicon thickness and t_{left}, t_{middle}, and t_{right} are the measured silicon thicknesses on the left, middle, and right of one edge, respectively. The chip thickness was then calculated as the arithmetic average of the thicknesses of two chip edges.

5.2.1.4 *Chip thickness calculated from chip weight*

Weight measurements are an interesting alternative to the above described optical chip thickness measurements. High precision scales by Mettler (type HK 60) with an accuracy of 10 µg were used to measure the chip weight. If the geometries and densities of the metal structures on the chips were known, their weight m_{metal} could be easily calculated and subtracted from the measured mass m_{meas}. Taking into account the chip area A_{chip} and the density of silicon ρ_{Si} = 2.3 g/cm³ the silicon thickness was calculated using

$$\mathrm{t_{Si}} = \frac{\mathrm{m_{meas}} - \mathrm{m_{metal}}}{\mathrm{A_{chip}} \cdot \rho_{\mathrm{Si}}}. \tag{5.2}$$

The density of electroplated layers depended on the plating conditions. Because this density data was not available for the electroplated metals, the bulk metal values found in the literature were used. This lead to an error, which was, however, assumed to be small, because the electroplated layers were found to be compact (and not porous) in optical inspections. Also, the error should have been very similar for all chips, because the plating conditions were not varied between samples. Using the weight method to determine the silicon thickness had the disadvantage that no data on the thickness distribution across the chip was collected. On the other hand, this method returned an average thickness of the chip, also taking into account the silicon thickness away from the chip edges. It will later be shown that the thickness distribution played an important role in determining deformations. Therefore, the weight method was preferably used for chips with a homogeneous silicon thickness, while the optical measurement technique was preferred in chips with a notable thickness distribution.

Based on the resolution of the scales, the theoretical accuracy Δt_{Si} for the weight method applied to square chips with an edge length of 5 mm or 10 mm was found to be ± 0.17 µm and ± 0.043 µm, respectively. These values were obtained from:

$$\Delta t_{Si} = \frac{\Delta m_{scales}}{A_{chip} \cdot \rho_{Si}}, \tag{5.3}$$

where Δm_{scales} is the resolution of the scales.

In a reproducibility study for the weight method involving 28 chips with 10 µm edge length and a total calculated metal weight of 5.38 mg/chip, the accuracy of the scales was found to be about ten times less exact than expected. The correlation of weight measurements done on two different days is shown in Figure 5.6 a). Transforming the measured weight difference into a silicon thickness difference yielded that 22 of the 28 measurements showed a deviation between 2 measurements of the same chip of less than 0.5 µm and the maximum deviation was 2.1 µm. The correlation between optical and weight thickness measurements is shown in Figure 5.6 b) for 12 of the 28 chips from above. The deviation was generally between 1 µm and 2 µm. For individual thicker chips, however, it reached almost 5 µm.

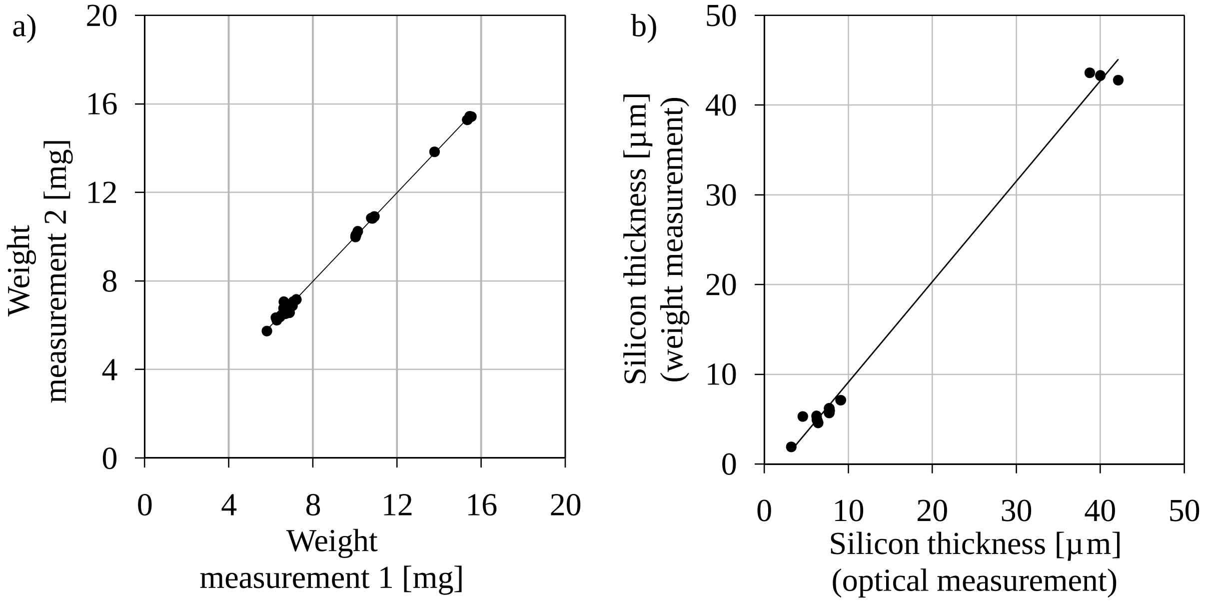

Figure 5.6: a) Reproducibility of weight measurement. A regression line was added; b) correlation between optical silicon thickness measurement as described in Section 5.2.1.3 and silicon thickness determined from weight measurement results. As can be seen by the regression line the weight method yielded a silicon thickness slightly smaller than the optical measurement for chips with a thickness below 10 µm, while for chips with a thickness around 40 µm the weight method yielded larger than expected silicon thicknesses.

5.2.2 Measurement of deformations

5.2.2.1 Chip bow measurement at room temperature

Two measurement techniques are discussed in this section:

- using a surface profiler and
- using optical focus depth measurements.

Measuring the chip bow with the surface profiler offered the advantage of an automated measurement with continuous bow readout over the whole chip area. However, as in the thickness measurement, the lightweight chips were moved by the scanning needle of the profiler or, if a vacuum fixation was used, the chips were deformed by the fixation, falsifying the measurement.

The optical focus depth measurement technique was much better suited to determine the chip's curvature. The chips were lying on a stepper motor actuated microscope table with digital position readout. The xy table was sequentially moved over three defined points lying on a circular slice of the cylindrically deformed chip. At each point, the microscope focus was adjusted to the chip's surface and the z-position was determined with an electronic Heidenhain linear encoder (type ND 221) attached to the microscope table. Using the trilateration formula derived in Appendix B, the radius of curvature was determined. Although this method was very time-consuming and yielded the radius of curvature in only a single plane, it was the preferred method for bow measurements, because of its excellent reproducibility. Individual z-coordinates could usually be measured with a reproducibility ≤ 2 µm, typically resulting in a deviation of the radius of curvature between two measurements of less than 1%. In another trial, the bow was first measured in the original orientation and then after the chip was flipped over. No significant difference between these measurements was found, indicating that the chip weight had a negligible influence on the chip bow. Alternatively, the three points for the trilateration could be obtained by placing the chip upright onto the microscope table and using only the xy table readout to obtain the point coordinates. The disadvantage of this procedure was that only points on the chip edge could be measured. Also, the chips required a fixation to prevent them from falling over, which modified the bow.

Apart from the bending radius of a chip, the bending angle was a good indicator for the degree of curvature. Because for most causes of chip deformations, the silicon thickness had a more linear relation to the bending angle than to the bending radius, the bending angle was the preferred method for comparing bows in chips with different silicon thicknesses. The bending angle α can be easily calculated from the bending radius R and the edge length b_{Si} of the chip using

$$\alpha = \frac{b_{Si}}{R} \quad [\mathrm{rad}] \tag{5.4}$$

$$= \frac{360°}{2 \cdot \pi} \cdot \frac{b_{Si}}{R} \quad [\text{deg}] .$$

5.2.2.2 *In-process chip bow measurement*

The reflow was generally carried out in a nitrogen/formic acid atmosphere inside a reflow oven. However, using the heating plate of a flip chip bonder equipped with a digital camera, the chip deformations could be observed during the reflow process. In order to take sideways images of the chips during reflow, the forming gas module chamber was removed and the reflow was carried out in ambient air. During the reflow process, sideways images were taken and later the bow was determined by analyzing the images. In Figure 5.7 an example of a thermally bent chip with gold/tin lines is shown.

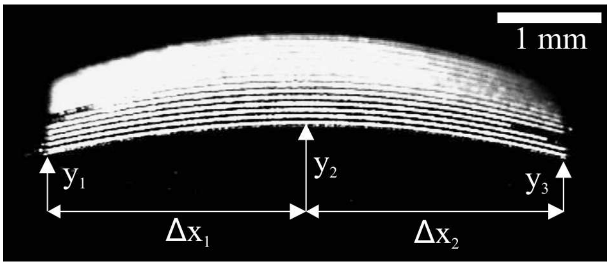

Figure 5.7: Optical side view image of thermally bent chip taken with a digital camera during reflow at 210°C. The measured values y_1*,* y_2*,* y_3*,* Δx_1*, and* Δx_2 *are used to calculate the radius of curvature and the bending angle with a trilateration formula. The bending angle of the chip in the image is defined as positive.*

Because a digital camera with a large focus distance had to be used to keep a safe distance from the heating plate, the images had a low resolution of approximately 5 µm. Compared to the total deformations of typically several hundreds of microns, this error was negligible. In Figure 5.8 an example of in-process bending angle measurements on 3 chips with silicon thicknesses of 6 µm, 12 µm, and 18 µm is shown. Forthe sake of argument, the bending angle was defined such that it increased with temperature (for an example of a chip with a positive bending angle see Figure 5.7). The dependence of the deformation on temperature and silicon thickness is clearly visible. For a detailed description of the different parts of the diagram refer to Section 5.4.2.

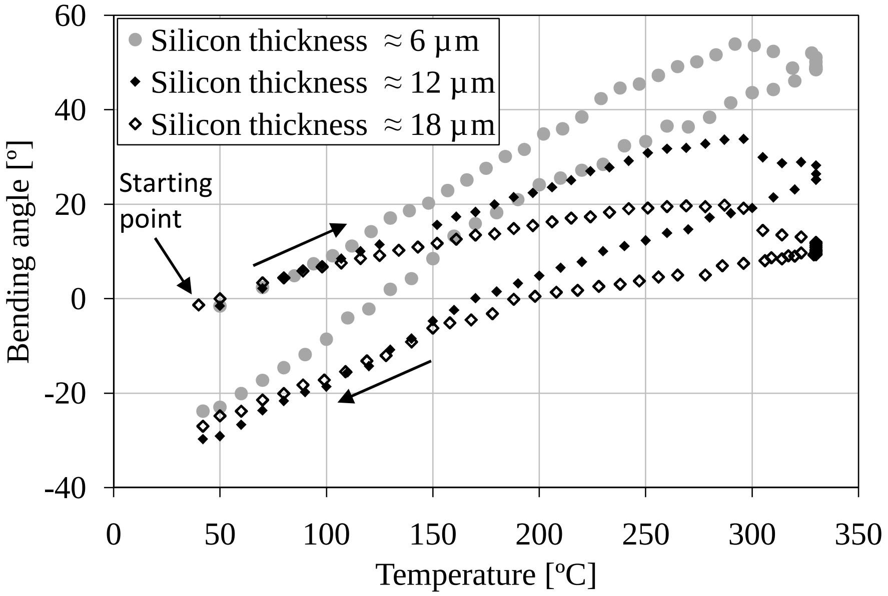

Figure 5.8: Optical in-process bow measurements on 3 silicon chips with gold/tin lines as a function of silicon thickness and temperature during reflow. The arrows indicate the sequence of measurements during the reflow cycle.

The measurements were well suited to pick up small differences between chips. In Figure 5.9 two chips with only a slight thickness difference of 3 µm are compared. The thinner chip is slightly more bent at the peak temperature of 330°C and after reflow at 42°C,

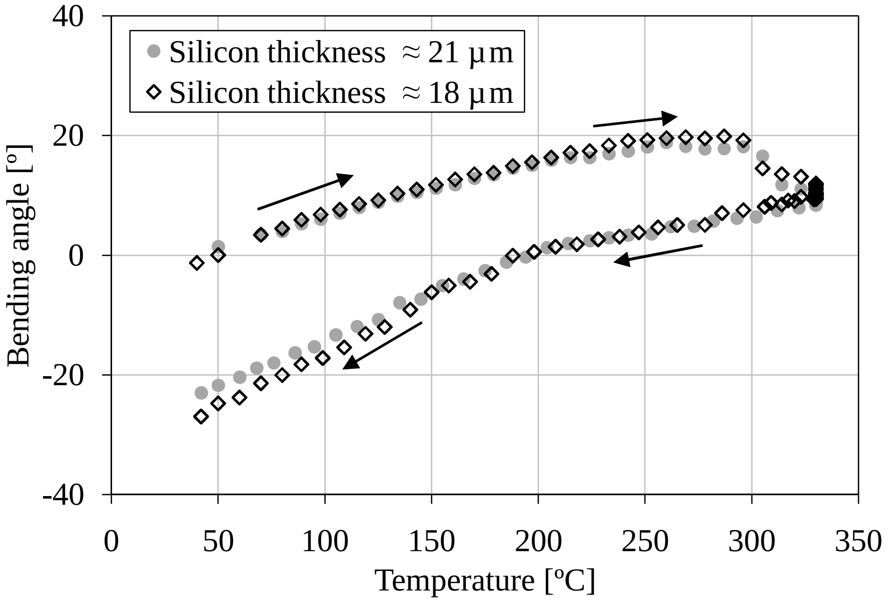

Figure 5.9: Optical in-process bow measurements on 2 silicon chips with gold/tin lines as a function of temperature during reflow. The thicknesses of the 2 chips were very similar, resulting in only slightly different deformation curves. As expected, the thinner chip had a larger concave deformation at high temperatures and also larger convex deformation after ramp-down.

The accuracy of bow measurements based on sideways images is a point of interest. In Figure 5.10 results for bow measurements of 6 bent chips using an optical microscope and images taken with a digital camera are compared. The chips had different silicon thicknesses, resulting in different bending radii.

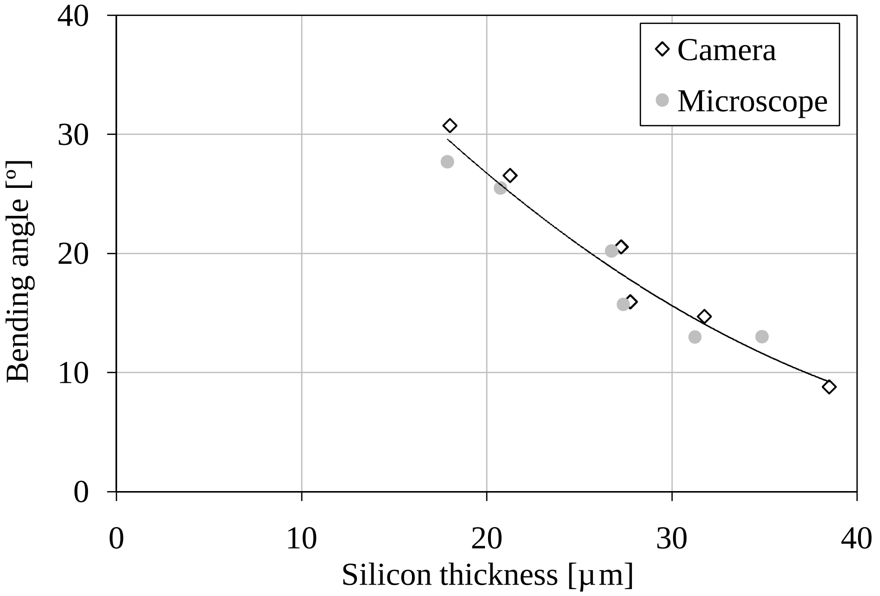

Figure 5.10: Bending angle of 6 thin silicon chips with long Au/Sn lines after reflow. Two different methods were used: direct measurement using an optical microscope and indirect measurement by analyzing images taken with a digital camera. A common regression curve was added.

The microscope measurements were carried out along a line through the chip center, while the camera measurements were performed on the chip edge. Because of non-uniform silicon thickness distributions within the chips, the silicon thickness at the measurement location had to be taken into account in Figure 5.10. Although there is a slight tendency for bending angle values obtained from camera measurements to be larger than those obtained with the microscope, the distribution of the data points around the regression curve is fairly tight. Possible causes for the observed deviations from the regression curve were:

- temperature differences at the time of measurement: this effect, however, had only a small influence, leading to a bending angle differences between 0.2°/°C (for the thinnest chip) and 0.1°/°C (for the thickest chip),
- viewing angle of the camera: the camera's angle relative to the heating plate was about 9°. This lead to a small error d_{error} in the measured distances d in the images of

$$d_{error} = \cos(9°) - 1 \approx -1.25\% \,, \tag{5.5}$$

- silicon thickness measurement: the silicon thickness in the chip middle could not be directly measured, but was calculated by the average of 6 measurements on the chip edge,

- mechanical coupling between edges in a chip with an inhomogeneous thickness distribution: the thinner measured edge was influenced by the bow of the opposite thicker edge, generally leading to reduced bow readings on the thinner side and increased bow readings on the thicker side.

Because the measured deformations are strongly dependent on the silicon thickness on the edge, where the measurement was carried out, bow measurement data was preferably related to the actual silicon thickness at the measurement site. In Figure 5.11 a silicon chip before, during, and after reflow is shown. The chip had a silicon thickness of 12 µm in the front and 2 µm in the back, leading to larger deformations in the back at 330°C. During cool-down, the concave deformation in the front was inverted into a convex deformation, while the concave deformation in the back was not completely reversed.

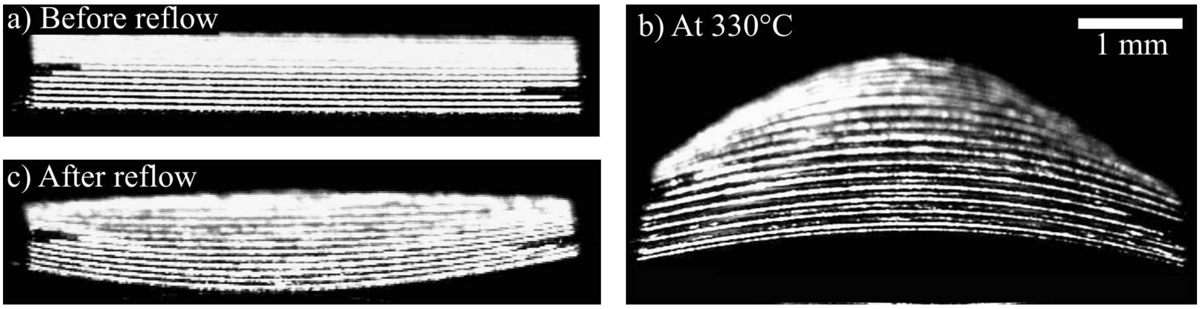

Figure 5.11: Silicon chip with Au/Sn solder lines (running from left to right) and Au-Ni UBMs. The front part of the chip had a silicon thickness of 12 µm and the back part of only 2 µm; a) before reflow (at 40°C) → no deformation visible; b) at peak temperature during reflow (at 330°C) → medium deformation in the front, strong deformation in the back; c) after reflow (at 42°C) → convex deformation in the front, slight concave deformation in the back.

5.2.2.3 Local deformations in chips with gold/tin lines

Apart from the macroscopically visible bow, the chips with gold/tin lines also exhibited local deformations between the silicon under the gold/tin lines and the areas between the lines. In Figure 5.12 both the chip bow and local silicon deformations are illustrated. Because local deformations are usually in the sub-micron range, they could not be reliably measured using measurement principles based on optical microscopes. Surface profiler scans were much more accurate and could easily pick up deformations around 10 to 50 nm.

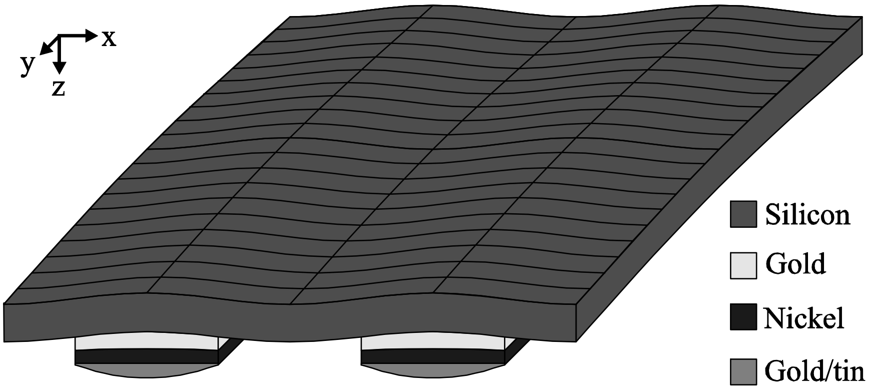

Figure 5.12: Schematic of reflowed chip with gold/tin lines. The solder lines are facing down and are oriented in the y-direction. The macroscopically visible, stress induced chip bow was around the x-axis. Because this schematic is not to scale, the local deformations are also visible as a wavy surface profile along the x-axis. The amplitude of this wavy pattern was defined as the local deformation D for chips with gold/tin lines.

However, as discussed earlier, this required the use of a vacuum chip fixation, which forced the flexible chip to assume the contour of the underlying chuck, modifying its global and local shape. Also, the vacuum lead to substantial deformations of the chip at and around the vacuum holes, which also modified the profiler scans and often destroyed the chip. In Figure 5.13 an image of a three-dimensional profiler scan of an ultrathin silicon chip is shown that visualizes these problems. To minimize the effect of the vacuum holes and the uneven chuck surface, a polished aluminum chuck was fabricated with a vacuum hole diameter of only 0.5 mm. While the influence of the vacuum hole could be reduced this way, the quantitative influence of the vacuum on local deformations away from the hole was not known. A better alternative to measure local deformations, was the reduction of the vacuum to almost ambient pressure and using the Bernoulli effect of streaming air to very slightly pull down the chip. Because most chips had a cylindrical or conical shape after reflow, this technique worked well and was extensively used to measure local deformations. The vacuum reduction was achieved by opening bypass holes in the chuck. The number of bypass holes was adjusted, so that the chip was just tight enough not to be moved by the profiler needle. This way the adaptation of the chip shape to the chuck's surface was avoided. However, global vacuum induced deformations could still occur. Therefore, the optical focus depth measurement principle was still the better alternative to measure global deformations, while the profiler setup with very slight vacuum was the best method to determine local deformations.

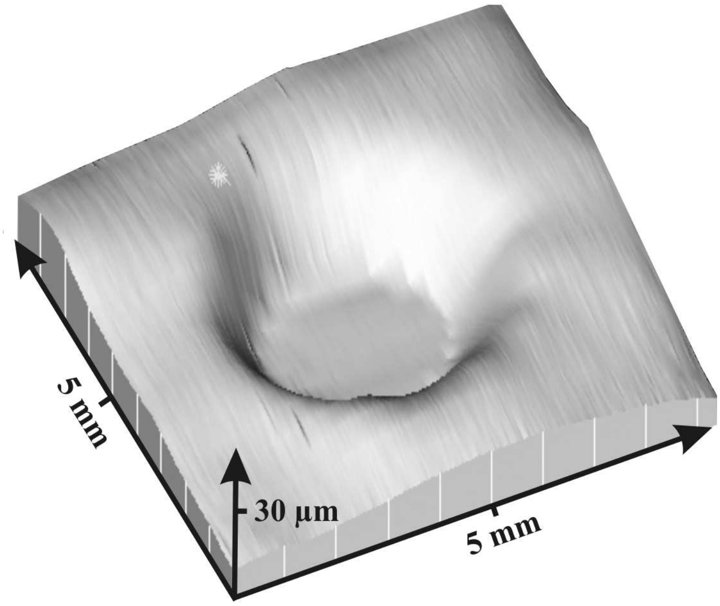

Figure 5.13: 3D profiler scan of the silicon surface of a chip with an approximate thickness of 20 µm. The vacuum hole of the underlying chuck lead to silicon deformations. The chuck's uneven surface that was impressed onto the chip is also visible.

Before the simulation results and measurements of the local deformation could be compared, disturbances in the profiler measurement data had to be eliminated or minimized. These disturbances generally were global chip deformations, non-leveled profiler output data, and the silicon surface roughness. Because the global deformations in chips with long gold/tin lines were generally cylindrical, with the gold/tin lines bent around the imaginary cylinder's axis, a single profiler scan perpendicular to the gold/tin lines should have yielded only the wavy profile illustrated in Figure 5.12. In practice, however, the data had to be leveled and the silicon surface roughness needed to be eliminated by superimposing many individual measurements.

In Figure 5.14 a) an example of original profiler data points from a single surface scan on the backside of a reflowed chip is shown. The 3000 µm scan was oriented perpendicular to the gold/tin lines, which had a width and spacing of 200 µm (i.e. a 400 µm pitch). Hence, the scan resulted in the measurement of the silicon deformation of 7 and a half parallel gold/tin lines. Due to data export limitations of the profiler system the saved data points had a spacing of 20 µm. Although the 400 µm pitch of the silicon deformations was well visible in the data, quantitative results required further data analysis. Global deformations were reduced by calculating a regression line and subtracting this line from the profiler data. The resulting data points are shown in Figure 5.14 b). The local deformations were more clearly visible now. However, the signal was still jittery due to the roughness of the silicon surface. Hence, a much better quantitative result could be obtained if the 400 µm periodicity of the signal was taken into account and averages were calculated from the data based on the superposition of seven and a half 400 µm segments. The resulting graph is shown in Figure 5.14 c). Here the local deformation D for the single scan could be extracted by simply subtracting the minimum z-value from the maximum z-value.

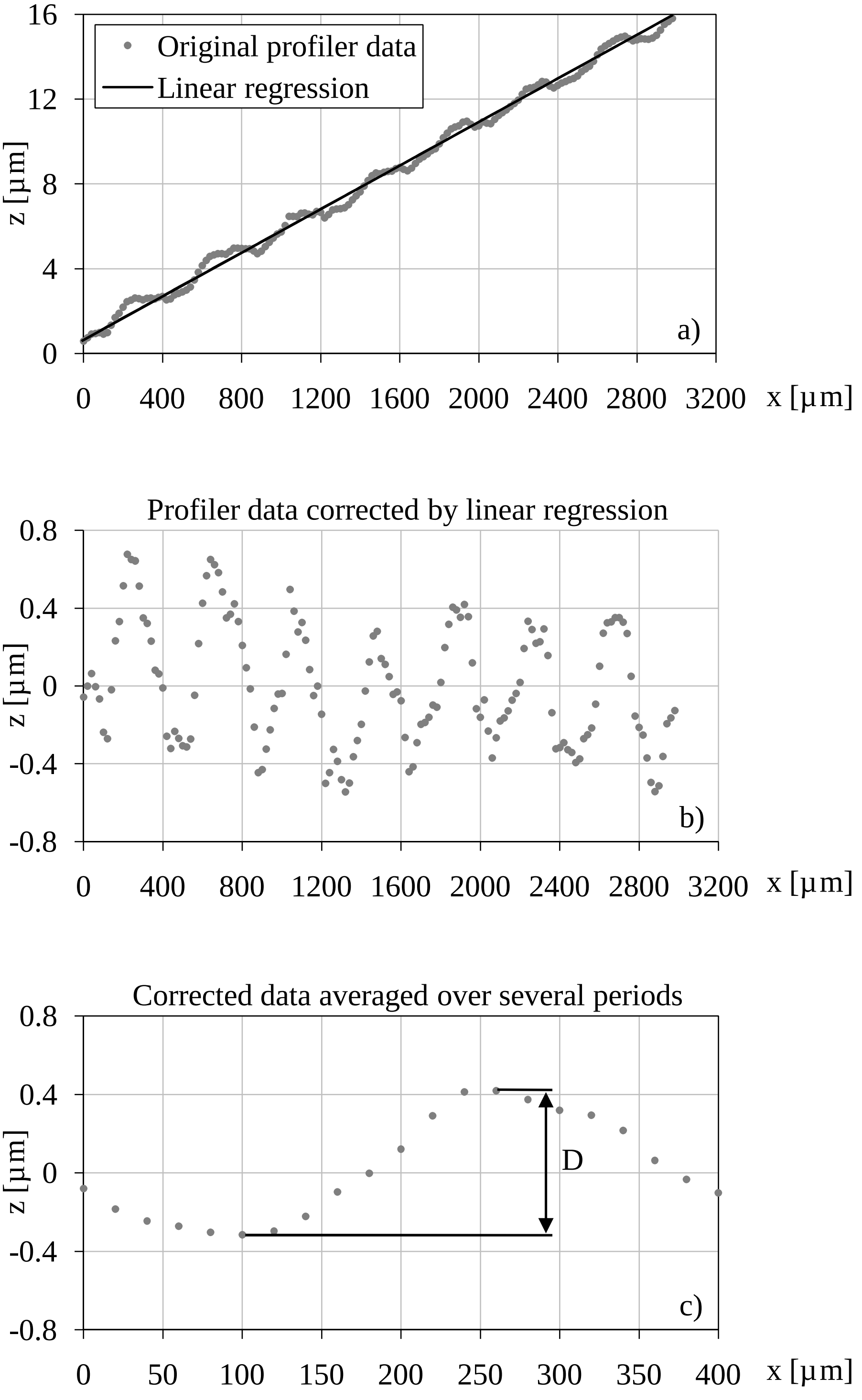

Figure 5.14: Extraction of local deformation data from profiler measurements; a) original profiler data and regression line; b) original profiler data corrected by regression line; c) 400 µm periodicity of signal used to superimpose the z-values of all 400 µm segments and calculate D from this data.

To improve data quality, 150 scans with a sideways pitch of 20 µm were taken on each chip and individually analyzed in the described way. Therefore, the scanned area comprised a square with 3 mm edge length with a total of 22,500 data points per chip. The 150 individual results for the local silicon deformation D for each line scan were then averaged per chip.

5.2.2.4 Local deformations in chips with round gold/tin bumps

Local silicon deformations under and around circular reflowed gold/tin bumps were even more difficult to measure, because the chips assumed a complex three-dimensional shape, which needed to be compensated to extract local deformation data. Like in chips with gold/tin lines, profiler scans were the best measurement technique to detect these local deformations. In Figure 5.15 step-by-step data extraction of local silicon deformations from a thin chip with a thickness of about 20 µm is shown.

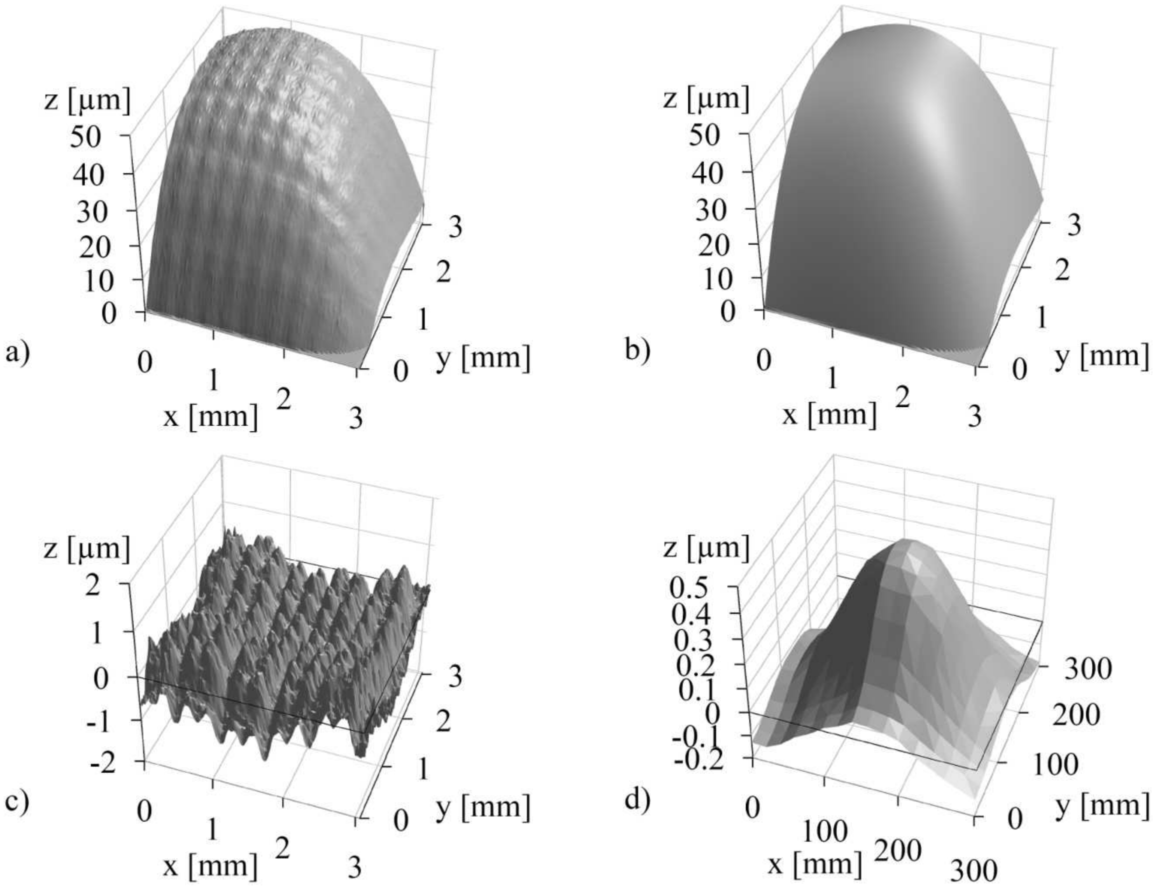

Figure 5.15: 3 x 3 mm² profiler scan of the backside of a reflowed thin chip with gold/tin bumps on Au-Ni UBMs and gold tracks. The local silicon deformation data was extracted from this scan step by step; a) original profiler data; b) three-dimensional approximation function; c) difference between original profiler data and approximation function; d) average of 81 local deformations.

5.3 Deformations based on stress in as-plated layers

The gold and nickel electroplating required temperatures of 50°C and 40°C degrees, respectively, leading to tensile stress during cooling to room temperature after deposition. Apart from this thermal stress, the plated layers also exhibited intrinsic stress, resulting from the crystal growth process during plating. After wafer thinning and detachment from the carrier substrate, the total stress (that is the combined thermal and intrinsic stress) lead to a small convex deformation. The intrinsic stress and hence the deformation depended on many parameters such as bath temperature, current density, sample and feature size, convection, sample agitation, electrolyte composition, and surface roughness [163]. Apart from this, the room temperature electroplating of tin also influenced deformations. Because the tin layer exhibited very little total stress, it merely stiffened the bumps, hence reducing deformations after chip detachment from the carrier substrate.

5.3.1 Determination of plating stress from wafer bow

In an experiment based on three thinned 4” silicon wafers, each plated with either a nickel, gold, or tin layer of 10 µm thickness, deformations were found, which could be explained by the presence of intrinsic plating stress and stress in the thinned silicon back surface. The wafers were first lapped, then CMP polished, and finally etched with HNA etching solution to relieve stress in the silicon surface. During the HNA wet etching of the silicon, the plated metal layers were protected using standard dicing tape. The wafer bow of each wafer was measured before and after the HNA etching step. Before HNA etching all wafers showed cylinder-like deformations with bending around the [100] crystal orientation of the silicon. The bending orientation indicated tensile stress in the plated layers. After HNA etching the bow decreased in all wafers. In the wafer plated with tin the bow was even decreased to negative values indicating compressive stress in the plated tin layer.

The bending radius measurements of all 3 wafers before and after HNA etching, as well as calculated film stress values, are listed in Table 5.1. The required material properties for calculating the film stress are given in Table 5.2.

Table 5.1: Bending radius R and calculated metal film stress before and after HNA etching. The resulting stress reduction $\Delta\sigma_t$ is also given.

Plated layer	R [mm]		Calculated metal film stress σ_t [MPa]		$\Delta\sigma_t$ [MPa]
	before etching	after etching	before etching	after etching	
Au	306	354	96.8	58.5	-38.3
Ni	386	490	95.4	54.7	-40.7
Sn	939	-3366	43.3	-10.1	-53.4

Table 5.2: Material and process parameters relevant to the wafer bow of 10 µm thick metal layers plated onto silicon wafers, which were thinned to approximately 100 µm after metal deposition.

Layer	ΔT [°C]	α_m (at RT) [ppm/°C]	Young's modulus E_m [GPa]	Poisson's ratio ν_m
Au	- 30	14.36	78	0.42
Ni	- 20	13.10	195	0.312
Sn	0	23.5	49.9	0.357

Because the metal layer thickness was very small compared to the silicon substrate thickness, the expected thermal stress could be calculated using

$$\sigma'_{\Delta T} = \Delta T \cdot \Delta\alpha \cdot E_m \, , \tag{5.6}$$

where ΔT is the difference between plating bath temperature and room temperature, $\Delta\alpha$ is the difference of the coefficients of thermal expansion of the metal layer and silicon, and E_m is the Young's modulus of the metal layer. Because the resulting wafer bow was generally cylindrically shaped, the elongation of the metal layers along the y-axis (i.e. the cylinder's axis) was close to zero. Taking into account the biaxial stress state, the stress was increased by a factor based on the Poisson's ratio ν_m of the metal layer [182]:

$$\sigma_{\Delta T} = \frac{\sigma'_{\Delta T}}{(1 - \nu_m)} = \frac{\Delta T \cdot \Delta\alpha \cdot E_m}{(1 - \nu_m)} \, . \tag{5.7}$$

The total stress (that is the combined thermal and intrinsic stress) could be calculated from the bending radius R, the layer thicknesses, Young's moduli, and Poisson's ratios of silicon (t_{Si}, E_{Si}, and ν_{Si}, respectively) and the plated metal (t_m, E_m, and ν_m, respectively) using Stoney's formula [145]:

$$\sigma_R = \frac{E_{Si}}{6 \cdot (1 - \nu_{Si}) \cdot R \cdot t_m^2 \cdot \left(1 + \frac{t_{Si}}{t_m}\right)} \cdot \left(t_{Si}^3 + \frac{1 - \nu_{Si}}{1 - \nu_m} \cdot \frac{E_m \cdot t_m^3}{E_{Si}}\right) . \tag{5.8}$$

Because the substrate thickness t_{Si} was much larger than the plated metal thickness t_m (and because the respective Young's moduli and Poisson's ratios were similar) formula (5.8) simplifies into

$$\sigma_R = \frac{E_{Si} \cdot t_{Si}^2}{6 \cdot (1 - \nu_{Si}) \cdot R \cdot t_m} \, . \tag{5.9}$$

Because silicon is an anisotropic material, the bending stiffness depended on the direction of the bending axis. The wafers used in this study were (100) oriented and therefore bending around the <100> or <110> axis was expected. These axes were oriented at an angle of 45° to one another. The relative stiffnesses around these axes were calculated in an FEM simulation and the stiffness for bending around the <110> axis was found to be approximately 22.6% larger than for bending around the <100> axis. These simulation results are in good agreement with the Young's moduli of silicon in the <100> and <110> directions of 130.2 GPa and 168.9 GPa, respectively. Therefore, bending of the wafers around the <100> axis was expected to dominate. Indeed all wafers exhibited the expected bending direction. Accordingly, the Young's modulus and Poisson's ratio for the <100> direction had to be used in (5.9). Because the wafer bow was relatively small, different samples were selected to illustrate the dependence of deformations on crystal orientations in Figure 5.16.

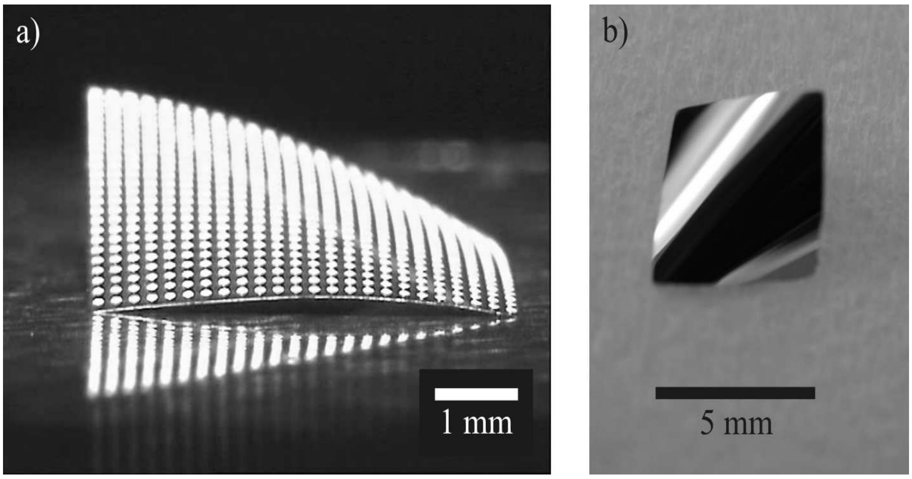

Figure 5.16: Two examples of thin silicon samples bent around the <100> axes because of stress of different origins. In both samples the chip edges were <110> oriented; a) thermal stress in a chip with a homogeneous matrix of bumps and a silicon thickness of about 15 µm at 200°C; b) thermal stress between Si and SiO_2 at room temperature caused by the cool-down after the SiO_2 growth at 1200°C. The chip had a thickness of only 5 µm.

The coefficients of thermal expansion in Table 5.2 were taken from the literature for the relevant temperature ranges (α_{Au} and α_{Ni} were taken from [166] and α_{Sn} was taken from [183]). For gold and nickel the Young's modulus and Poisson's ratio were the same as in Section 4.1, while for tin the values were taken from [167]. From the data in Table 5.2 and $E_{Si<100>} = 130.2$ GPa [171], $\nu_{Si<100>} = 0.064$ [171], and $\alpha_{Si} = 2.45$ ppm/°C [176] the thermal stress, total stress, and intrinsic stress were calculated in Table 5.3. It should be noted that using Stoney's formula for stress calculations in this experiment was not entirely correct, because the observed large wafer bow caused strain (and consequently stress) reductions in the stressed layers. However, the calculated stress values are useful as a first approximation.

Table 5.3: Thermal plating stress $\sigma_{\Delta T}$ calculated from the temperature drop after removal of the wafers from the plating bath, total stress σ_t calculated from the bending radius and layer thicknesses, and calculated intrinsic stress σ_i resulting from the crystal growth during plating.

Plated layer	$\sigma_{\Delta T}$ calculated from ΔT [MPa]	σ_t calculated from R [MPa]	Intrinsic stress $\sigma_i = \sigma_t - \sigma_{\Delta T}$ [MPa]
Au	27.87	58.51	30.64
Ni	41.54	54.71	13.18
Sn	0	-10.08	-10.08

5.4 Reflow of chips with gold/tin lines

5.4.1 Simulation and measurement of chip bow in as-plated samples

The relative amount of plating stress with respect to total stress after reflow gave an idea of the impact of plating stress on the final deformation. In Table 5.4 this ratio was determined for 3 chips with different silicon thicknesses. A ratio of 7% to 10% was found, that could substantially influence chip bows after reflow and therefore needed to be analyzed in more detail.

Table 5.4: Silicon thickness, chip bow before and after reflow for 3 chips with gold/tin lines on Au-Ni UBMs. From these values the plating stress before reflow was calculated as a percentage of the stress after reflow.

t_{Si} [µm]	Bending radius [mm]		Initial stress rel. to stress after reflow [%]
	Before reflow	After reflow	
17.9	104.6	10.24	9.9
20.8	118.2	11.24	9.5
27.8	198.3	14.17	7.1

Using the plating stress σ_t for gold, nickel, and tin calculated in Section 5.3.1, the expected deformations for thin chips with gold/tin lines were simulated using the FEM software ANSYS and then compared to bow measurements of 12 thin chip samples after electroplating of Au-Ni-Au-Sn lines. The results are shown in Figure 5.17. In order to correlate different analyses for the same group of chips, the samples in this and the next section are referred to by their respective sample group labels (e.g. A1, A2,…, A7 for the 7 different groups of chips with Design A).

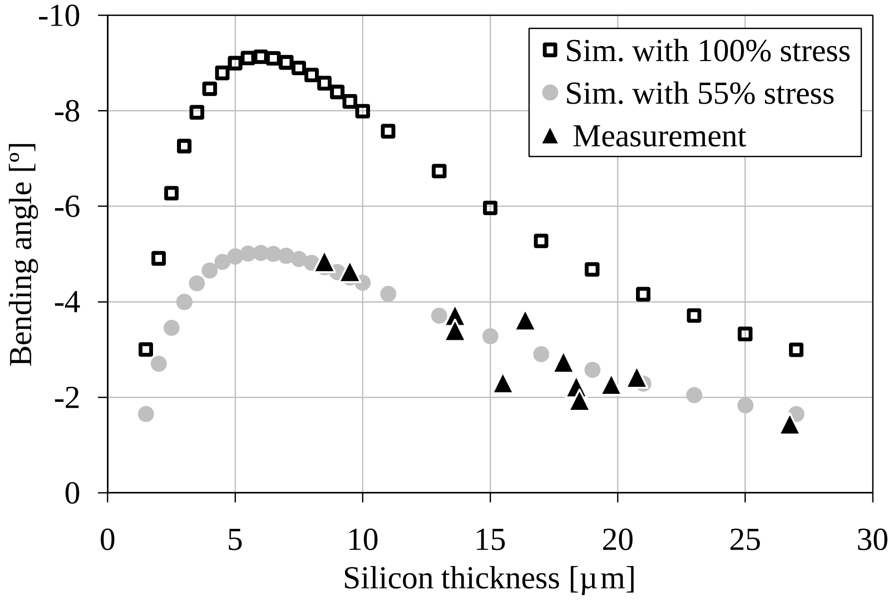

Figure 5.17: Simulated and measured bending angle as a function of silicon thickness for ultrathin as-plated chips with gold/tin lines on an Au-Ni metallization (sample group A1). The simulation was carried out with the calculated film stress values from the wafer bow experiment (squares in the diagram) and with film stress values reduced by 45% (gray circles in the diagram).

The simulation based on the film stress values from the wafer bow experiment yielded deformations that qualitatively matched the measured chip bows. However, the absolute bow values were about twice as large in the simulations. Obviously, the different plating conditions (whole wafer plating versus microstructure plating) lead to much lower film stress in the flex chip samples. By reducing all film stresses in as-plated metals by 45% in the simulation, a very good fit between simulation and measurement was achieved. As expected, the deformations increased with smaller silicon thickness for the prepared samples. For extremely thin silicon ($t_{Si} < 5$ µm) the simulations predicted a reduction of deformations. An explanation for this effect is that in very thin chips the plated metal layers were very stiff compared to the silicon substrate and hence the plating stress merely lead to a contraction and not to a bow of the whole system (also see the thought experiment on page 50).

The simulations in this section were based on FEM models with a gold/tin line width of 100 µm. The line width only had a minor influence on the chip bow, as long as the combined footprint of all lines per chip was the same. In Figure 5.18 simulations of the bending angle in as-plated samples due to plating stress are shown for a line width of 25 µm and 200 µm. Results for all other line widths were in between these results.

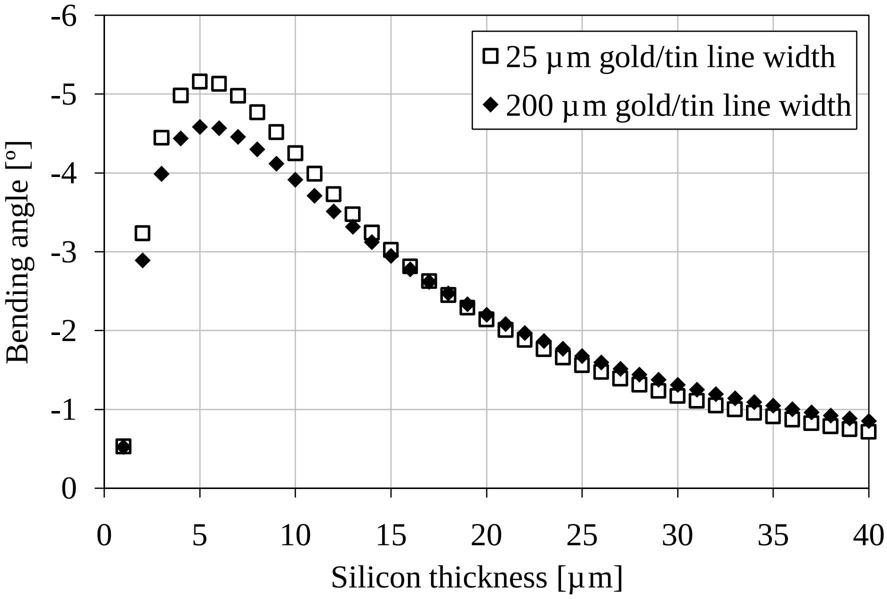

Figure 5.18: Simulation results for plating stress induced bending angle of thin chips with Au-Ni-Au-Sn lines. The bending angle is given as a function of silicon thickness for line widths of 25 µm and 200 µm.

For large silicon thicknesses (t_{Si} > 17 µm) the deformation was slightly larger for 200 µm lines than for 25 µm lines. This was expected and could be explained by transverse contraction effects. The plating stress was assumed to be isotropic and therefore comprised a sideways component (i.e. a stress component in the wafer plane, perpendicular to the Au-Sn-lines). This lateral stress lead to the narrowing of the lines. Because of their large aspect ratio, which was 8 times larger than in 200 µm lines, the 25 µm lines could narrow more easily. Accordingly, the wide lines showed less sideways contraction, but rather transferred the sideways plating stress component via transverse contraction into the main direction of the gold/tin lines. For smaller silicon thicknesses (t_{Si} ≤ 17 µm) this effect was inverted. A possible explanation was that thinner silicon showed larger local deformations leading to a stiffening with respect to global deformations (i.e. the macroscopic chip bow). In 200 µm lines the local deformations were larger, possibly leading to increased stiffness and reduced bending angles.

The deformations resulting from the plating stress in chips with different gold/tin line widths could not be compared directly, because all chips had different silicon thicknesses, which had a much larger influence on the chip bow than the line width. If the silicon thickness had been precisely known, its influence on the chip bow could have been calculated and a comparison of the bows of chips with different line widths would have been possible. For lack of sufficiently accurate silicon thickness data a different analysis technique was used, which eliminated the need for silicon thickness data. At room temperature the chips typically had a convex shape. By increasing the temperature slightly (to about 40°C to 70°C) the bow was reversed resulting in nearly flat chips, a state which is called *zero bending* in the following. At this temperature the silicon was nearly stress-free and hence its thickness only had a minor (if any) influence on the bow. The thermal stress (which was compressive in the metal structures) resulting from the slight temperature increase compensated the tensile stress from plating. A larger temperature required for zero bending therefore indicated larger plating stress (in absolute values). In Figure 5.19 the temperature required for zero bending of 22 thin chips is presented as a function of gold/tin line width. Even though there was some fluctuation in the data a slight tendency towards higher required temperatures for larger line widths was found. This observation is in accordance with the above findings, that the plating stress of smaller structures (i.e. gold/tin lines) was found to be 55% smaller than the stress in large plated structures (i.e. whole wafers).

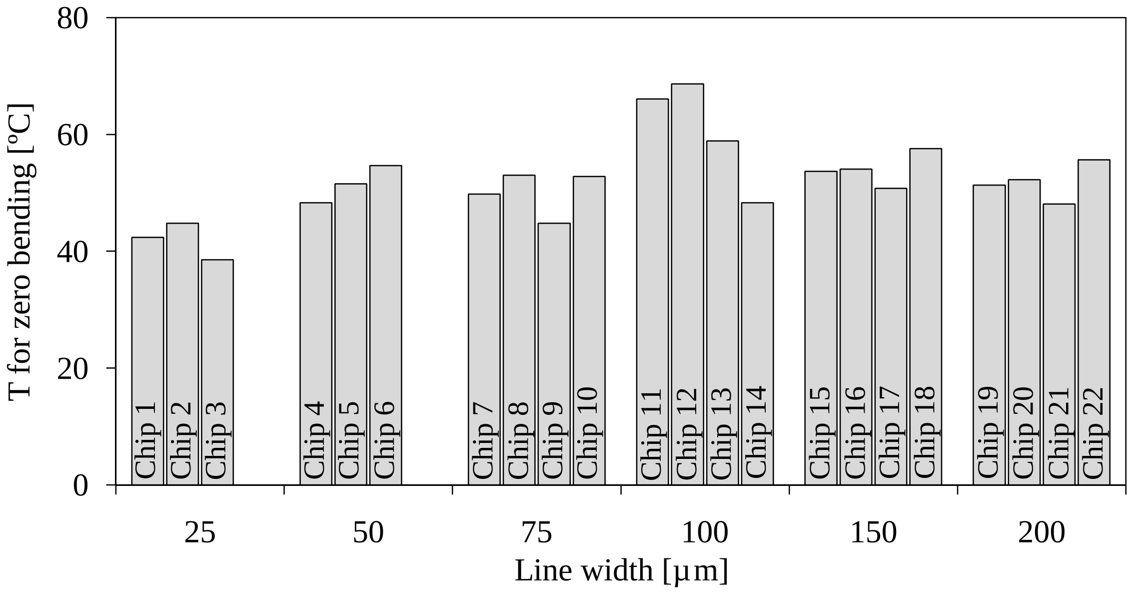

Figure 5.19: Diagram showing the temperature required to compensate the mechanical stress from plating for chips with different gold/tin line widths. The data was obtained by heating 22 as-plated chips (sample group A2) and measuring the bending angle as a function of temperature. Using a linear regression for each chip, the temperature resulting in zero bow was calculated. Because the results were independent of the silicon thickness of each chip, this method allowed the direct comparison of chips with different gold/tin line widths without compensating for silicon thickness. Although there is slight fluctuation in the data, an overall tendency to higher temperatures for wider lines was observed.

5.4.2 Bending angle measurement during reflow in air

As described in Section 5.2.2.2 the bending angle could be measured in-process, if the reflow was carried out in ambient air and sideways images were taken of the bent system. In Figure 5.20 results from bow measurements on a single thin chip during reflow are shown.

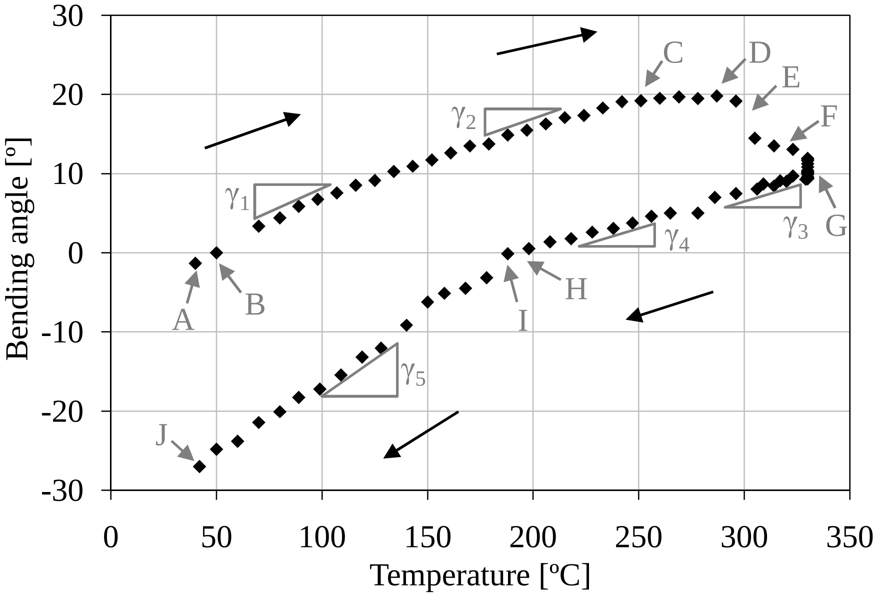

Figure 5.20: Example of in-process bending angle measurements during reflow of a chip with a silicon thickness of 18 µm. Each data point was obtained by analysis of an image taken with the flip chip bonder camera during the reflow process in air. Notable gradients and points are labeled with Greek letters γ_1 to γ_5 and Roman letters A to J, respectively, and are explained in Table 5.5 and Table 5.6.

Table 5.5: Explanation of distinctive results from the in-process bending angle measurement of an 18 µm thick chip with gold/tin lines during reflow in Figure 5.20.

	Temperature [°C]	Bending angle [°]	Explanation
A	40	-1.3	Small convex bending angle from plating stress
B	50	≈ 0	Tensile plating stress compensated by thermal expansion
C	251	19.2	Tin top layer melts, reducing the bending angle gradient
D	287	19.8	Maximum concave deformation
E	296 → 305	19.2 → 14.5	Rapid gold dissolution in molten tin
F	305 → 330	14.5 → 11.9	Slow gold dissolution in molten tin and/or stress relaxation
G	330	11.9 → 10.1	Stress relaxation at constant temperature held for 40 s
H	188	≈ 0	ζ → ζ' phase transition at approximately 190°C
I	188	≈ 0	Thermal tensile solder stress compensated by thermal compressive stress in Au-Ni UBM.
J	42	-27.0	Large convex bending angle after reflow

The different process sections in the diagram are labeled and are explained in detail in Table 5.5 and Table 5.6. The image-based bow measurement technique was suitable to determine macroscopic deformations in-process. The following sections deal with several distinctive parts of the diagram and how they could be simulated using the FEM tool ANSYS 11.0. For better comparison of the analysis with the diagram in Figure 5.20, the labels of the different process sections that best fit the observed temperature range are included in the title of each subsection.

Table 5.6: Temperature range, bending angle gradient, metal layers, and explanation of different gradients in the diagram in Figure 5.20.

	Temperature [°C]	Bending angle gradient [° / 1000°C]	Metal layers	Explanation
γ_1	70 → 107	113	Au-Ni-Au-Sn	Steep gradient at low temperatures
γ_2	179 → 215	93	Au-Ni-Au-Sn	Reduced gradient due to softening of metal layers at elevated temperatures
γ_3	330 → 286	53	Au-Ni	Shallow gradient because molten solder has no impact
γ_4	256 → 218	74	Au-Ni-Au/Sn	Shallow gradient for ζ phase
γ_5	128 → 99	178	Au-Ni-Au/Sn	Steep gradient for ζ' phase

5.4.3 Chip bow between 20°C and 80°C (process section γ_1)

The silicon thickness of 18 as-plated thin chips with lines of Au-Ni-Au-Sn (sample group A3) was measured using optical techniques. Seven of the chips were reflowed using the standard reflow profile and the bending angle was determined by optical means during the reflow process. Eleven more chips were not run through the complete reflow cycle, but analyzed at only three temperatures: 40°C, 70°C, and 100°C. In order to reduce the impact of measurement errors a linear regression was fitted to each set of three bow measurements and the bending angle at 80°C was calculated from this regression. By using a maximum temperature of 100°C, significant metallurgical changes of the tin layer were prevented, allowing the chips to be analyzed a second time, with the bow at the opposite chip edge now being observed. Apart from these measurements, the bending angle was also simulated for the given temperature of 60°C above room temperature. In Figure 5.21 the simulation and measurement results are compared. By only taking the deformation resulting from the temperature increase into account, no sufficient accordance between simulated and measured bending angles was achieved. However, by including the simulated bow caused by the plating stress (based on 55% of the whole wafer plating stress values) the measured and simulated bows showed good accordance.

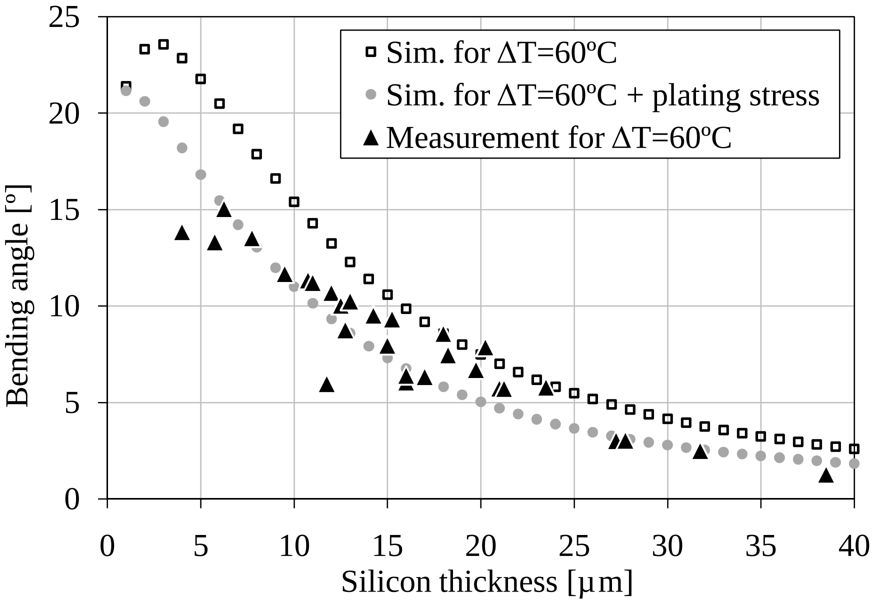

Figure 5.21: Simulated and measured chip bow at 80°C (sample group A3). The original simulation (squares in the diagram) was corrected by the plating stress resulting in a good fit between measurement (triangles in the diagram) and simulation (circles in the diagram).

These results indicated that the plating stress could not be neglected and that the adapted FEM model with all 4 electroplated layers (Au-Ni-Au-Sn) well reproduced actual deformations during the early part of the ramp-up.

5.4.4 Chip bow at 330°C (process section G)

Because the tin and underlying gold layer were transformed into liquid Au/Sn eutectic during reflow, their original plating stress no longer influenced the chip bow at 330°C. The plating stress in the nickel diffusion barrier and gold layer underneath, however, still had an influence. In Figure 5.22 simulation results for the influence of the plating stress in these layers on the bending angle as a function of silicon thickness and gold/tin line width are shown. The qualitative curve shapes were similar to the Au-Ni-Au-Sn simulation results in Figure 5.18. However, the bending angle difference between 25 µm and 200 µm lines at large silicon thicknesses was less pronounced. A possible explanation for this result is, that the Au-Ni layers were about 70% thinner than the Au-Ni-Au-Sn layers, leading to a fairly small aspect ratio for all line widths.

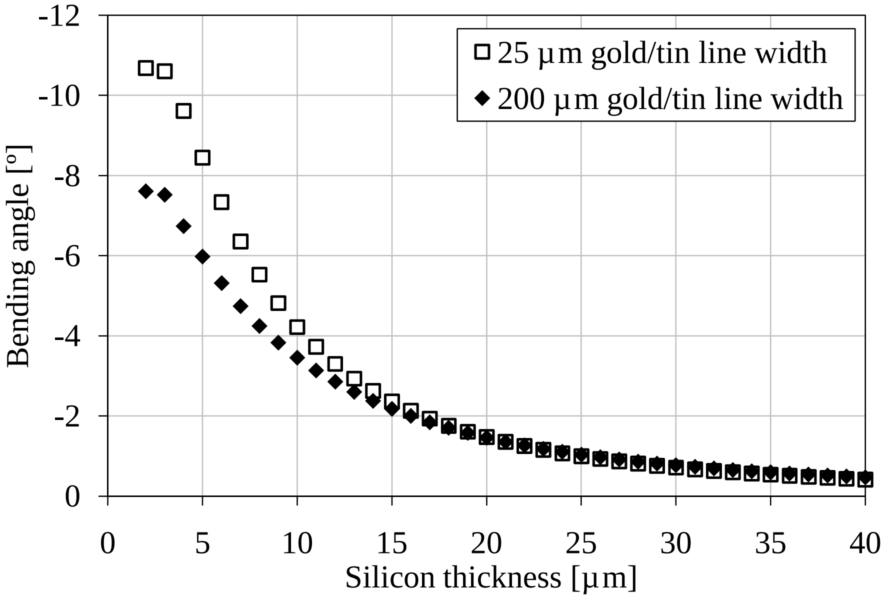

Figure 5.22: Simulation results for plating stress induced bending angle of thin chips with Au-Ni lines. The bending angle is given as a function of silicon thickness for line widths of 25 µm and 200 µm.

Because of excessive processor time required for the numerical solution of models with very thin silicon thicknesses, the simulations were limited to 100 µm lines in the following. In earlier simulations, the results from 100 µm line models were similar to averaged results for line widths between 25 µm and 200 µm. Figure 5.23 compares the bending angle in thin chips due to plating stress in Au-Ni and Au-Ni-Au-Sn lines with a widths of 100 µm as a function of silicon thickness.

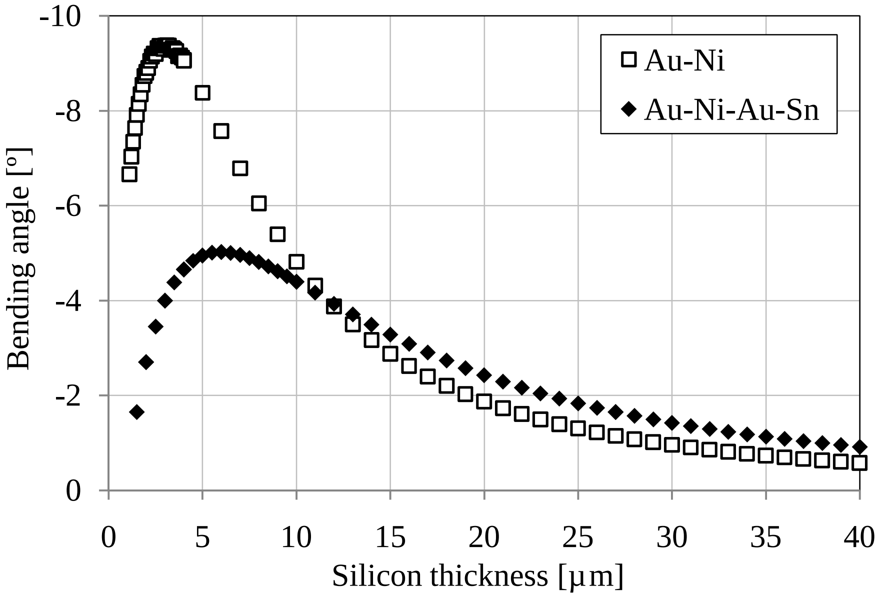

Figure 5.23: Comparison of simulated plating stress induced bending angle in thin chips with Au-Ni and Au-Ni-Au-Sn lines as a function of silicon thickness. To achieve reasonable simulation times at very small silicon thicknesses, only the 100 µm line width was considered.

For large silicon thicknesses ($t_{Si} \geq 12$ µm) the Au-Ni lines showed less bow than the Au-Ni-Au-Sn lines. This was expected and was due to the large tensile stress in the 5 µm gold layer, which dominated the combined stress in the Au-Sn layer. For small silicon thicknesses ($t_{Si} < 12$ µm), however, the bow increased substantially, reducing tensile stress in the outer layers (away from the silicon substrate). For the Au-Ni-Au-Sn simulations this effect was more pronounced, because here the total metal thickness was about three times larger than in the Au-Ni simulations. Therefore, Au-Ni simulations showed larger bow than Au-Ni-Au-Sn simulations at small silicon thicknesses. For very small silicon thicknesses ($t_{Si} < 2.7$ µm for Au-Ni and $t_{Si} < 6$ µm for Au-Ni-Au-Sn) the silicon became less important for deformations and the bow was ultimately determined by the metal layers and their relative stress, thickness, and Young's modulus. Interestingly, the maximum deformation was about twice as large in Au-Ni simulation compared to Au-Ni-Au-Sn simulation.

In Figure 5.24 the measured bending angle at 330°C for 9 chips with gold/tin lines is compared to simulation results with and without plating stress.

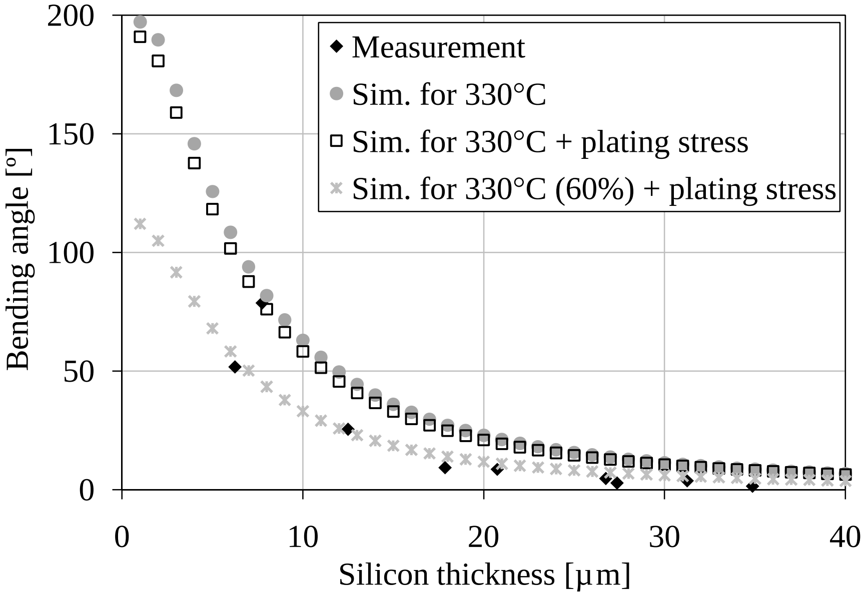

Figure 5.24: Measured bending angle at 330°C for 9 thin chips (sample group A4) with different silicon thicknesses and gold/tin lines. Simulations for silicon thicknesses between 1 µm and 40 µm and gold/tin line widths of 100 µm are also included in the diagram. The measured samples were reflowed in air and only 1 chip edge was considered. The original simulation results, simulation results corrected for plating stress, and simulation results based on plating stress and only 60% of thermal film stress are presented in the diagram.

Even though the qualitative similarity between original simulation and measurement was good, the absolute measured bending angles were about 50% smaller than the simulation results. For one single chip the results showed very good similarity. This was, however, considered a random result, because this chip had an unusually large thickness range on the measured edge between 4 µm and 9 µm. Taking into account the plating stress improved the agreement between simulation and measurement only very slightly. Reducing the thermal stress in the simulation by 60% and including the plating stress, however, lead to a good agreement between measurement and simulation. Possible explanations for the observed deviation between measurement and original simulation are relaxation processes in and softening of the nickel and gold layers at elevated temperatures. In Figure 5.25 the gradient of the bending angle (with respect to a temperature change) is given as a function of temperature (i.e. $\Delta\alpha/\Delta T=f(T)$). The graph is based on data for temperatures just below the melting point of tin. Although showing large fluctuations, the data suggests that the bending angle changed less for a constant temperature increase at elevated temperatures. The regression line fitted to the data yielded a reduction of the gradient $\gamma = \Delta\alpha/\Delta T$ by about 50% between 50°C and 230°C. This large decrease of stiffness was not expected, but was possibly due to structural reorganization of the electroplated metal layers. The softening was believed to continue when the temperature was raised to 330°C. On the other hand, at 330°C the relevant metal layer system was different, consisting of only 2 µm of gold and 2 µm of nickel. At this temperature, the proportion of the high melting element nickel was much larger and the very soft tin was missing completely. These arguments suggested a smaller gradient than in the Au-Ni-Au-Sn layer system. Altogether, a stiffness reduction by 40% in the Au-Ni layer system at 330°C was reasonable.

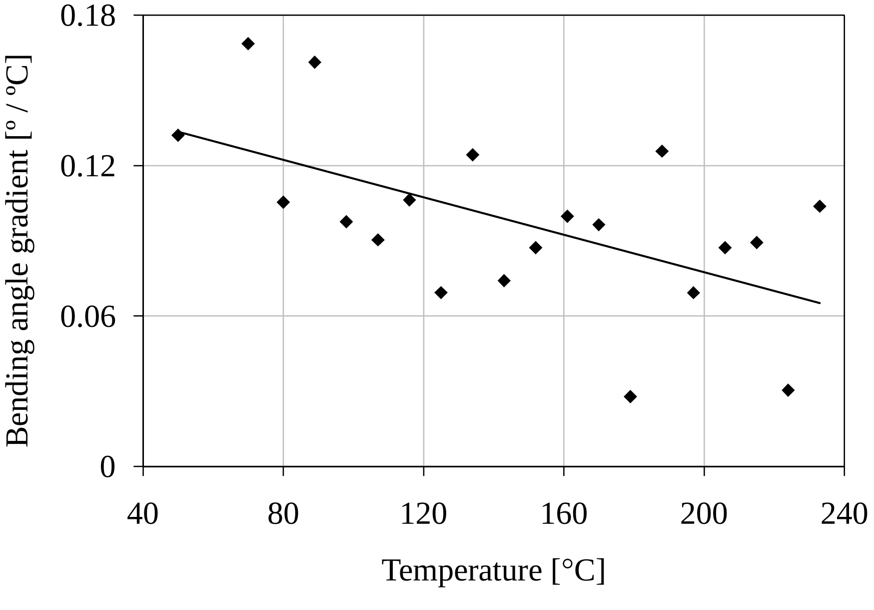

Figure 5.25: Measured bending angle gradient as a function of temperature for the ramp-up of a thin chip with gold/tin lines. A regression line was fitted to the data in order to analyze the results quantitatively.

5.4.5 Chip bow during ramp-down from 330°C to 280°C (process section γ_3)

Although deformations during this part of the reflow cycle should, in theory, not have an impact on deformations after reflow, they were used to verify the simulation model. The gradient of the bending angle (see γ_3 in Figure 5.20) was measured for 10 thin chips during reflow and correlated to FEM simulations. The results in Figure 5.26 showed good agreement between simulation and measurement. Two chips with small silicon thicknesses had slightly lower gradients than calculated in the simulation. This could, however, be explained by the very inhomogeneous silicon thickness distribution in these two chips. The range was between 4 µm and 9 µm for the one chip and between 2 µm and11 µm for the other. The simulation, however, was carried out for an average silicon thickness, possibly causing the observed deviation between measurement and simulation.

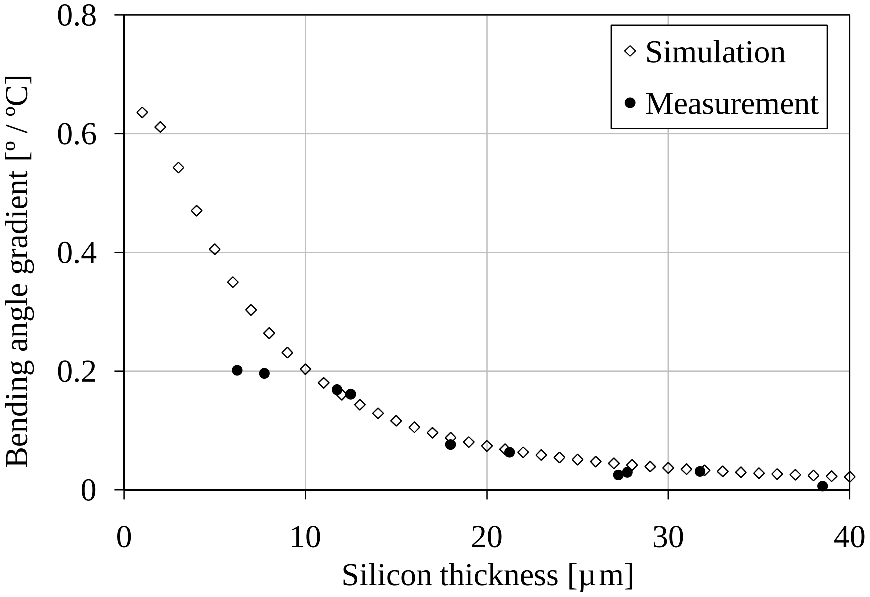

Figure 5.26: Simulation and measurement of bending angle gradient ($\Delta\alpha/\Delta T$) as a function of silicon thickness for thin chips with gold/tin lines at temperatures above the melting point of gold/tin. The measured data was based on 10 chips (sample group A6), for which the bending angle was measured between 330°C and 280°C.

5.4.6 Chip bow during ramp-down from 190°C to 20°C (process section γ_5)

During ramp-down from 190°C to 20°C, the bending angle showed a linear dependence on the temperature. Therefore, bending angles in this temperature range were analyzed for the representative temperature segment between 100°C and 40°C .

19 chips that were reflowed in a formic acid atmosphere were later heated to temperatures of 40°C, 70°C, and 100°C using the flip chip bonder and the bending angle was measured in-process on two sides of each chip. From this data the bending angle gradient was calculated and is drawn as a function of the silicon thickness in Figure 5.27. For comparison, simulation results were also included in the diagram. Because of mechanical coupling between the two measured edges of the chip, the bending angles were likely to influence each other. Therefore, silicon thickness values and bending angles were averaged per chip and are presented in the bottom diagram of Figure 5.27. With the exception of two outliers with silicon thicknesses above 40 µm, the measurements averaged per chip showed good agreement with the simulated bending angle gradients.

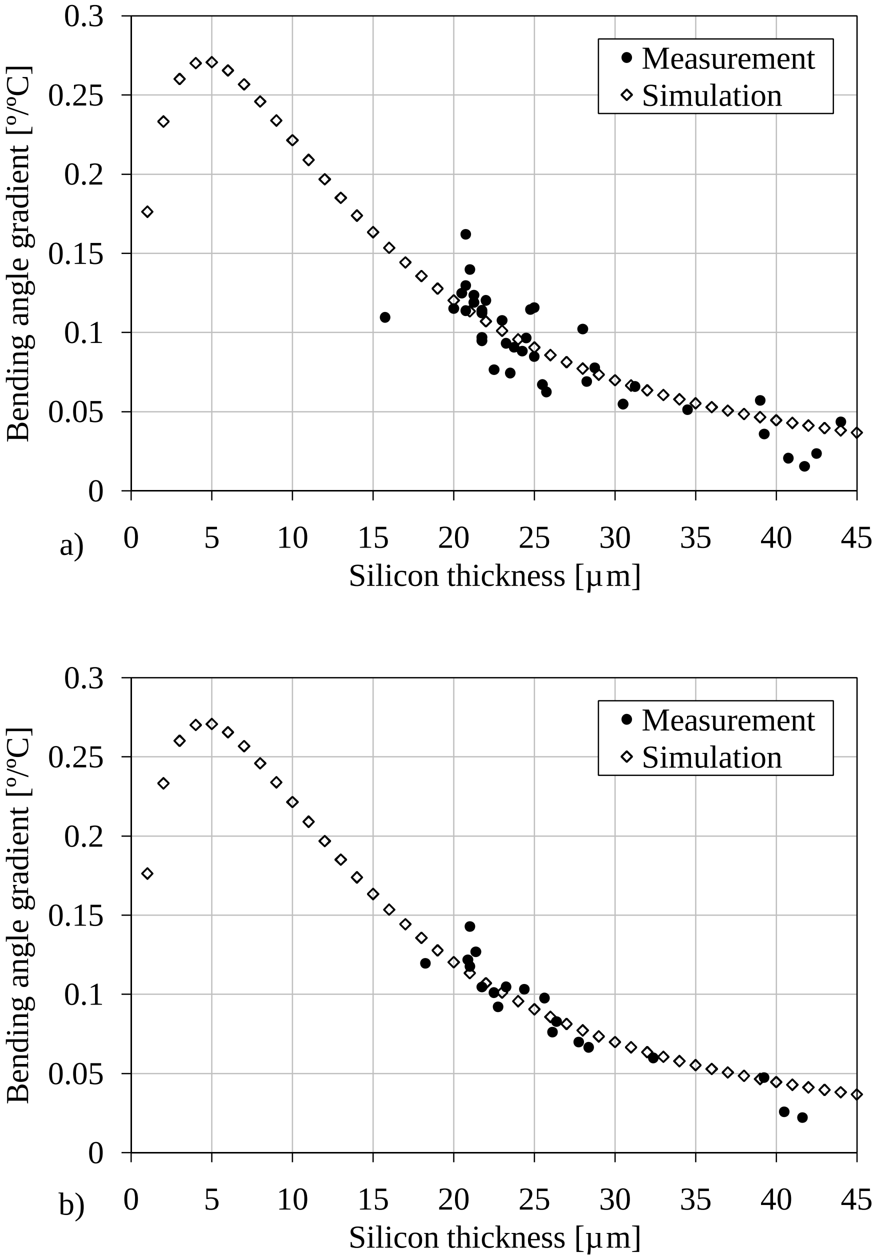

Figure 5.27: Bending angle gradient as a function of silicon thickness for reflowed chips. The bending angle was simulated for chips with 100 µm wide gold/tin lines and silicon thicknesses between 1 µm and 45 µm. The measurement results were based on 19 chips (sample group A5) reflowed in a formic acid atmosphere. Each chip was analyzed after reflow by measuring the silicon thickness on 2 edges and the according bending angles at temperatures of 40°C, 70°C, and 100°C. From these bending angle measurements the bending angle gradient was calculated; a) silicon thickness and bending angle calculated separately for the 2 edges of each chip; b) silicon thickness and bending angle averaged per chip.

5.4.7 Chip bow during ramp-down from 280°C to 190°C (process section γ_4)

During ramp-down, eutectic gold/tin solder undergoes the $\zeta \rightarrow \zeta'$ phase transition at 190°C [122], [170]. This effect can be seen in the bending angle measurements during cooling shown in Figure 5.28. To calculate the actual transition temperature from these measurements, two regression lines were added for data points with a temperature above or below the expected 190°C transition temperature. The intersection of these regression lines yielded a transition temperature of 194°C, which was reasonably close to the literature value. The deviation between the measured and the literature values for the transition temperature may be explained by errors in the bending angle measurement. To reduce these errors, the bending angle data of 10 chips was superimposed, regression lines were determined, and a transition temperature of 186°C was calculated. The deviation of 4°C from the ideal value of 190°C can be explained by thermal lag between the heating plate and the chip.

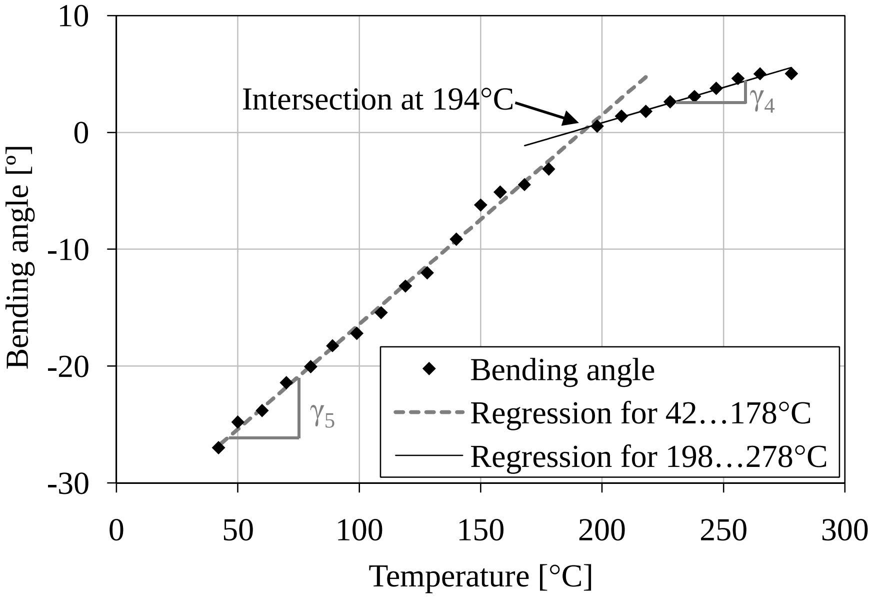

Figure 5.28: Optical in-process bow measurements for a single chip with gold/tin lines as a function of temperature during ramp-down. Regression lines, based on the data below and above the ζ→ζ' transition temperature of 190°C, were added. The intersection of these two regression lines was determined as 194°C, which is only slightly off the expected value of 190°C. The two phases have largely differing mechanical properties, visible as the two different bending angle gradients γ_5 and γ_4.

The regression lines in Figure 5.28 were used to calculate material parameters for the FEM simulation above 190°C. For γ_5 the simulation was previously shown to agree well with the measured bending angle gradient. The gradient γ_4 is 66% smaller compared to γ_5 for the chip in Figure 5.28. Taking into account the temperature ranges for the two different gradients, a combined gradient of gold/tin between 20°C and 280°C was calculated:

$$\gamma_{4,5} = \frac{\gamma_5 \cdot (190°\mathrm{C} - 20°\mathrm{C}) + \gamma_4 \cdot (280°\mathrm{C} - 190°\mathrm{C})}{280°\mathrm{C} - 20°\mathrm{C}}. \tag{5.10}$$

Here, the gradient γ_5 was assumed to be valid between 20°C and 190°C (and not only for the measured range of 42°C to 190°C). The data plot in Figure 5.28 supports this assumption, because for temperatures below 100°C the data shows nearly linear behavior.

The resulting gradient $\gamma_{4,5}$ was 23% smaller compared to γ_5 for the chip in Figure 5.28. The reduction of $\gamma_{4,5}$ and γ_4 compared to γ_5 depended on the silicon thickness. In Figure 5.29 measurements of the gradients for 10 thin chips during cooling are shown. Exponential regression curves were added for better comparison. To transform this reduced gradient into adapted material parameters for the simulation, the correlation between individual material parameters and the bending angle had to be known. A set of simulations was carried out, where the effect of a small change of individual material parameters by 1% on the bending angle was determined. The results are shown in Figure 5.30.

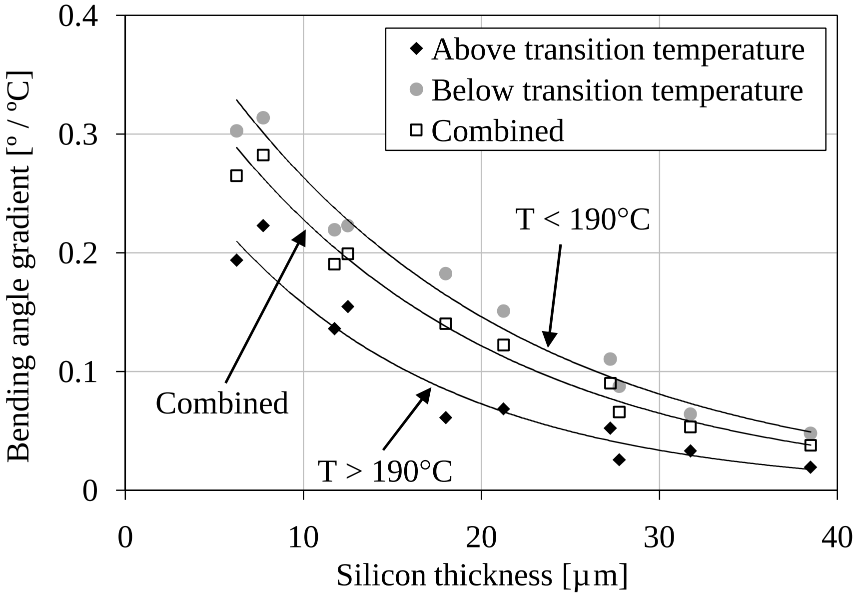

Figure 5.29: Bending angle gradient (Δα/ΔT) as a function of silicon thickness. 10 thin chips (sample group A6) were reflowed in air and the bending angle was measured during ramp-down between 280°C and 42°C. Because eutectic gold/tin undergoes a phase transition at 190°C (resulting in the change of mechanical properties), the measured data was plotted separately for temperatures above and below the transition temperature. The combined data was also plotted and exponential regression curves were added for each of the 3 data sets. The combined data plot is closer to the data points for T < 190°C, because this temperature range spanned 170°C compared to only 90°C for temperatures above 190°C.

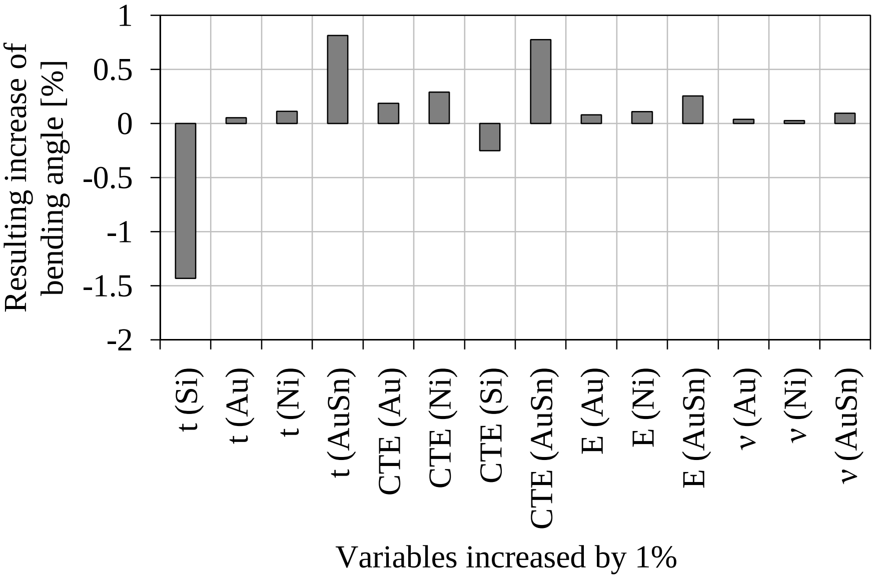

Figure 5.30: Correlation between material parameter changes and the resulting bending angle change during ramp-down for a silicon chip with a thickness of 28.5 µm. The material parameters were increased by 1% and the resulting bending angle change was calculated in individual FEM simulations for each parameter. Acronyms used in the diagram: t = thickness, CTE = coefficient of thermal expansion, E = Young's modulus, ν = Poisson's ratio.

Both the coefficient of thermal expansion and the Young's modulus of eutectic gold/tin had a positive correlation with the bending angle. This correlation was, however, dependent on the silicon thickness as can be seen in Figure 5.31, where the CTE respectively Young's modulus of gold/tin were reduced by 10% for a whole range of silicon thicknesses.

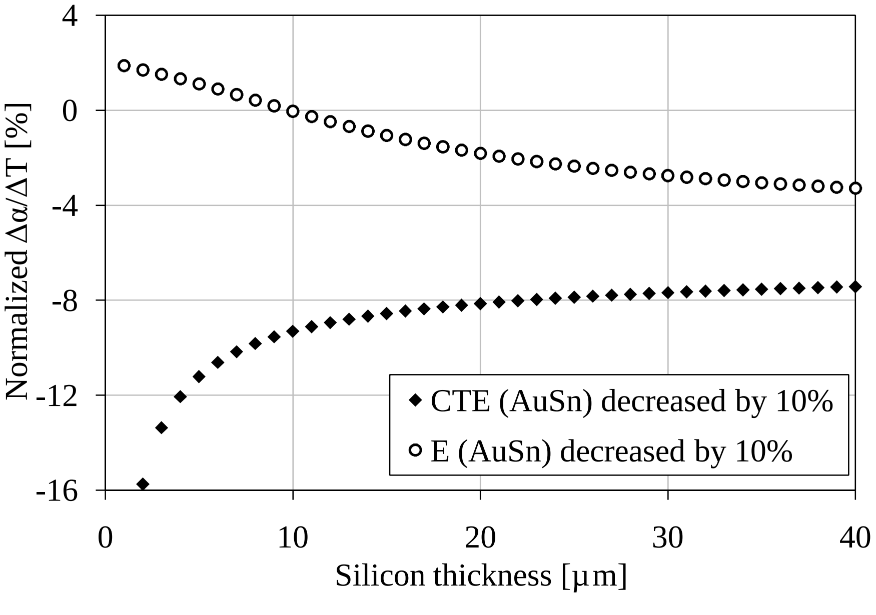

Figure 5.31: Simulation results for the influence of a 10% decrease of the coefficient of thermal expansion or Young's modulus on the bending angle gradient during cooling as a function of silicon thickness. The bending angle gradient was normalized to its values for unchanged CTE and Young's modulus. The graph was based on simulation results for a chip with 100 µm wide gold/tin lines.

A CTE decrease of 10% resulted in the reduction of the gradient ($\Delta\alpha/\Delta T$) for all silicon thicknesses, while a Young's modulus decrease by 10% resulted in smaller gradients at large silicon thicknesses and larger gradients at very small silicon thicknesses. In [170] the Young's modulus of eutectic gold/tin was shown to be slightly larger at temperatures above 190°C than below 190°C (70.6 GPa vs. 69.0 GPa). This very small increase of about 2.5% suggested a bending angle gradient increase of about 1% for thicker silicon chips ($t_{Si} \approx 40\ \mu m$) and a bending angle gradient decrease of about 0.5% for thinner chips ($t_{Si} \approx 1\ \mu m$). These percentages were calculated by multiplying the values in Figure 5.31 with -1/4. The actual gradient change at 190°C was about -66% and could not be explained by the slight increase of the Young's modulus. The observed gradient change at 190°C required a larger CTE below 190°C. In [170] it was found that the CTE of annealed (250°C for 4 days) eutectic gold/tin increased from 15.6 ppm/°C to 16.8 ppm/°C during cooling at 190°C. This small increase by 7.6%, however, could also not explain the large gradient differences observed during reflow of thin chips. Assuming a CTE increase of 25% during cooling at 190°C resulted in a better approximation between measurement and simulation as can be seen in Figure 5.32. At small silicon thicknesses, the agreement was good, while at large silicon thicknesses the simulated gradients were still too large. Contrary to the findings for eutectic gold/tin in [128], the CTE of solder generally rises with temperature, while the Young's modulus decreases. According to [55], for example, the CTE of several common solders typically increases by 10 to 15% when comparing the 20°C to 100°C range with the 100°C to 150°C range, while the Young's modulus is reduced by 10% to 50% at 150°C compared to its value at room temperature. Recalling the different influences of a Young's modulus decrease and a CTE decrease as calculated in Figure 5.31, a better approximation was expected by decreasing the Young's modulus and the CTE at the same time in simulations for the temperature range between 190°C and 280°C. Due to the large number of possible combinations and consequently required simulation time, this avenue was not further pursued. However, considering the gradient over the complete temperature range from 280°C to 42°C also allowed for a good approximation between simulation and measurement by adapting only the CTE. A CTE reduction for eutectic gold/tin by 10% was suitable for this temperature range. The results are shown in Figure 5.33.

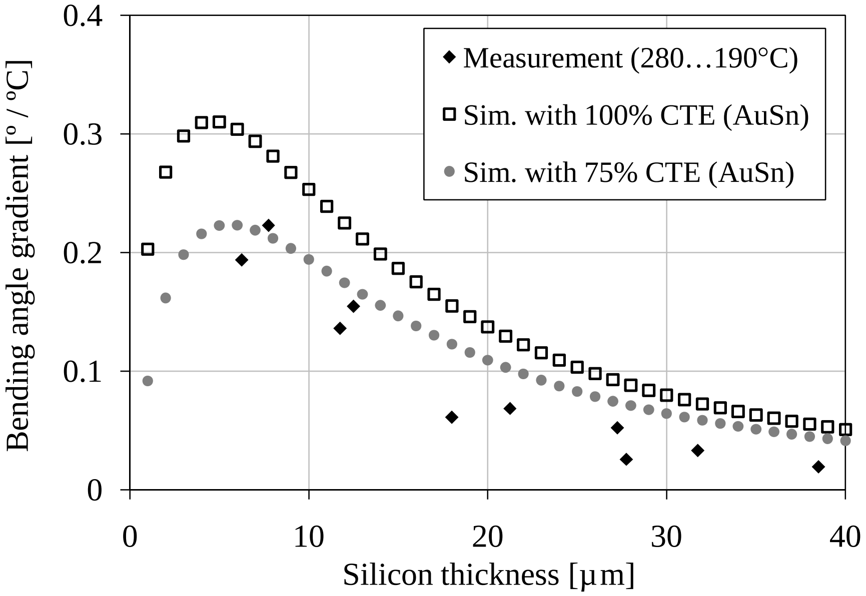

Figure 5.32: Comparison of simulated and measured bending angle gradients (Δα/ΔT) as a function of silicon thickness for 190°C < T < 280°C for 10 chips (sample group A6). The simulation was carried out with the original CTE of Au/Sn and with a CTE reduced by 25% to 12.15 ppm/°C.

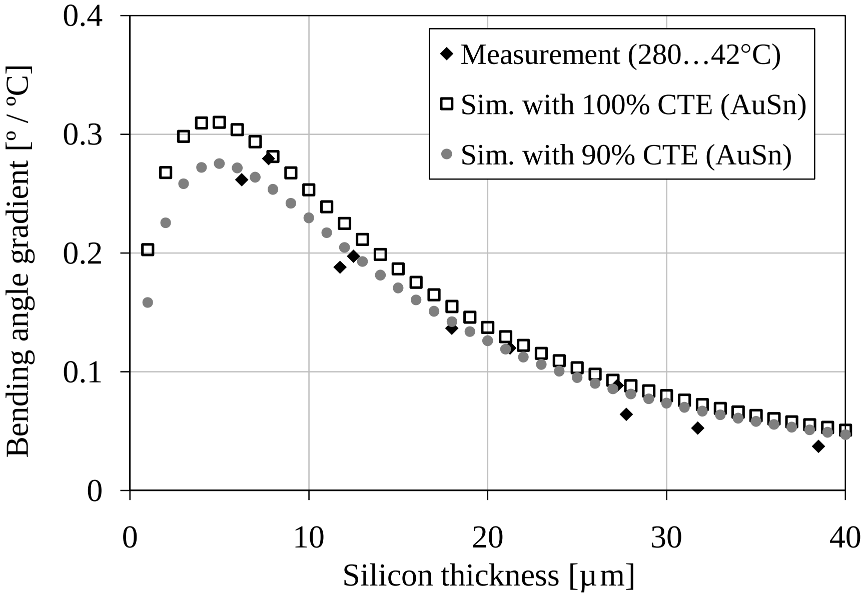

Figure 5.33: Comparison of simulated and measured bending angle gradient (Δα/ΔT) as a function of silicon thickness for 42°C < T < 280°C for 10 chips (sample group A6). The simulation was carried out with the original CTE of Au/Sn and with a CTE reduced by 10% to 14.58 ppm/°C.

5.4.8 Summary of bow measurements and simulations of thin chips with gold/tin lines

The bending angle evolution during reflow of thin chips with gold/tin lines was analyzed intensively in the previous subsections. A multitude of results was found and a good understanding of the macroscopic deformations during the reflow process of chips with electroplated gold and tin layers was gained. The results are summarized in Table 5.7.

Table 5.7: Summary of major results regarding the agreement between simulation and measurement of deformations in thin chips during reflow. The letters in the first column refer to the different parts of the bending angle diagram in Figure 5.20. The term "qualitative" refers to the similarity of the curve shapes of simulated and measured deformations as a function of silicon thickness.

	Description	Agreement between Sim. and Meas.		Graph in
		Qualitative	Quantitative	
A	Bending angle α in as-plated samples	good	good (using 55% of the wafer level plating stress)	Figure 5.17
γ_1	Δα/ΔT for Au-Ni-Au-Sn between 40°C and 100°C	good	good	Figure 5.21
G	Bending angle α at 330°C	good	good (including plating stress and using 60% thermal stress)	Figure 5.24
γ_3	Δα/ΔT for Au-Ni between 330°C and 280°C	good	good	Figure 5.26
γ_4	Δα/ΔT for Au-Ni-Au/Sn between 280°C and 190°C	good	medium (using a CTE for Au/Sn that is reduced by 25%)	Figure 5.32
γ_5	Δα/ΔT for Au-Ni-Au/Sn between 190°C and 42°C	good	good	Figure 5.27
$\gamma_{4,5}$	Δα/ΔT for Au-Ni-Au/Sn between 280°C and 42°C	good	good (if CTE of Au/Sn is reduced by 10%)	Figure 5.33

Nearly all parts of the reflow were simulated with good agreement between simulation and measurement. The ramp-down between 280°C and 190°C was the only section, where good qualitative similarity between simulation and measurement was obtained, but where significant quantitative differences around 50% remained. This inaccuracy of the model, however, could be circumvented by simulating the complete cool-down of solidified gold/tin in one step. This simulation mode yielded good qualitative and quantitative results. Finally, the simulation results for individual parts of the reflow were merged to obtain a resulting simulated bending angle after reflow. The simulation results were merged in two ways. In both cases the bow in as-plated samples and the deformations during cooling from 280°C to 20°C were included. The deformations at 280°C, however, were either obtained by simulating the ramp-up from 20°C to 330°C and adding the cooling results from 330°C to 280°C (*Simulation 1*) or by directly simulating the bow at 280°C (*Simulation 2*). Results for both simulation paths and actual measurements of reflowed chips are compared in Figure 5.34. The two simulation paths lead to similar results and the overall agreement with measurements was very good. Because Simulation 2 is less complex, it was favored. For very small silicon thicknesses, only one data point was available, which was gained by measuring the bow locally for a chip with a silicon thickness of 4 µm at one end of the analyzed chip edge. This data point proved that the FEM model was valid even for very small silicon thicknesses and that the bow reduction predicted by the simulation in such chips after reflow was indeed present after reflow. At a silicon thickness of 2 µm, the simulation predicted a positive bending angle (that is a concave deformation) after reflow. One chip was fabricated with a local silicon thickness on the edge as small as 2 µm and indeed this chip showed a slight concave deformation on this thin side after reflow (for an image of this chip see Figure 5.11).

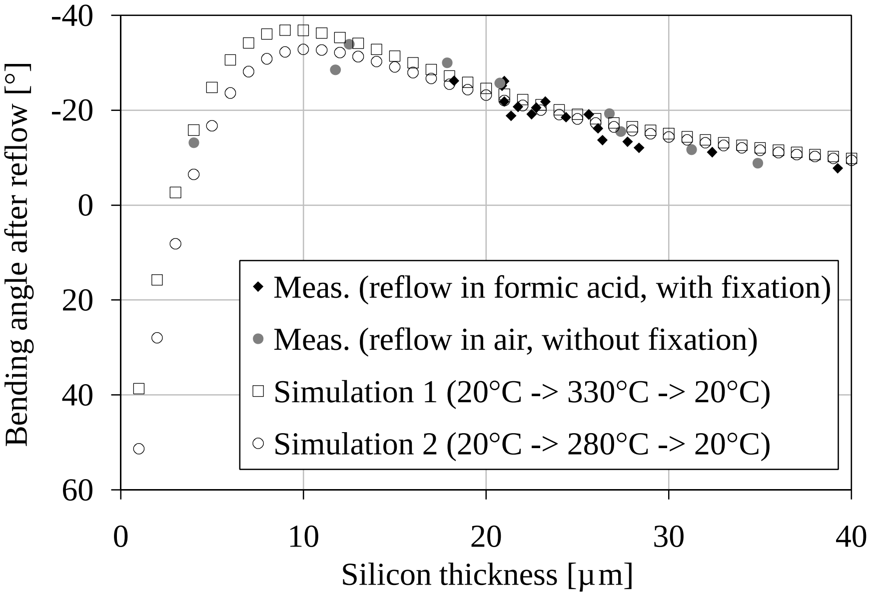

Figure 5.34: Measurement of bending angle after reflow and results of the complete simulation of the reflow process. One group of measurements was carried out on chips that were reflowed in a formic acid atmosphere and held down by a fixation, while the other group of chips was reflowed in air without use of a fixation. The simulations were carried out in two ways. Both simulations included the plating stress and ramp-down between 280°C and 20°C. In Simulation 1 the bow at 280°C was simulated by calculating the bow for 330°C and then adding the deformation reduction between 330°C and 280°C, while in Simulation 2 the bow at 280°C was directly calculated.

5.4.9 Local deformations

A total of 30 chips with gold/tin line widths between 25 µm and 200 µm were reflowed in a formic acid atmosphere and the local silicon deformations were measured using a profiler. Three chips broke during handling, so that 27 chips could be measured. The individual results are shown in Figure 5.35. Chips with the same line width showed similar local deformations, proving that the analysis technique for local silicon deformations described in Section 5.2.2.3 was reliable. Regarding different line width, a quadratic dependence of the local deformations on the line width was found.

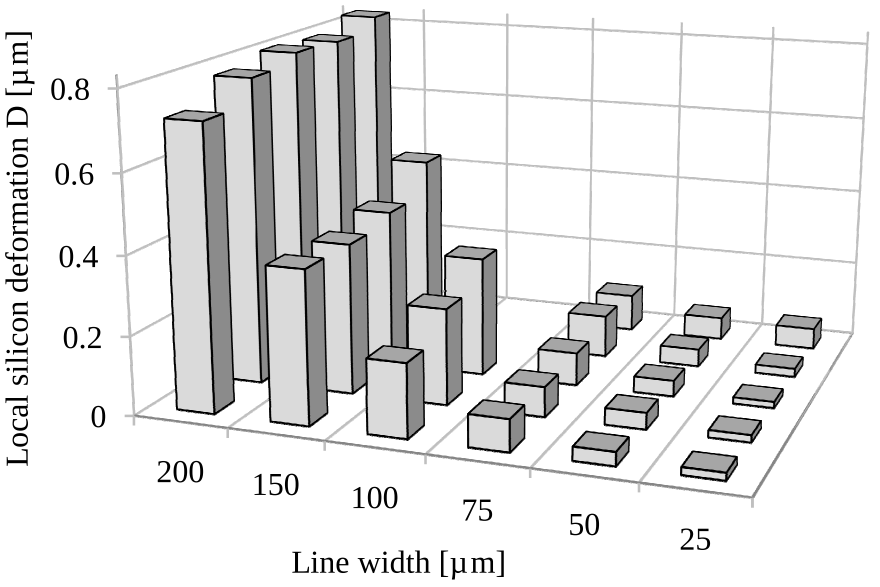

Figure 5.35: Measured local silicon deformations in 27 thin chips (sample group A7) with an average silicon thickness of 28 µm and gold/tin line widths between 25 µm and 200 µm.

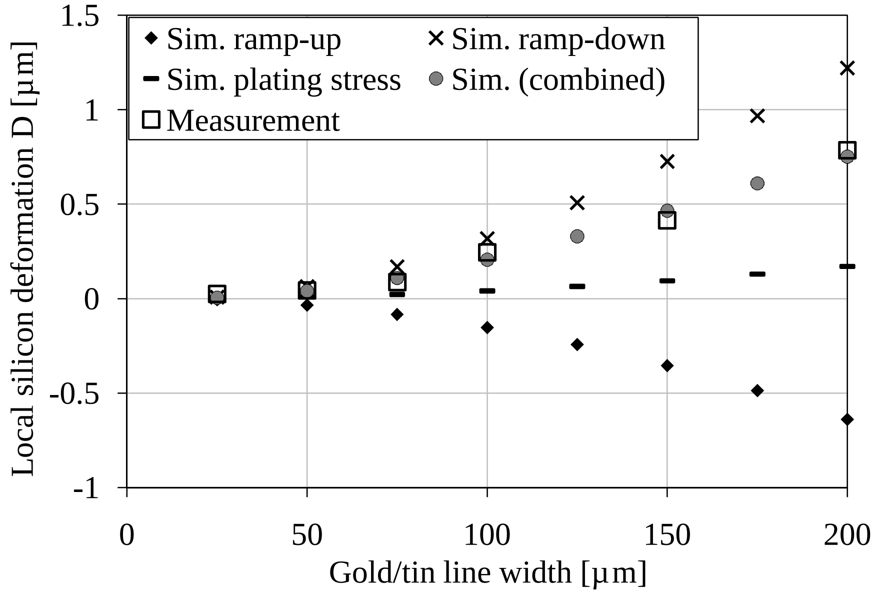

Figure 5.36: Simulated and measured local silicon deformations in reflowed chips with an average silicon thickness of 28 µm and gold/tin lines of varying width (sample group A7). The simulation was carried out in three steps according to the sequence of Simulation 2 in the previous section. This simulation sequence was composed of separate simulations for the ramp-up, cool-down, and the plating stress influence on deformations. Without any further parameter tuning, very good agreement between simulation and measurement was obtained.

In order to compare the measured local deformations with simulation results, the measured data was averaged for each line width. These averaged values and according simulation results are shown in Figure 5.36. The simulation results were obtained using the same optimized simulation sequence (Simulation 2) as in the previous section, consisting of the plating stress (using 55% of the wafer level plating stress), the ramp-up to 280°C (using 60% thermal stress), and the ramp-down (using a CTE for eutectic gold/tin reduced by 10%). Without further parameter adjustments, very good agreement between simulation and measurement is obtained.

5.4.10 Influence of chip fixation during reflow on bending angle

When the reflow was carried out in the solder oven in a formic acid and nitrogen atmosphere, the thin chips needed to be fixated by a fine needle tip as shown in Figure 5.37. This needle tip prevented the chip from deforming macroscopically during heating, because the rising of the chip center was blocked. During cooling, however, the edges of the chip started rising, which was not impeded by the needle tip.

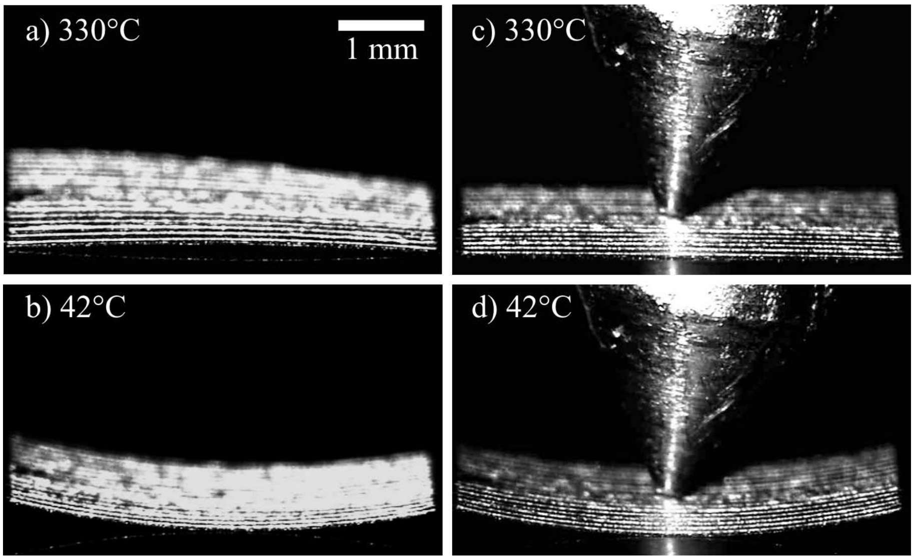

Figure 5.37: Bow of a thin silicon chip with long gold/tin lines at 330°C and 42°C, with fixation (left images) and without fixation (right images); a) at 330°C without fixation; b) after ramp-down at 42°C without fixation; c) at 330°C with fixation; d) after ramp-down at 42°C with fixation. The fixation prevents concave deformations at 330°C (c), but does not impede convex deformations after cooling (d). All 4 images show the same chip for better comparison.

Quantitative measurements of the bending angle with and without fixation are shown in Figure 5.38. The chip was reflowed prior to the measurements and already had a strong bow at the beginning of the reflow process. During ramp-up, the deformations were reduced and at approximately 210°C the chip was entirely flat. Further heating allowed the chip without fixation to lift its center off the heating plate resulting in a concave bow. The chip with fixation, however, was restrained and remained flat even at 330°C. At the eutectic point of gold/tin of 280°C, there was a notable bow difference of 3.5° between the process runs with and without fixation. This meant that the bow at the solidification point was different if a fixation was used. After ramp-down to 42°C this difference was still visible as a 2.3° gap between the two process runs. Using a fixation increased the bow by 9.1% in the case of the chip in Figure 5.38.

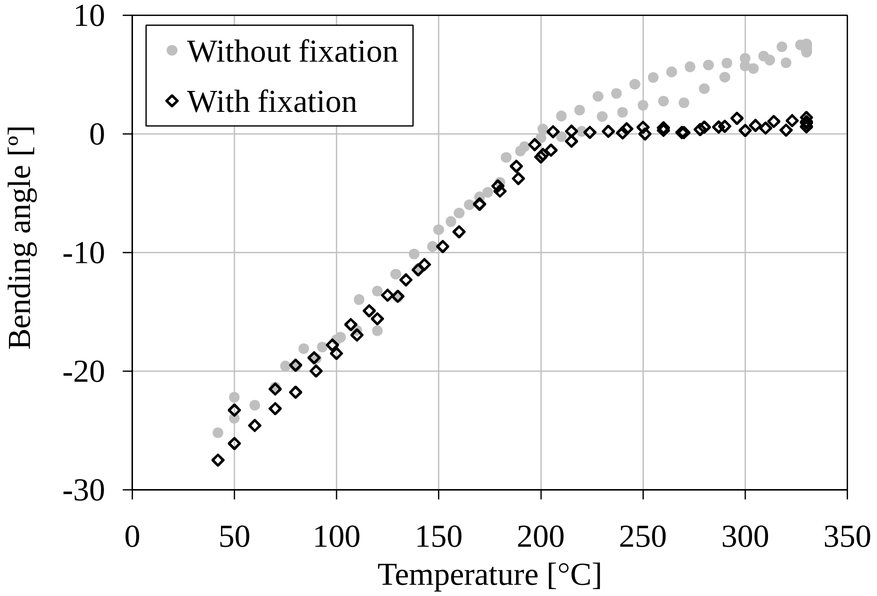

Figure 5.38: Bending angle measurement results during reflow of the chip in Figure 5.37. During ramp-up, the chip showed different behavior above 210°C, where the fixation prevented further deformation in the one chip. During ramp-down, when the eutectic temperature of 280°C was crossed, the fixated chip was about 3.5° less bent compared to the case without fixation. This bow difference partly remained after ramp-down to 42°C, where the remaining bow difference was about 2.3°.

5.5 Reflow of chips with Kelvin contacts and Daisy chains

The design of the thin chip samples considered here was described in Section 3.1. The chips were four times larger in area than the chips with gold/tin lines in the previous section. Therefore, silicon thickness fluctuations across individual chips were expected to be larger. Extremely thin chips with thicknesses between 3 µm and 7 µm were fabricated, where small thickness deviations had a large impact on deformations. In order to reduce the influence of errors resulting from the measurement of the silicon thickness on the chip edge in samples with an inhomogeneous thickness distribution, the thickness was calculated from the measured weight of each chip. Each chip comprised 3 major areas with different bumps diameters. These areas showed different deformations, but, due to their mechanical coupling, could not be considered separately. In Figure 5.39 bending angle measurements as a function of bump diameter and silicon thickness are presented for 3 chips with varying thickness distributions.

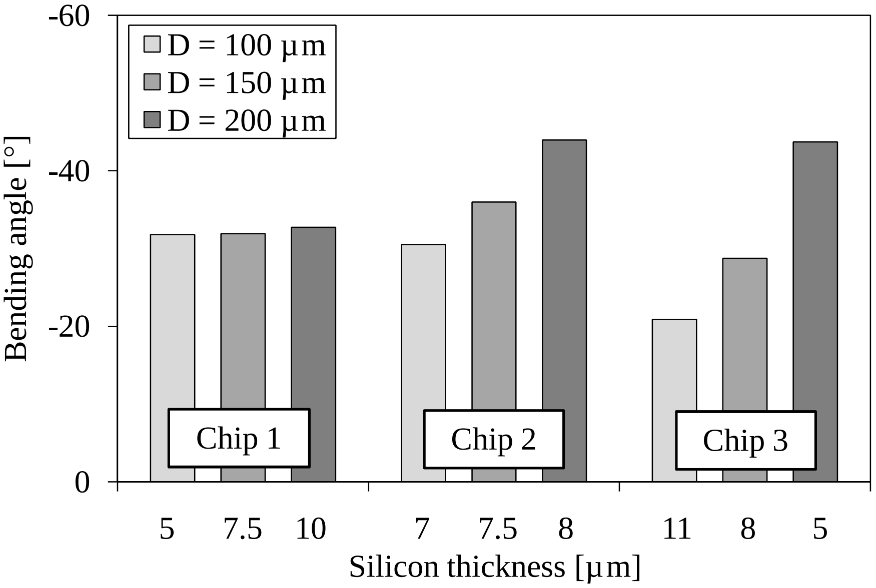

Figure 5.39: Measured bending angle in 3 reflowed chips as a function of the local silicon thickness. In each chip the bending angle was measured separately for each of the 3 areas containing only 100 µm, 150 µm, or 200 µm bumps. The silicon thickness on the chip edge was measured on two sides (near the 100 µm and 200 µm bumps). To obtain the silicon thickness for the area in between (containing the 150 µm bumps), the thickness measurements were averaged.

Chip 1 had a thickness of 10 µm on the side with the 200 µm bumps and of 5 µm on the side with 100 µm bumps. This combination yielded almost identical bending angles in the different areas. Chip 2 on the other hand had a homogeneous silicon thickness, resulting in the continuous increase of the bending angle (in absolute values) from the 100 µm to the 200 µm area. In Chip 3 this effect was even more pronounced, because the silicon thickness in the 200 µm area was about half of that in the 100 µm area. Therefore, in order to better compare the measured and simulated chip bows, averages had to be calculated per chip (based on three bow measurements in the three different chip areas) and also for each set of simulations (consisting of three individual simulations for the 100 µm, 150 µm, and 200 µm bumps). Unlike in thin chips with gold/tin lines, the bending axis was not generally parallel to the chip edge in chips with gold tracks and round gold/tin bumps. In the following section this observation is analyzed in detail and an explanatory model is given.

5.5.1 Angular shift between bending axis and chip edge

In this thesis, reflowed chips with gold tracks and electroplated gold-tin bumps on a gold-nickel UBM generally showed cylindrical deformations after reflow. The axis of this deformation cylinder (also called bending axis) usually was not parallel to the chip edge. By projecting the bending axis onto the chip surface such that the projection runs through the chip center, an angle θ was formed between this projection and the chip edge. This angle was a good quantitative measure for the direction of bending. In Figure 5.40 this angle is illustrated for θ = 0° and θ = 45°. Measuring the z coordinates of the chip at nine locations marked M1, … , M9 in Figure 5.40 allowed the calculation of the angle θ using a complex algorithm. In Figure 5.41 the nine measured z coordinates of the silicon surface in each of four bent chips are depicted as a function of the xy coordinates. The calculated directions of the different bending axes can be qualitatively verified by the diagrams.

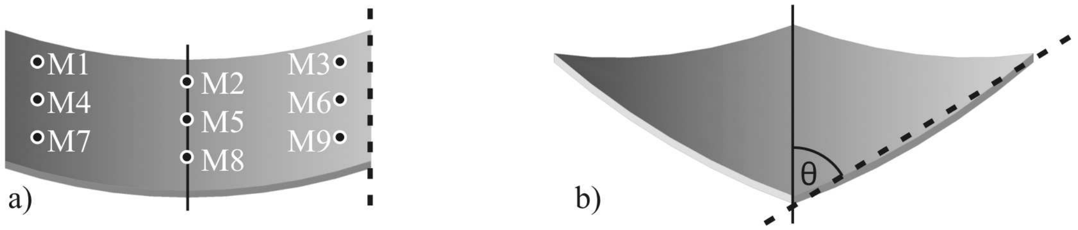

Figure 5.40: Schematic illustration of the angle θ between the bending axis projection (continuous line) and the chip edge (dashed line); a) bending axis direction and chip edge are parallel (θ = 0°). The 9 measurement locations M1,…, M9 later used to determine the bending axis direction are also marked; b) bending axis projection and chip edge are at an angle of θ = 45°

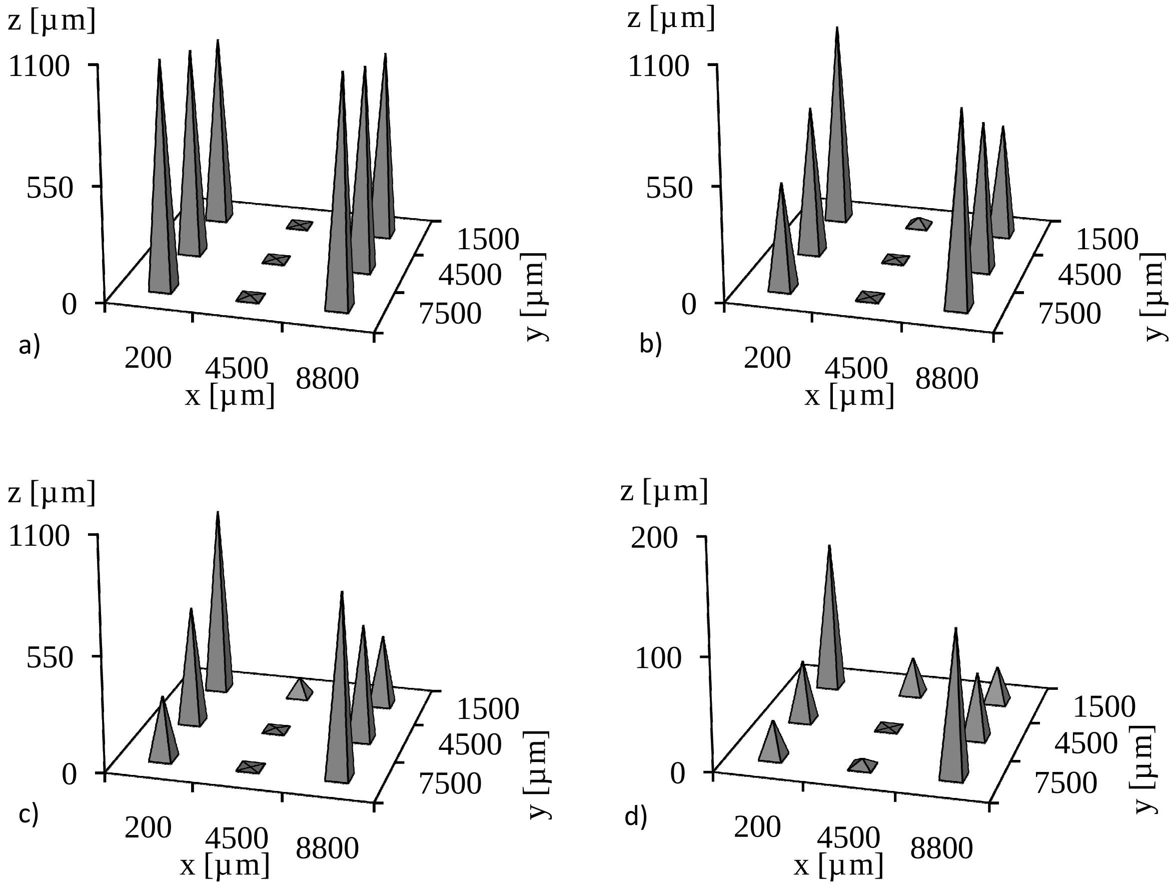

Figure 5.41: Measured z coordinates of the silicon surface at 9 locations in each of 4 reflowed chips with different silicon thicknesses. All chips showed roughly cylindrical deformations with different angles θ between the bending axis projection and the chip edge (which was identical to the y-axis in these measurements). a) θ = 6°, $t_{Si} \approx 3$ µm. The increasing bow for increasing bump diameters (from 100 µm at y = 1500 µm to 200 µm at y = 7500 µm) can also be seen; b) θ = 12.5°, $t_{Si} \approx 6$ µm; c) θ = 22.7°, $t_{Si} \approx 8$ µm; d) θ = 42.5°, $t_{Si} \approx 40$ µm.

The direction of the bending axis and the measurement data at the nine locations were used to determine the radius of curvature for each chip using the trilateration formula in Appendix B based on corrected x coordinates resulting from the angle θ (see Appendix C for details). In Figure 5.42 the angle θ between the chip edge and the bending axis projection is given as a function of the silicon thickness for 27 reflowed chips. Discarding 4 outliers, the remaining 23 chips suggested that for larger silicon thicknesses a 45° angle was more likely, while for smaller silicon thicknesses an angle of θ = 0° was more likely. Linear and logarithmic regression curves were added in the diagram. Because the θ gradient was larger at small silicon thicknesses, the logarithmic approximation was more accurate.

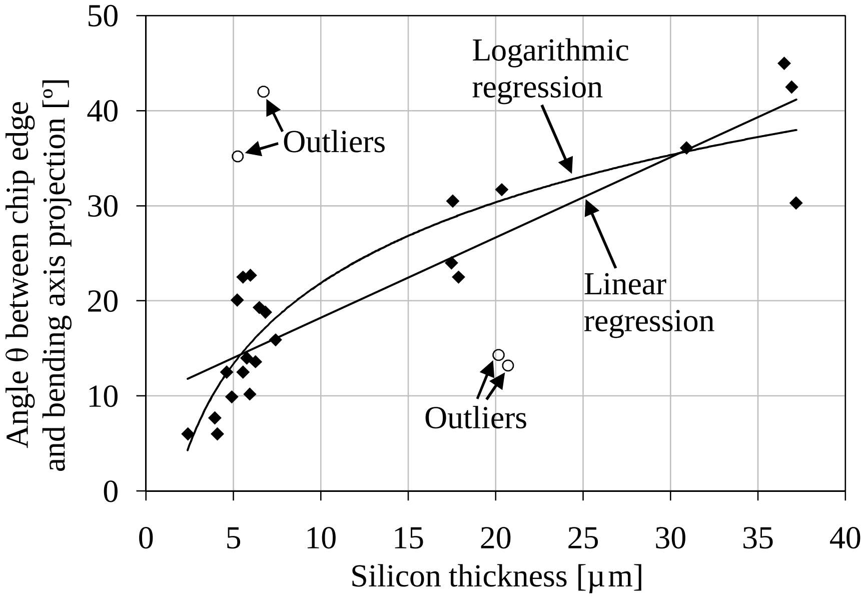

Figure 5.42: Angle θ between chip edge and bending axis projection as a function of silicon thickness for 27 chips (sample group C1). 4 of these chips were outliers and were not included in the further analysis. For the remaining 23 data points a linear and a logarithmic regression were calculated.

The observed behavior can be explained by two competing effects. Because of its anisotropic mechanical properties, the bulk silicon preferably bent around the <100> axes (which corresponded to θ = 45°). The track design, however, resulted in a minimal stiffness for bending around the <110> axes (corresponding to θ = 0°). In Figure 5.43 the major bending axes are shown. The proportion of gold in the tracks that was deformed by the stress in the reflowed solder was minimized for Axes B' and B" and maximized for Axes A' and A". In between, the gold track cross section area changed steadily. Although cross contraction of the gold tracks also favored bending around Axes B' and B", this effect was considered negligible due to the small height of the gold tracks compared to their widths (2 µm vs. 100 µm to 200 µm).

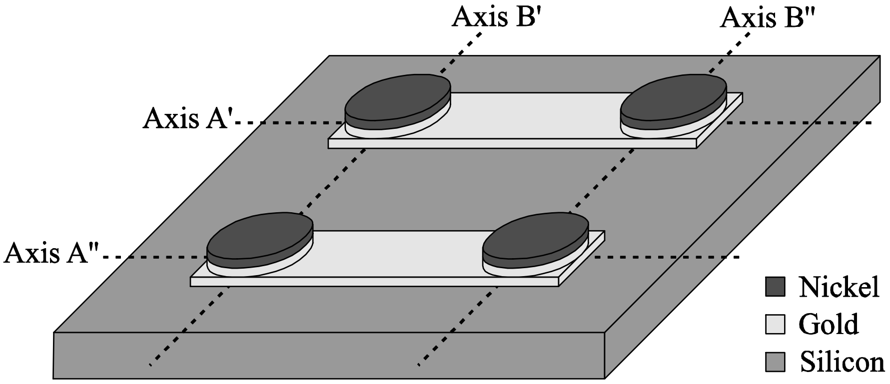

Figure 5.43: Schematic of silicon chip area with 2 gold tracks and 4 gold/nickel UBMs. The bending stiffness for bending around axes B' and B" was smaller than for bending around Axes A' and A" due to the smaller gold track cross section that needed to be compressed when bending occurred around B' and B".

For thick chips ($t_{Si} \geq 30$ µm) the anisotropic mechanical properties of silicon dominated, resulting in the bending axis being closer to the <100> direction (that is at an angle of 45° relative to the designated axes in Figure 5.43), while for thinner chips ($t_{Si} < 10$ µm) the track design dominated, resulting in θ being closer to 0°.

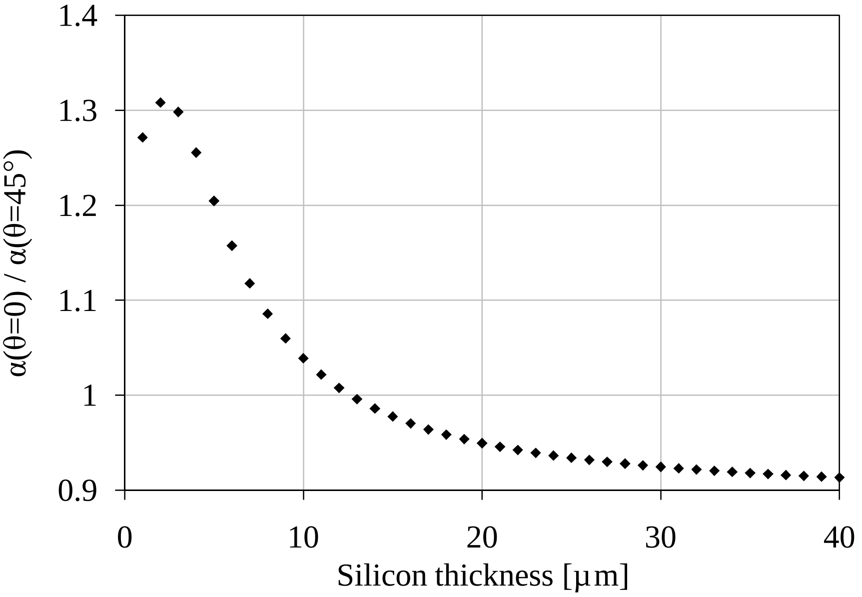

Figure 5.44: Simulated bending angle ratio for bow parallel to the chip edge and diagonally across the chip as a function of silicon thickness. At a silicon thickness of about 12.65 µm, the bows were equal, resulting in a predicted bending axis direction of 22.5° (that is half-way between the chip edge and the diagonal). For thinner chips the bending axis was expected to be shifted towards the chip edge, while for thicker chips the bending axis was expected to be shifted towards the diagonal.

When the simulation model was devised, the observed dependence of θ on the silicon thickness was not expected. Either only 0° or only 45° values for θ were expected for all chips. Therefore, the model only contained reference points to read out the bow for θ = 0° and θ = 45°. This data was available for a range of simulated silicon thicknesses and was used to verify the correctness of the simulation model. In Figure 5.44 the ratio of the simulated bending angle α around the <110> axes and the <100> axes is given. As expected the ratio was larger at small silicon thicknesses (resulting in preferred bending around the <110> axes in very thin silicon samples). For larger silicon thicknesses, the preferred bending axes are the <100> axes. The gradients of α(θ=0°) / α(θ=45°) in Figure 5.44 are also in accordance with Figure 5.42, where θ changes more steeply at small silicon thicknesses than at large silicon thicknesses. Finally, the silicon thickness for which α(θ=0°) equals α(θ=45°) was calculated as t_{Si} = 12.65 µm from the simulation data. For this silicon thickness the angle between the bending axis and the chip edge was expected to be θ = 22.5° (that is half-way between θ=0° and θ=45°). Calculating the angle values of the linear and logarithmic regression curves in Figure 5.42 for t_{Si} = 12.65 µm yielded θ values of 20.4° and 24.7°, respectively. These values were gained from regression curves of actual measurements and were very close to the value of θ = 22.5° predicted by the simulation for t_{Si} = 12.65 µm, indicating a correct simulation model.

5.5.2 Bending angle measurements and simulation results

In Figure 5.45 simulated and measured bending angles after reflow as a function of silicon thickness are compared. The silicon thickness was calculated from a weight measurement and the optically measured bending angle was corrected by taking the angle between the chip edge and the deformation cylinder axis into account.

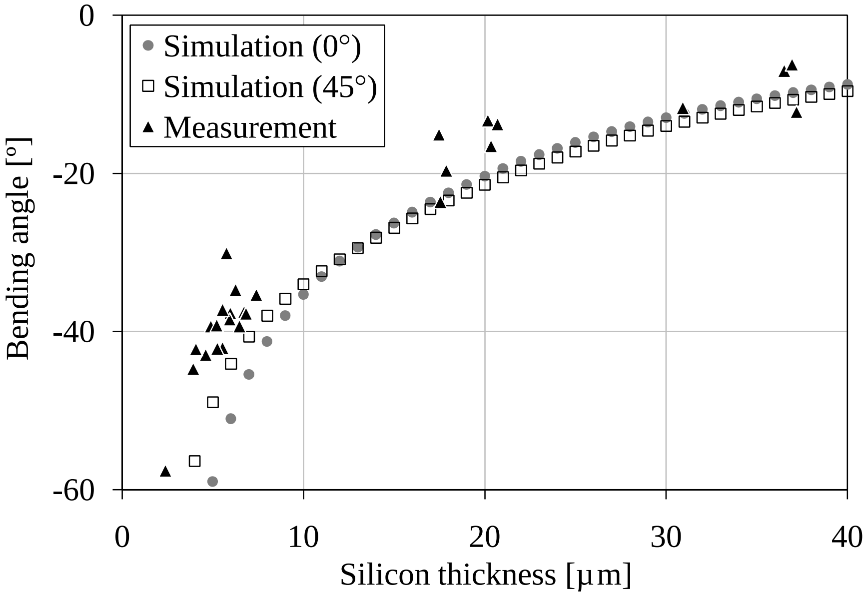

Figure 5.45: Measured and simulated bending angle as a function of silicon thickness in 27 chips (sample group C1). The simulation model allowed results to be extracted for bows parallel to the chip edge (0°) or diagonal across the chip (45°). For silicon thicknesses ≥ 8 µm the difference between the simulated curves is only a few percent, while below 8 µm substantial differences exist.

All bending angles (simulation and measurement) were normalized to a chip size of 10 mm. The simulation was carried out in 3 steps according to the optimized simulation sequence in Section 5.4.8, where the plating stress, ramp-up and ramp-down deformations were considered separately and then superimposed. Because the chips comprise 3 different bump diameters, each diameter was simulated separately and then the results were averaged. The simulation results were extracted both for θ = 0° and θ = 45°. As expected, the absolute simulated bow for thinner silicon was larger at θ = 0° and for thicker silicon it was larger at θ = 45°. The differences between these simulation data sets were generally in the range of a few percent. However, thin chips with $t_{Si} \leq 8\ \mu m$ tended to show larger bending angle differences between θ = 0° and θ = 45°. The qualitative agreement between the simulated and measured bending angles was very good. However, nearly all measured data points exhibited smaller bow (in absolute values) than predicted by the simulation. A possible explanation for this deviation is the ramp-up part of the 3-step simulation sequence, which may require different parameters. In Section 5.4.8 the metallic structures on the silicon chip (that were relevant for the bow at 280°C) consisted of equal volumes of gold and nickel. Here, however, the total nickel volume was about 6 times smaller than the gold volume relevant for the bow at 280°C. This significant difference justified an adaptation of the 60% stress proportion used for the ramp-up in Section 5.4.8. By increasing this proportion to 68%, the agreement between simulation and measurement was improved substantially. These results are shown in Figure 5.46.

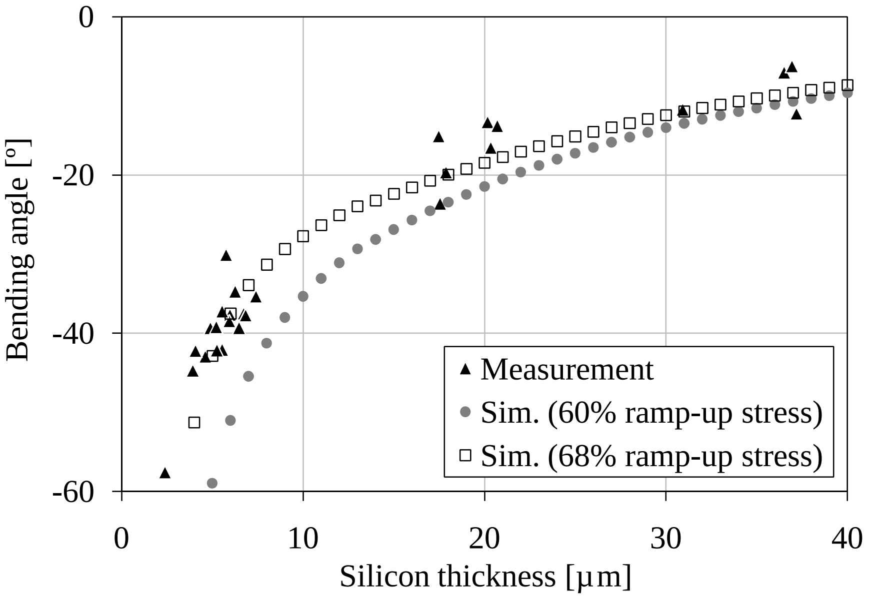

Figure 5.46: Measured and simulated bending angle as a function of silicon thickness for 27 chips (sample group C1). For silicon thicknesses ≥ 13 µm the simulation data for θ = 45° was used and for smaller silicon thicknesses the simulation data for θ = 0°. The simulation was carried out with the optimal simulation sequence from Section 5.4.8 (using 60% ramp-up stress) and with an adapted simulation sequence based on 68% ramp-up stress. The latter yielded a very good agreement between simulation and measurement.

5.5.3 Local deformations

The local deformation profile in a reflowed chip with Daisy chains and Kelvin contacts was measured using the surface profiler. The design was slightly different from the earlier samples, because a gold base thickness of 5 µm underneath the nickel diffusion barrier was implemented (which is two and a half times thicker than the regular 2 µm). The measurement was based on an average value for 81 bumps to compensate for the surface roughness. According simulations were carried out using three different approaches. 3D surface profile diagrams of the measurement and 3 different simulation types are shown in Figure 5.47. A different presentation of the same data sets is given in Figure 5.48 in the form of 2D diagrams. Here, the position of the gold track and bump are marked as dashed lines in the measurement data set.

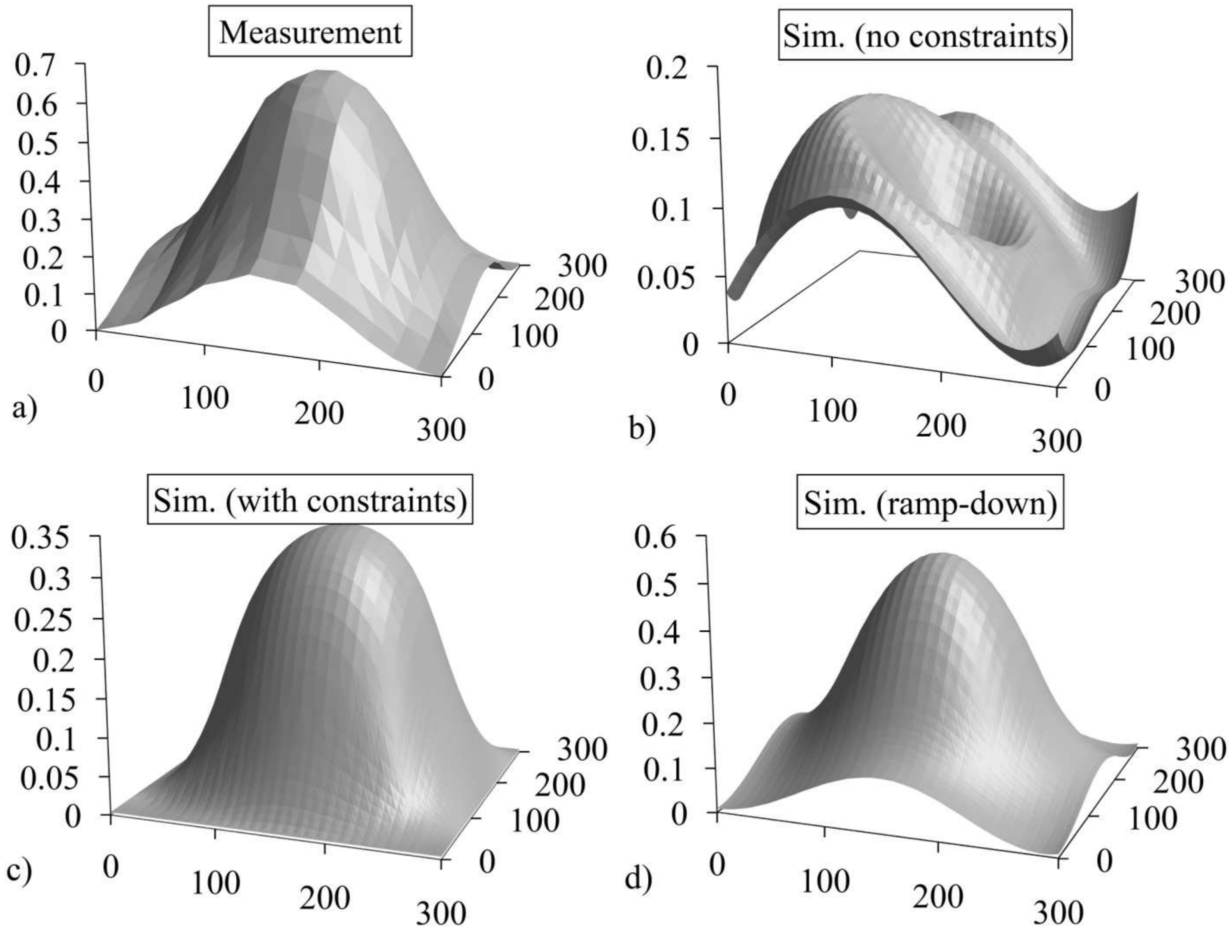

Figure 5.47: 3D diagrams of (a) measured and (b), (c), and (d) simulated local silicon deformations in a reflowed silicon chip with a thickness of approximately 10 µm and a bump diameter of 200 µm. The bumps consisted of approximately 3 µm of tin on top of 5 µm of gold. The bumps were separated from the gold tracks by a 2 µm nickel diffusion barrier and a 5 µm gold base. The axis units are all in µm and were excluded in the diagrams; a) original profiler measurement based on the average of 81 bumps; b) simulation without line constraints, based on a three-step simulation; c) simulation with 4 line constraints, also based on a three-step simulation; d) single step simulation based on a CTE reduction for gold/tin by 30%.

Directly underneath the bump, the silicon surface profile resembled a rotated parable. Outside this core area the shape was more complex: near the sides of the gold track, the deformations were stronger than near the corners of the gold track. This can be explained by the larger distance between the bump edge and the track corners. The deformations at the gold track edges on the side toward the next parallel Daisy chain were larger than for the gold track edges toward the next link of the same Daisy chain. This distortion is visible in Figure 5.48 a) as slightly oval contour lines.

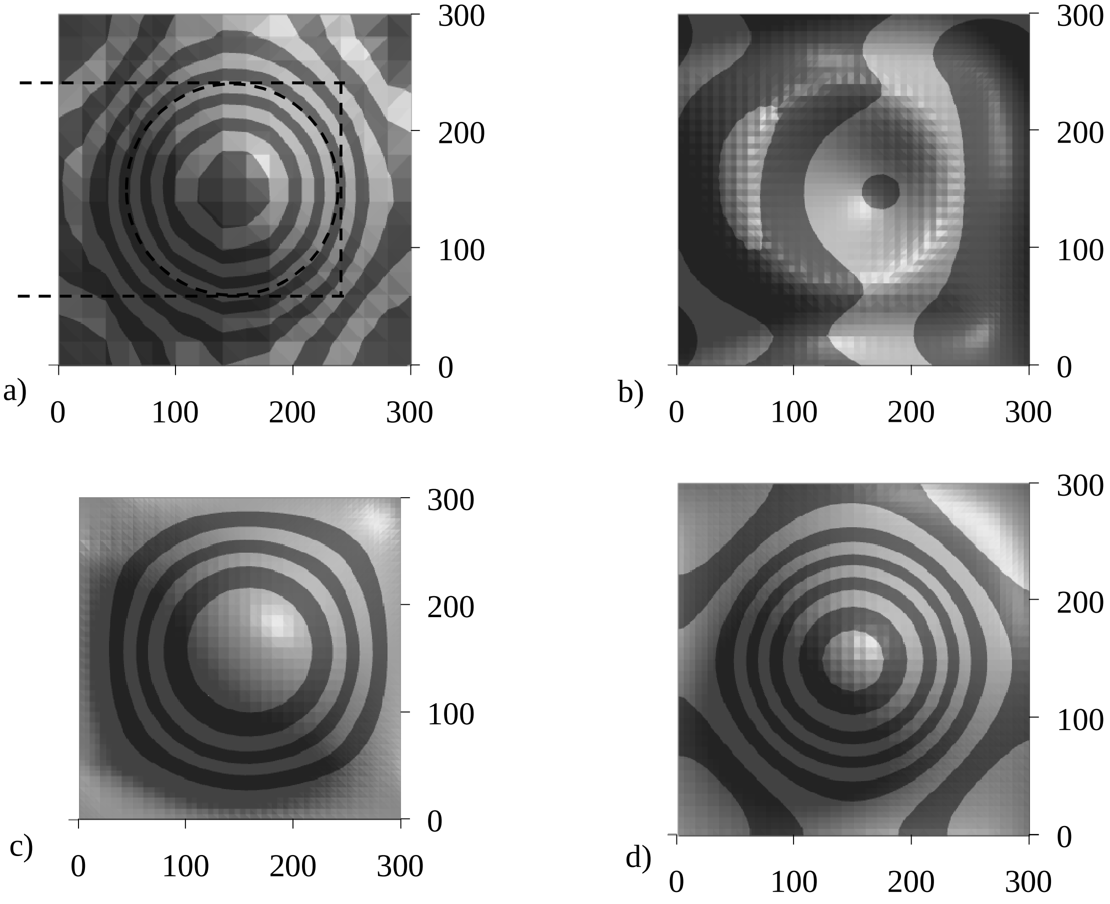

Figure 5.48: 2D diagram of (a) measured and (b), (c), and (d) simulated local silicon deformation in a reflowed silicon chip with a thickness of approximately 10 µm and a bump diameter of 200 µm. The bumps consisted of approximately 3 µm of tin on top of 5 µm of gold. The bumps were separated from the gold tracks by a 2 µm nickel diffusion barrier and a 5 µm gold base. The axis units are all in µm and were excluded in the diagrams. Each ring (dark or light gray) represents a 50 nm height change; a) original profiler measurement based on the average of 81 bumps. The approximate position of the gold track and Au-Ni-Au/Sn bump structures were included as dashed lines; b) simulation without constraints, based on a three-step simulation; c) simulation with 4 line constraints, also based on a three-step simulation; d) single step simulation based on a CTE reduction of gold/tin by 30%.

The simulation was carried out in three different ways. First, the same simulation sequence as in the previous sections was implemented and local deformation data was superimposed for all three simulations (based on 55% of the wafer level plating stress, 68% of the ramp-up stress, and a ramp-down based on a CTE reduction for gold/tin by 10%). The global deformation (i.e. the radius of curvature) was calculated and subtracted from the simulation results for clarity of presentation. The resulting diagrams are shown in Figure 5.47 b) and Figure 5.48 b). The similarity between the measured and simulated diagrams was very poor. The subtraction of the large global deformations from the simulated data probably induced errors resulting in the unusual simulated silicon surface profile with a generally convex shape and local concave indentation underneath the bump. By varying the parameters of the simulation, the indentation could be reduced, but significant deviations between the simulated and measured surface profile remained nonetheless.

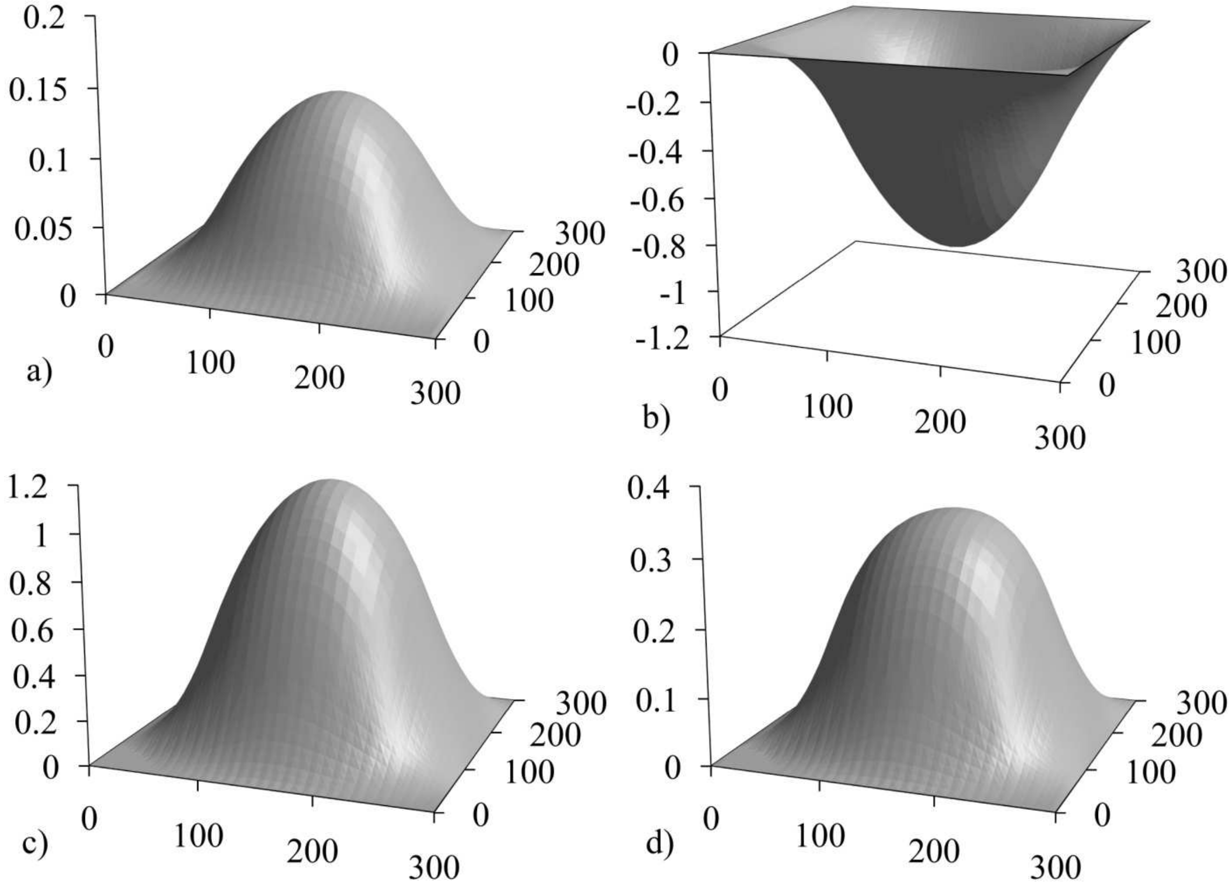

Figure 5.49: Individual and superimposed simulation results for the silicon surface profile during reflow of a chip with 200 µm bumps. The edges of the silicon surface were constrained in the z direction to avoid global deformations. All axis units are in µm; a) simulation of plating stress, based on 55% of the wafer level plating stress; b) simulation of ramp-up, using 68% stress; c) simulation of ramp-down using a CTE reduction for gold/tin by 10%; d) superposition of all three simulations.

To analyze the influence of the correction of the global deformations on local deformation profiles, the three-step simulation sequence was carried out once more with a slight model variation. The four lines making up the edge of the silicon surface were constrained in the z direction. This way, no global deformations could occur, but local deformations were influenced slightly as well. The results of the individual and the superimposed simulations are shown in Figure 5.49. The final result is more similar to the measured surface profile than the simulation without the four partially constrained lines. However, the measured complex shape of the edge could not be reproduced by the simulation this way. Also, the height difference between the outside edge and the tip in the center was about 37% smaller for the simulation with 4 line constraints compared to the measurement.

Finally, a single step simulation was tried based on the ramp-down geometry only. The CTE for gold/tin was reduced to 70% of its standard value and all other CTEs were set to zero. The CTE reduction for gold/tin was necessary to compensate for concave deformations during ramp-up and for the contraction of the silicon, gold track, and gold-nickel UBM during cool-down. The resulting 3D and 2D diagrams are given in Figure 5.47 d) and Figure 5.48 d), respectively. The qualitative agreement between this single step simulation and the measured silicon surface profile was very good. Comparing the height difference between the edge and the center point underneath the bump for the simulation and measurement, a deviation of less than 13% was found. This good agreement between simulation and measurement is also visible in Figure 5.50, where cross sections of the measured and simulated 3D profiles were compared for all three bump diameters. The cross sections were perpendicular to the x-axis and included the point of maximum deformation of the silicon surface. The single step simulation with a CTE reduction for gold/tin by 30% yielded a good agreement between simulation and measurement for 100 µm bumps. For 150 µm bumps the agreement is even slightly better and for 200 µm the similarity is nearly perfect.

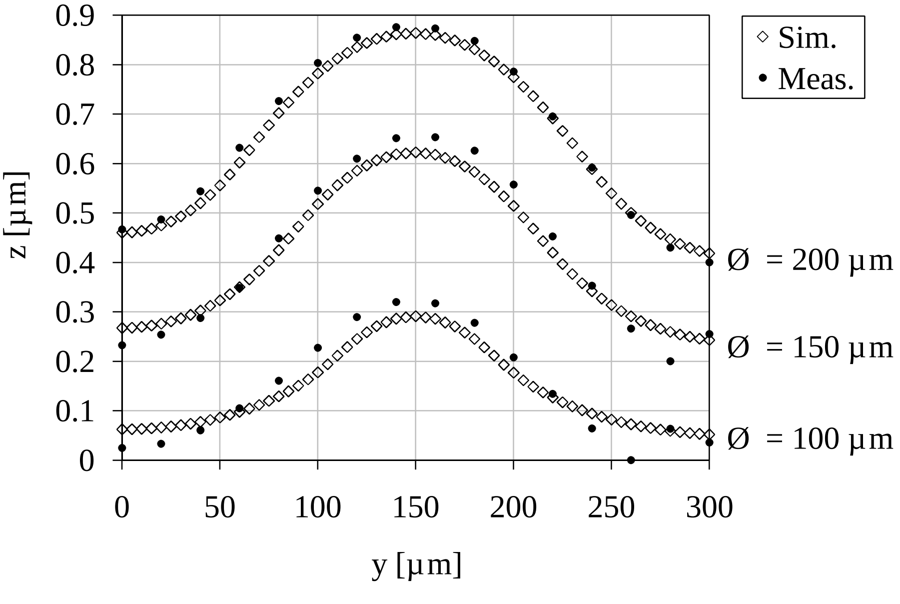

Figure 5.50: Measured and simulated cross sections of silicon deformations in areas with bump diameters of 100 µm, 150 µm, and 200 µm. The cross sections were parallel to the yz plane and ran through the point of maximum silicon deformation. For clarity of presentation, the curves for 150 µm and 200 µm were shifted in the z direction by 0.2 µm and 0.4 µm, respectively.

5.6 Reflow of bumps on a polyimide tape with Kelvin contacts and Daisy chains

As described in Section 3.3 the solder connection between two electroplated gold-tin bumps was nearly void-free and hence mechanically more stable if both bumps were reflowed before soldering. Hence reflow of the bumps on the polyimide tape was necessary for good quality solder connects. In this section the reflow of a single flex tape is analyzed which was fabricated using spin-on polyimide with a thickness of 5 µm. The tape was fabricated with the silicon chip masks (and not the polyimide substrate masks) in order to obtain a polyimide tape with an edge length of 10 mm and gold/tin bumps close to the sample's edge. The results in this section were used to gain a basic understanding of the bending behavior of polyimide tapes during bump reflow.

5.6.1 Bending angle measurement during reflow in air

The bow of the sample was measured in-process using the flip chip bonder camera and is shown in Figure 5.51.

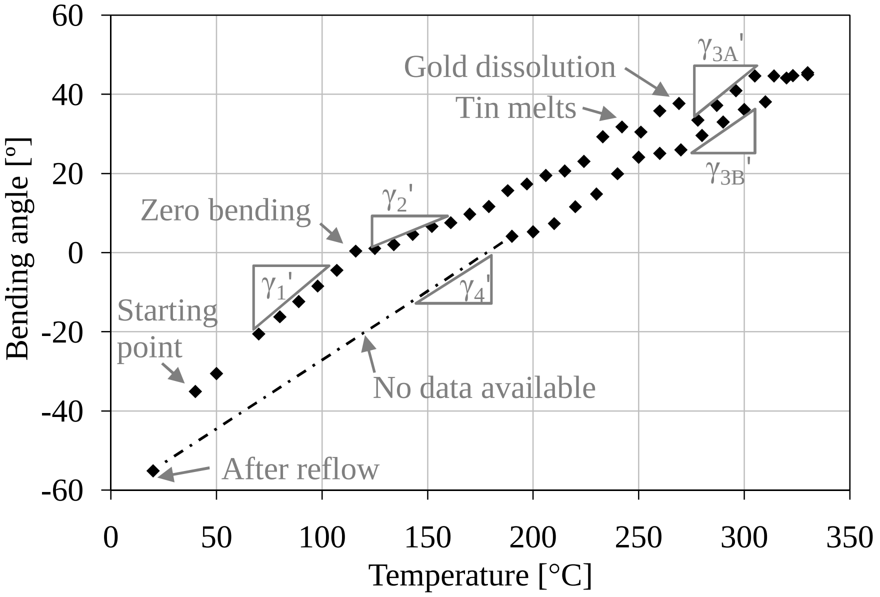

Figure 5.51: In-process bending angle measurement during reflow of a polyimide tape with a polyimide thickness of 5 µm and an edge length of 10 mm. Each data point was obtained by analysis of an image taken with the flip chip bonder camera during the reflow process in air. Notable gradients are labeled with Greek letters γ_1' to γ_4'.

The diagram exhibits many differences compared to the reflow diagram for a thin silicon chip with the same track and bump design (see Figure 5.20). The major differences are:

- strong convex deformation in as-plated samples (possible cause: intrinsic compressive stress in the polyimide tape; for details see Section 5.6.3),
- large gradient decrease at 120°C during ramp-up ($\gamma_1' \approx 2 \cdot \gamma_2'$). This decrease was centered around the temperature of zero bending. A possible explanation were temperature dependent material properties. However, this assumption did not answer the question, why these material properties changed abruptly at 120°C. A more likely explanation were differences in the ability of the sample to deform, depending on the mode of contact between the sample and the heating plate. When the sample's bow changed, the distance between the left and right edge changed as well. Below 120°C, these edges were not in contact with the heating plate and could easily slide together

or move apart. Above 120°C, however, the edges touched the heating plate and the sample's weight rested upon these two edges. Therefore, deformations above 120°C required sliding of the left and right edge over the heating plate surface, resulting in a slow-down of deformations. For an illustration of this effect see Figure 5.53,

- momentary almost horizontal gradient around the point of zero bending during ramp-up. This may be explained by two effects:

  1.) at zero bending the whole chip was lying flat on the heating plate. This possibly caused significant sticking (also called *stiction*), retarding the lifting of the sample's center and hence the bending angle change,

  2.) below 120°C, the sample's weight supported bending, while above 120°C the sample's weight slowed down bending. At 120°C, this effect might have been observed as a bending angle gradient of zero.

  Unfortunately, bending data for the point of zero bending could not be measured during ramp-down. Independent of the nature of the cause, a similar bow retardation was expected during ramp-down,

- the tin melting is clearly visible as a step (and not just a gradient reduction as in the silicon sample) in the reflow diagram between 240°C and 250°C. The delay relative to the melting point of tin of 232°C was similar to the silicon sample,

- the gold dissolution started very early at 280°C, but the resulting bending angle step was relatively small. However, at 305°C the bending angle gradient changed abruptly, possibly indicating further gold dissolution. This retardation was probably caused by local temperature differences, resulting from the low thermal conductivity of polyimide and the different thermal mass density in the 100°C, 150°C, and 200 µm bump areas. Another possible explanation is that the gold dissolution accelerated again upon reaching the peritectic point of gold/tin at 309°C (see the phase diagram of gold/tin in Figure 2.1),

- the polyimide tape was mechanically bistable after reflow with 2 possible bending states (predominant bending around the x-axis or around the y-axis; see Figure 5.54 b) and c) for an illustration). Because the 100 µm bump area deformed faster during ramp-down, the back of the chip (where the 100 µm bumps were located) tried to assume a convex shape, while the front (where the 200 µm bumps are located) was still concavely shaped. As a compromise the back started rolling forward, resulting in global bending around the x-axis after reflow. Images of this observation are shown in Figure 5.52. Due to this bending behavior, no in-process bending data was available during ramp-down from the zero bending point onwards. After reflow, the tape was manually flipped into the second stable bending state and an individual image was taken at 20°C, resulting in this data point being available for the reflow diagram.

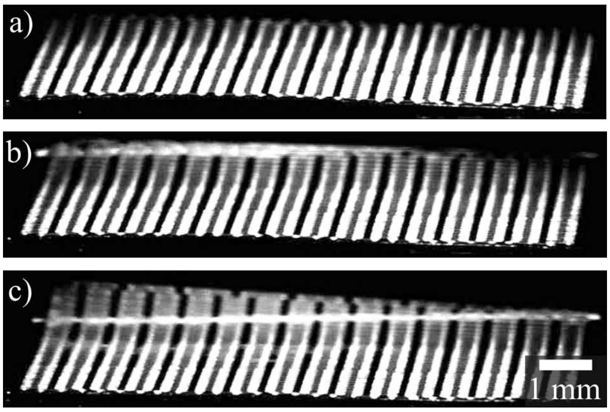

Figure 5.52: Deformation of polyimide tape during ramp-down. The 100 µm bumps were in the back and the 200 µm bumps were in front; a) at 200°C the polyimide tape was flat in the back and concavely shaped in the front; b) at 180°C the back started to roll forward, while the front was still concavely shaped; c) at 160°C the back was more visibly rolling forward, while the front now had reached its zero bending state.

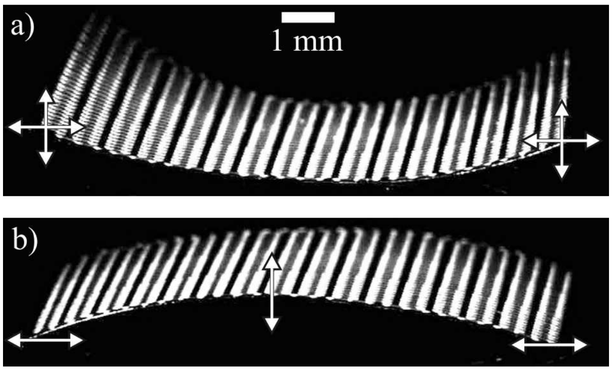

Figure 5.53: Comparison of the influence of bow orientation on deformation changes during reflow; a) convex shape: the edges on the left and right could move freely when bow changes occurred, marked by the crossed arrows; b) concave shape: the edges on the left and right rested on the heating plate (the heating plate is barely visible in the image), carrying the samples weight, and therefore experienced increased mechanical resistance to bending induced sideways movement.

5.6.2 Bending angle differences between bumps with different diameters

Because of its small thickness and soft material properties, the polyimide film showed large bow differences between the areas with different bump diameters. In Figure 5.54 a) and b) this effect is well visible as a much stronger bow in the back of the sample, where the smaller bumps were located. In Table 5.8 bow measurement results for the front and back of the as-plated and reflowed sample are presented.

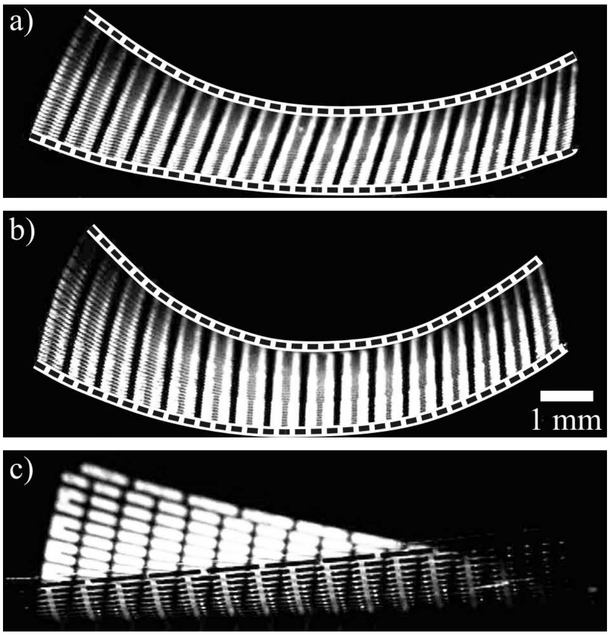

Figure 5.54: Polyimide tape before (a) and after (b), (c) reflow. Because the tape was mechanically bistable after reflow, two images were taken with the same edge in front. The edges of the tape in a) and b) are highlighted with dashed lines for better illustration. The tape had an edge length of 10 mm and was equipped with 1014 gold/tin bumps with underlying Au-Ni UBMs and gold tracks. The 200 µm bumps were in the front and the 100 µm bumpswere in the back.

Table 5.8: Measured bending angles of the polyimide sample in Figure 5.54 before and after reflow. The bending angles were measured in the back part of the chip, where the 100 µm bumps were located, and in the front of the chip, where the 200 µm bumps were located.

Stage	Bending angle [°]	
	at the front edge (Ø = 200 µm)	at the back edge (Ø = 100 µm)
a) Before reflow	-39	-67
b) After reflow	-61	-88

This behavior is in contrast to the observations on reflowed silicon samples, where the 200 µm bumps generally lead to larger chip bow during reflow. However, the inverted behavior in polyimide samples was confirmed by simulation results shown in Figure 5.55. For the bow in as-plated samples and all gradients during reflow it was found that the 100 µm bumps lead to larger bow than the 200 µm bumps. A possible explanation for these simulation results and observations was that the main cause for bending in this case was the bimetal-like stack of polyimide and gold tracks. The bumps merely reduced the bending caused by the polyimide + gold track layer system, because they had a comparatively large stiffness. In a simulation with the geometry consisting of the gold tracks and polyimide only, this physical model was confirmed. The bending angle gradient was found to be 25% larger for the simulation model based on gold tracks and polyimide alone, than for the simulation model with the 100 µm bumps.

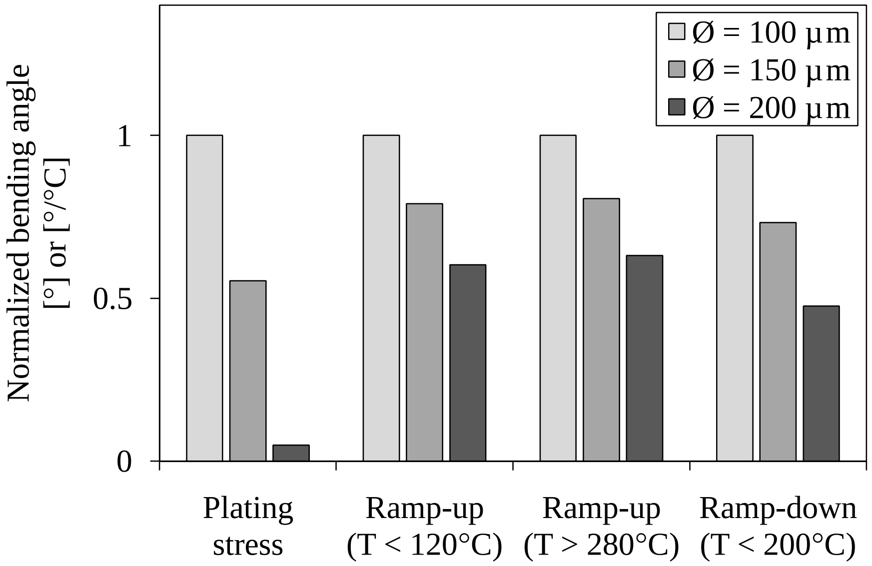

Figure 5.55: Simulation results for the bending angle α (for the plating stress) or the bending angle gradient Δα/ΔT (for ramp-up and ramp-down) of polyimide tapes with gold tracks and Daisy chains. The data is given separately for bump diameters of 100 µm, 150 µm, and 200 µm and was normalized to the 100 µm value for each triplet.

5.6.3 Bending angle simulation in as-plated polyimide sample

In Table 5.9 the measured bending angles of an as-plated polyimide tape are compared to simulations with different intrinsic stress in the polyimide layer.

Table 5.9: Measured and simulated bending angles of an as-plated polyimide tape. The values for 100 µm and 200 µm bumps are given separately and were also averaged. The simulation was carried out without intrinsic stress in the polyimide layer and with an intrinsic stress of -5 MPa (which was a compressive stress). The stress in the metal layers was set to 55% of the wafer level plating stress.

	Measured bending angle [°]	Simulated bending angle [°] with intrinsic polyimide stress of	
		0 MPa	-5 MPa
Ø = 100 µm	-67	-31	-80
Ø = 200 µm	-39	-2	-25
Average	-53	-16.5	-52.5

With a measured bending angle of approximately -39° and -67° for 200 µm and 100 µm bumps, respectively, the as-plated sample was strongly bent. The initial simulation (using 55% of the wafer level stress for the metal layers and no intrinsic stress for polyimide) resulted in a bending angle of -2° and -31° for the 200 µm bumps and 100 µm bumps, respectively. The differences between simulated and measured bending angles in as-plated samples were almost the same in absolute terms for the 100 µm and 200 µm bumps (about -37° and -36°, respectively). Therefore, a compressive stress component in the polyimide could explain these "homogeneous" deviations. A simulation with a stress of -5 MPa in the polyimide tape resulted in bending angles of -25° and -80° for the 200 µm and 100 µm bumps, respectively. The increase, resulting from this additional intrinsic compressive stress in the polyimide was smaller for the 200 µm bumps compared to the 100 µm bumps. The reason for this was that the 200 µm bumps exhibited a larger bending stiffness. Although much weaker than in silicon chips, mechanical coupling between different bump areas was still expected in polyimide tapes and could explain the smaller differences in the measured bows of 200 µm and 100 µm bumps compared to the simulated bows. The average values, however, were almost the same for the simulation and measurement when an intrinsic stress of -5 MPa in the polyimide layer was used in the simulation.

5.6.4 Comparison of bending angle gradients at different stages during reflow

In the reflow diagram in Figure 5.51 bending data of the polyimide edge on the side with the 200 µm bumps is presented. Due to the mentioned low degree of mechanical coupling between areas with different bump diameters in polyimide tapes, according simulations were carried out for 200 µm bumps only. In Table 5.10 several gradients from these simulations and from the measurements in the reflow diagram are compared.

Table 5.10: Measured and simulated bending angle gradients ($\Delta\alpha/\Delta T$) of a polyimide tape with Kelvin contacts and Daisy chains during different parts of the reflow process. The data is based on 200 µm bumps only. The simulations were carried out with the standard value of E_{PI} = 8.5 GPa for the Young's modulus of polyimide and also with a reduced value of 2.0 GPa. For each simulation the deviation with respect to the measured bending angle was also calculated.

Gradient	Metal layer sequence	Measurement [°/°C]	E_{PI} = 8.5 GPa		E_{PI} = 2.0 GPa	
			Sim. [°/°C]	Dev. [%]	Sim. [°/°C]	Dev. [%]
γ_1'	Au-Ni-Au-Sn	0.46	0.61	24	0.50	8
γ_{3A}'	Au-Ni(-Au*)	0.41	0.64	36	0.37	-10
γ_{3B}'	Au-Ni	0.35	0.64	45	0.37	5
γ_4'	Au-Ni-Au/Sn	0.35	0.46	24	0.35	0

*) At this stage the gold layer may or may not be dissolved in the molten tin on top.

The gradients γ_1' (ramp-up between 40°C and 120°C) and γ_4' (ramp-down between 200°C and 20°C) both showed deviations from the simulated values of 24%. These results were achieved with the original set of material parameters from the literature (see Section 4.1). For higher temperatures, where the gold/tin solder was liquid (that is for γ_{3A}' and γ_{3B}'), the deviations were much larger (36% and 45%, respectively). A possible explanation for these deviations is that the polyimide softened at higher temperatures. The polyimide softening at elevated temperatures was simulated by using a value for the Young's modulus for polyimide of $E_{PI} = 2.0$ GPa, which was about four times smaller than the room temperature value of $E_{PI,RT} = 8.5$ GPa. As a result, a bending angle gradient of 0.37 °/°C was found, which was in between the measured values of $\gamma_{3A}' = 0.41$ °/°C and $\gamma_{3B}' = 0.35$ °/°C. Interestingly, a Young's modulus of 2.0 GPa also resulted in a better approximation for γ_1' and γ_4', suggesting that the polyimide had a Young's modulus that was fairly independent of temperature, but by about a factor of 4 smaller than in the supplier's data sheet.

5.7 Soldering of polyimide-silicon-flex-systems

When flip chip soldering silicon chips onto polyimide substrates, misalignment may lead to deformed interconnects with low conductance and in severe cases even to open circuits. Therefore, the misalignment during soldering of thin chips to polyimide tapes is analyzed by FEA in the next section. Then, simulations and measurements of the bending angle of soldered thin chip/polyimide systems are compared. A new FEM model is proposed that is based on a single step simulation with adapted CTEs for gold and nickel.

5.7.1 Thermal misalignment in soldered polyimide-silicon systems

In this section analytical considerations are presented regarding thermal misalignment during soldering. FEM simulations are used to gain a better understanding of how the misalignment of thin silicon/flex tape systems are influenced by the wiring and UBM dimensions.

Although the polyimide and silicon bulk materials had almost the same coefficient of thermal expansion, significant thermal misalignment could occur during soldering, caused by the presence of metal layers (gold tracks and gold-nickel UBMs). In chips and substrates, where the bulk material dominated, the mechanical behavior, i.e. the elongation ΔL during soldering could be calculated using

$$\Delta L = \Delta T \cdot CTE \cdot L\,, \tag{5.11}$$

where ΔT is the temperature difference (i.e. present temperature minus room temperature), CTE the coefficient of thermal expansion of the bulk material, and L the original length. If the CTEs of the bulk materials differed, the expansions for the chip and substrate were not the same, resulting in the misalignment of bumps at elevated temperatures. The misalignment was probably largest at the peak soldering temperature and then reduced again during cooling. However, at the eutectic temperature the solder solidified, locking the misalignment. Disregarding the influence of the metal structures (which were fabricated on the silicon chip and polyimide substrate) on the thermal expansion, the thermal misalignment Δm between two bumps was calculated by replacing the bulk CTE in formula (5.11) with the CTE difference of the two bulk materials:

$$\Delta \mathrm{m} = \Delta T \cdot (CTE_{Si} - CTE_{PI}) \cdot \mathrm{L}\,, \tag{5.12}$$

where $\mathrm{CTE_{Si}}$ and $\mathrm{CTE_{PI}}$ are the coefficients of thermal expansion of silicon and polyimide, respectively, and L the distance of the two bumps to the point of perfect alignment.

During soldering, misaligned bumps exhibit a larger solder surface area, resulting in an energetically unfavorable state. Therefore, the molten solder's surface tension counteracts the misalignment and tries to reduce the total solder surface area of all interconnects. Generally, this is achieved by good alignment in the center of the bump pattern and a radially increasing misalignment proportional to the distance to the center of the bump pattern [184]. If the CTE difference between the two bulk materials and the soldering temperature are known, different bump pitches may be used for the chip and substrate such that this designed misalignment is compensated by the thermal misalignment at the eutectic temperature.

Disregarding the metal layers on the silicon chip and polyimide substrate in this thesis, the maximum misalignment was obtained by substituting L in formula (5.12) with the distance between the chip center and a bump at the chip corner. This distance was approximately 6.4 mm, resulting in a theoretical maximum thermal misalignment of 0.2 µm, which was fairly small and did not pose a problem. However, the gold tracks and gold-nickel UBMs had comparatively large CTEs and thicknesses of the order of magnitude of the bulk materials and therefore had a significant impact on expansions and hence misalignments. In Figure 5.56 simulation results for the thermal strain of silicon chips and polyimide substrates during ramp-up are shown. The simulation was based on Daisy chains with a bump pitch of 300 µm and identical bump diameters and gold track widths. The model was derived from the previous model (for the reflow of silicon chips and polyimide tapes with Daisy chains) by adding z direction constraints, which prevented bending, but not the expansion within the xy plane. Apart from the actual polyimide thickness of 5 µm, a thickness of 7.5 µm was also simulated to account for the influence of the polyimide periphery on misalignments. Strain results were extracted from the model for different directions in the xy plane. Because the results were very similar, only the x-direction strain is given in Figure 5.56. For the smallest simulated track width of only 5 µm the strain for the silicon chip and polyimide tape was almost identical, resulting, if any, in a very small misalignment. With increasing track width, however, the thin silicon samples and the polyimide were extensively strained, while thicker silicon samples were moderately elongated. The actual misalignment between two bumps could be calculated from the simulation results by first calculating the strain difference for the silicon chip and polyimide tape and then multiplying this difference with the distance between the chip center and the bump pair of interest.

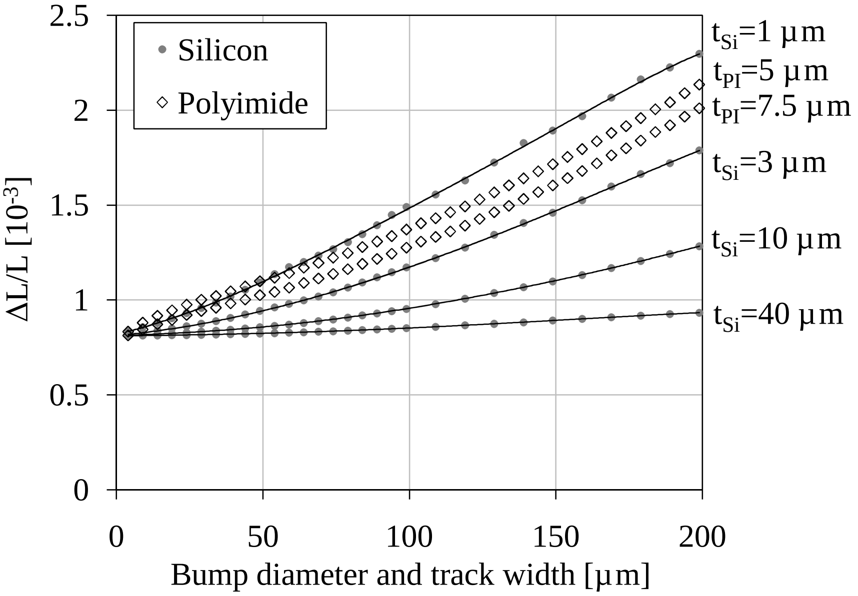

Figure 5.56: Simulated silicon chip and polyimide tape strain $\varepsilon = \Delta L/L$ *at 280°C as a function of bump diameter and track width. The bump diameter and track width were equal within each simulation and were varied between 5 µm and 200 µm from simulation to simulation. The silicon thickness was either 1 µm, 3 µm, 10 µm, or 40 µm and the polyimide thickness was either 5 µm or 7.5 µm. For better comparison polynomial regression curves were added for the silicon chip strain.*

According to the simulation results in Figure 5.56, the maximum misalignment was expected in the 200 µm bump area of soldered samples with a large silicon thickness. An upper limit for this misalignment was calculated as $\Delta\varepsilon_{max} \cdot L_{max} = 1.20 \cdot 10^{-3} \cdot 6155$ µm = 7.40 µm. Although much larger than the misalignment of 0.2 µm calculated from the bulk material properties only, the upper limit was still fairly small compared to the bump dimensions and was in the range of the accuracy of the prototyping masks used for sample fabrication. Moreover, the actual maximum misalignment was expected to be even slightly smaller, because only one third of the bumps had a large diameter of 200 µm and because the polyimide strips on the periphery of the soldered areas justified an increase of the polyimide thickness in the simulation to 7.5 µm. This stiffened the polyimide, reducing its elongation to values more similar to the silicon sample.

In the previous section it was found that the polyimide properties were more accurately described by a smaller Young's modulus (i.e. 2 GPa instead of 8.5 GPa). The stiffness reduction for polyimide increased its expansion under the thermal strain of the metal tracks and UBMs. This effect, however, was fairly small, leading to a maximum expansion of 8.73 µm between the center and corner bumps (compared to 7.40 µm for the regular Young's modulus of polyimide).

The analysis also showed that, by varying the silicon and/or polyimide thickness, the expansion and hence the misalignment could be influenced to a certain degree. For very thin chips (around t_{Si} = 2 µm) the misalignment relative to the polyimide substrate was very small and could ideally reach a value of zero.

5.7.2 Measurements and simulations of soldered polyimide-silicon systems

For the simulation of deformations in soldered polyimide-silicon flex systems the silicon thicknesses of 4 thin chips were determined using optical measurement techniques and used as input for the simulation model. Generally, the soldering process required individual simulations for the ramp-up and ramp-down part of the soldering cycle. The gold wires and nickel-gold UBMs have essentially inverted influences on the bow during ramp-up and ramp-down. Therefore, the two-step simulation was replaced by a single simulation with reduced gold and nickel CTEs [18]. The results of the bow measurements on 4 soldered samples and the according simulations based on a CTE reduction for gold and nickel by a factor of 1.7 are shown in Figure 5.57. Qualitatively, the results were very similar. The bow differences between the simulations of different bump diameters were larger than in the bow measurements of areas with different bump diameters within a single chip. This was expected and was due to the mechanical coupling of neighboring areas on the same chip.

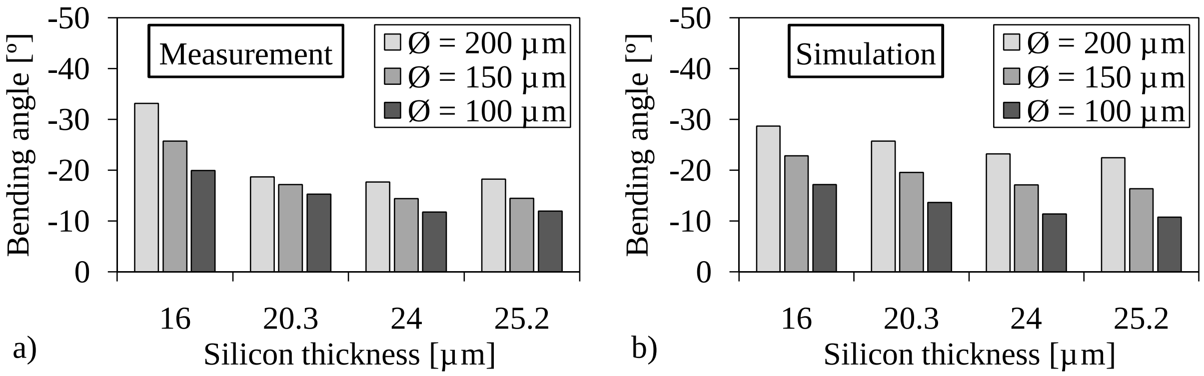

Figure 5.57: a) Measured bending angle in 4 soldered silicon-polyimide systems (sample group C2) with different silicon thicknesses. Each area with a predominant bump diameter was measured separately; b) according simulation results for a simplified one-step model with reduced gold and nickel CTEs.

Therefore, a quantitative comparison between simulation and measurement required the average bow per chip to be compared to the average bow in the simulations for all 3 bump diameters. These averaged results are presented in Figure 5.58.

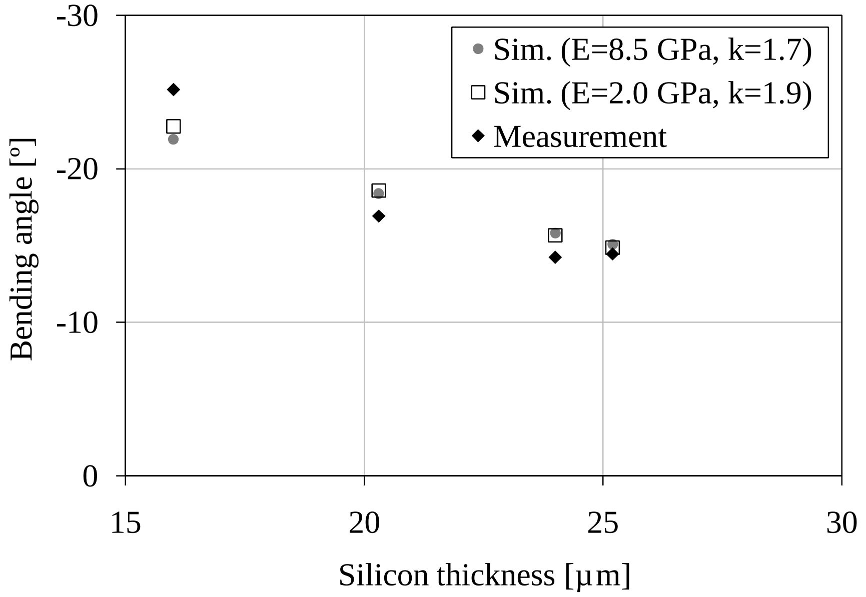

Figure 5.58: Comparison of measured and simulated deformations in flexible silicon soldered to polyimide tapes. For each data point three measurements (one for each 100 µm, 150 µm, and 200 µm bump diameter area) or three simulations (also one for each bump diameter) were averaged. The simulations were carried out for a Young's modulus for polyimide of 8.5 GPa and a CTE reduction for gold and nickel by k = 1.7 and for a Young's modulus of 2.0 GPa and a CTE reduction by k = 1.9.

The simulations were carried out for two sets of material parameters. For the standard Young's modulus for polyimide of E_{PI} = 8.5 GPa and a CTE reduction for gold and nickel by a factor of k = 1.7 the simulated and measured values showed a maximum deviation of 15%. For a value of E_{PI} = 2.0 GPa (see Section 5.6.4) and k = 1.9 the agreement was improved and a maximum deviation of 10% was found between the measured and simulated bending angles. This result supported the findings in Section 5.6.4, where a value of E_{PI} = 2.0 GPa yielded the best fit between simulation and measurement. However, it must be noted that the optimal values for k in this section were determined experimentally and that these may have caused or influenced the good agreement between simulation and measurement for E_{PI} = 2.0 GPa.

5.8 Summary

The comparison of measured and simulated deformations in thin silicon chips and polyimide tapes during reflow was at the heart of Chapter 5. To reduce simulation time, the FEM model size was reduced to a small section of each reflowed sample. Different types of deformation data were extracted from the simulation results and measurement data. Among these data were global bending radii (often converted to bending angles) and local silicon deformations around bumps or perpendicular to solder lines. In all analyses a profound dependence of deformations on the silicon thickness was found.

Based on the measured bow of thin silicon wafers that were plated either with gold, nickel, or tin, the intrinsic stress in the plated layers was determined and used as input for the simulations.

Special thin dummy chips with gold/tin solder lines on top of a gold-nickel base with varying width were fabricated. The chips had a silicon thickness between 6 µm and 45 µm and their bow was observed in-process by analysis of digital images taken during reflow. Based on these temperature dependent data and global deformation data collected after reflow, a mechanical FEM model for the reflow process of gold/tin solder bumps on thin silicon chips was established, generally showing good agreement with the measured deformations. The FEA consisted of three independent simulations for the deformations due to plating stress, the ramp-up, and the ramp-down. To obtain the actual deformation after reflow, the resulting three sets of simulation data were superimposed. The best agreement between simulation and measurement was achieved when the plating stress was set to 55% of the wafer level plating stress, and all CTEs were set to 60% of the literature values during ramp-up, and the CTE of the gold/tin solder was reduced to 90% during ramp-down. Without further adjustments, the model was also capable of predicting local deformations in thin chips with gold/tin lines with good accuracy.

Furthermore, it was found that the needle tip fixation that was used to prevent the light chips from being blown away by gas flows in the reflow oven, only had a small influence on the bow after reflow (of the order of 9% for the analyzed chip).

Next, the model was used to predict deformations during reflow of thin chips with gold tracks and gold-tin stacks on top of a gold-nickel base. The structures could be used to fabricate Daisy chains and Kelvin contacts. Because different bump diameters were present within each chip, the bending angle was found to vary across the chip. Similar (though slightly larger) differences were also found in the simulations. The measured bending differences for different bump diameters were smaller than in the simulations, because the areas on a chip with different bump diameters were mechanically coupled, whereas the simulations were carried out for completely decoupled models. By averaging deformation data in the simulation and measurement, good agreement could again be reached. Further improvement of the agreement between simulation and measurement was achieved by slightly increasing the CTEs during ramp-up to 68% of the literature values (instead of 60% in the previous simulations).

It was also found that the predominant bending axis depended on the silicon thickness. For thick chips the anisotropy of silicon dominated the bending behavior, leading to a bending axis tending to the <100> direction, while for thinner chip the metal structures dominated, leading to a bending axis tending to the <110> direction. Without further adjustments, the FEM model predicted the silicon thickness for which the transition from bulk-silicon-dominated to metal-dominated bending behavior was observed.

Next, different simulation strategies were tested to reproduce measured local deformation data for the silicon underneath reflowed bumps. A single-step simulation with the CTE of gold/tin reduced to 70% and all other CTEs set to zero resulted in very good qualitative and quantitative agreement between simulation and measurement, whereas the three-step simulation strategy did not work well. This was partly attributed to the fact that the local deformation data needed to be extracted from the total deformation, which was dominated by the global bow.

The reflow of a polyimide tape with the same metal structures as the silicon chips in the previous paragraph was analyzed and simulated. It was found that the mechanical coupling between different areas within the same tape was less strong. Moreover, the resulting bow after reflow was smaller for larger diameter bumps. This was explained by the good flexibility of the tape and gold tracks, which was reduced by the (comparatively) stiff bumps. It was also found that the as-plated sample was more strongly bent than predicted by the simulation. Assuming an intrinsic stress in the polyimide of -5 MPa resulted in a much better agreement between simulation and measurement. Furthermore, by reducing the Young's modulus of polyimide to about one quarter of the value in the manufacturer data sheet, the in-process bending angle gradients at higher temperatures could be much better approximated in the simulation.

Misalignment during flip chip soldering was also simulated. Although the polyimide substrate and the silicon chip had almost identical CTEs and were patterned with similar metal structures, a theoretical misalignment of up to 9 µm was found. This misalignment was caused by the expanding metal structures during ramp-up, which could stretch the soft polyimide much easier than the "stiff" silicon. The simulation results suggested, that by reducing the silicon thickness to about 2 µm, it should be possible to achieve similar stretching behavior for the silicon chip as for the polyimide tape, thus reducing misalignment.

Finally, the bending behavior of soldered systems was analyzed. 4 chips were soldered to polyimide tapes and a single-step simulation was carried out to reproduce the measured deformations. It was found that by using the literature value for the gold/tin CTE and reducing the CTEs for gold and nickel by a factor of 1.7, a good agreement between measured and simulated bending angles could be achieved.

5.9 Discussion

When different materials are joined by soldering, thermal stress develops during cooling from the eutectic to room temperature. Additionally, temperature changes occurring when a device is switched on or off may represent severe mechanical loads. The resulting stress is a major cause for failure. Therefore, many FEM simulations found in the literature focus on reliability aspects, with stress distribution, local stress maxima, and their correlation to the respective failure mode being at the center of the investigation. An example is given in [119], where stress concentration effects at solder voids were simulated. In very few publications have FEM analyses of gold/tin flip chip interconnects been presented, because, historically, the lead/tin system was mostly chosen for such analyses.

In flip chip soldered systems with standard substrate and chip thicknesses, the thermal mismatch between substrate and chip is partly compensated by the comparatively soft solder joints. The solder joints serve as a mechanical buffer between the joined parts, thereby often being exposed to severe deformations and/or stress. The load on the joints may be reduced by different measures:

- minimizing CTE differences between the joined parts,
- reducing the soldering temperature,
- optimizing the bump geometry,
- reducing the maximum lateral distances between two solder joints, or
- by applying an underfill, which fills the empty space between the bumps.

The underfill has to be chosen to exhibit desired material properties (that is a CTE similar to the solder material) and, if necessary, its properties may be adapted, for example by adding small SiO_2 particles to decrease the CTE. If such changes are implemented, this may, however, not only change the averaged material properties of the underfill, but also require an adapted (for example two-layered) geometry for the simulation model due to particle settling [54]. Apart from these geometric considerations, adapted material models may improve simulation results. For example, a soft solder like lead/tin may be more realistically represented by a material model with rate dependent plasticity, while copper tracks require an elastic-plastic material model, and polymers a viscoelastic material model. In this thesis, simplified linear material models were used. More complex material models may, however, be developed using the results in this thesis as additional input.

Often isotropic silicon material models are used in the literature to simplify the modeling process. In the case of silicon the resulting error is usually around a few percent [118]. However, in this thesis the anisotropic mechanical properties of silicon were found not only to influence the degree of bending, but also the direction of the bending axis and could therefore not be neglected. For the first time it was shown, that the direction of bending during bump reflow systematically varies with silicon thickness. For thinner silicon the metal tracks, UBM, and solder defined the orientation of the bending axis, while for thicker silicon the anisotropy of the bulk silicon dominated. In between these extremes, a continuous transition from metal-structure-dominated to bulk-silicon-dominated bending direction was observed.

Anisotropic material models may also be necessary for other materials than silicon. In [121], for example, the maximum strain in solder interconnects was found to be five times larger than suggested by an FEM simulation. The cause for this deviation was found to lie in the anisotropic properties of the polymer underfill that was used. Therefore, taking anisotropic properties of the polyimide films into account in this thesis may also lead to an improved simulation model. For lack of anisotropic material data for the polyimide used in this thesis, this was not implemented.

In contrast to most investigations found in the literature, the focus in this thesis lay on deformations during reflow and soldering (as opposed to maximum stress intensities). In order to obtain soldered systems with good flexibility, the joining partners were designed to be flexible. For the polyimide substrate with its naturally small Young's modulus and a spin-on thickness of only 5 µm the obtained flexibility was sufficient. The silicon chip, however, with its brittle properties and large Young's modulus was inherently inflexible. By reducing the thickness of the bulk silicon and avoiding crystal damage in the thinned chips, good flexibility was also achieved, allowing large deformations to occur and to be analyzed during reflow and soldering. By using flexible joining partners, the stress partially translated into deformations. Although these deformations are generally beneficial with regard to reliability (because they reduce maximum stress levels), they exhibit one important disadvantage: handling during the fabrication process has to be adapted to the new geometry. Also, the housing or fixation of the finished product may be more difficult to realize, when deformations are present. Therefore, the ability to predict deformations during processing, and especially during reflow and soldering may facilitate handling and packaging. Moreover, once the model for predicting deformations is established, it can also be used to optimize the sample design with respect to minimum (or, if required, predefined) deformations. Therefore, the analysis in this thesis goes beyond the approach of other research groups working with thin silicon chips, who mainly focus on stress or sometimes thermal dissipation.

One such example found in the literature is the FEM simulation of stress resulting from the gold/tin soldering process and thermal cycling of a silicon chip with a thickness of 150 µm that was flip chip attached to a GaAs chip with a thickness of 100 µm [114]. The system size with an edge length of 10 mm was similar to the sample size in this thesis, whereas the chip and substrate thicknesses were much larger. Also, the solder joints had a different geometry with a comparatively small height and diameter of about 35 µm. The solder joints were supported by a gold base on each side, but were missing a gold-nickel diffusion barrier. Different failure modes like fatigue or UBM delamination were investigated in the simulations. Because of the large length-to-thickness ratio of 1,000,000 : 1 (system size : minimum layer thickness) the FEM model also required a very large number of elements. To reduce the element count a so-called global/local approach was implemented. Therefore, the model was defined by two geometries: one for the individual solder joint and one for the arrangement of the solder joints. Although this approach leads to a more realistic model, it was shown in this thesis that simulations based on a small section of the sample deliver good results with regard to deformations. Similar to this thesis, it was found in [114], that intrinsic stress in electroplated layers has to be taken into account to obtain correct simulation results.

In certain cases a 2D model is sufficient to represent a 3D thin silicon sample. For example, in [117] the force required to bend silicon chips with a size of 1 x 5 mm² and a minimum thickness of 48 µm using a 3-point bending unit was measured and simulated. The deviation between simulation and measurement increases with the degree of bending but stays below about 5% for the considered maximum displacement of 600 µm. Another example for a successful 2D simulation of a 3D sample is the heat dissipation study of a thin chip (t_{Si} = 50 µm) integrated into a printed circuit board by resin moulding, resulting in a total system thickness of about 300 µm [115]. The thermal dissipation through heat vias was simulated using an axisymmetric model. Because the sample was not fully axisymmetric, material properties were averaged in some areas. This method allowed for faster simulations at the expense of accuracy. For more complex simulations like package deformations or stress concentration at the thermal vias, however, a 3D model (based on the simulation tool PATRAN) was preferred. Although no quantitative deformation data is given in [115], the package deformation could be estimated from a graph depicting the deformed state after cool-down to be around 11 µm for a chip size of 2.54 mm. This is about two orders of magnitude less than typical deformations observed in this thesis. Due to the fast relaxation of the plastic components at higher temperatures the samples were regarded as stress-free above approximately 120°C. This approach is similar to the assumption in this thesis that the deformations above the eutectic temperature of gold/tin are only temporary in nature and therefore can be neglected in the simulation. In order to reduce simulation time, parametric modeling was used in [115]. This technique allows to change parameters (for example material properties) without remeshing. Because the meshing in this thesis typically took less than 1% to 5% of the total simulation time, no significant time savings would have resulted from parametric modeling.

Of the few sources found in the literature, that investigated deformations in thin chips integrated into flex systems by FEM simulations, the analysis by Shin et al. [118] comes closest to the work in this thesis. A wireless blood pressure sensor chip with a silicon thickness of 30 µm was attached face-up to a polyimide substrate and contacted by electroplating wires over the chip and substrate. A second polyimide tape was added as encapsulation to realize a fully implantable device. Similar to this thesis polyimide PI2611 and electroplated gold were used among other biocompatible materials. The simulations, however, were carried out differently: a bending radius between 5 mm and 50 mm was forced onto the flex system, and the resulting maximum von Mises stress in the silicon was extracted from the simulation data in order to determine the minimum bending radius before breakage of the chip. It was found that the maximum von Mises stress is inversely proportional to the forced radius of curvature. For a radius of curvature of 10 mm a maximum von Mises stress of 0.6 GPa was calculated, which is about half the strength of silicon. In a different analysis the minimum radius of curvature just before breakage for chips with thicknesses of 25 µm, 40 µm, 80 µm, and 100 µm was measured and simulated For the simulations a silicon strength of 1 GPa was assumed. The results are shown in Figure 5.59. A quantitative comparison with results in this thesis is difficult, because maximum stress levels and fracture behavior were not analyzed in this thesis. However, several observations can be noted:

- The deviation between simulation and measurement is larger than typically in this thesis and is attributed to residual stress (which Shin et al. use as a synonym for surface damage) from the thinning process. Although the thinning process is not described in detail, the authors mention a wet etching step, that is used to remove possible surface damage. This is a similar strategy as in this thesis. So, surface damage may also be present in thin chips in this thesis directly after the final wet etching step.

- Only four simulation results are given in [118] (see Figure 5.59). These show a large spread of about +-10% around an exponential and linear regression curve. Unfortunately, Shin et al. do not give an explanation for these fluctuations, which are much larger than for simulations in this thesis. Possibly an optimization of the meshing procedure could have reduced these apparently random fluctuations and yielded a more precise correlation between the silicon thickness and minimum radius of curvature.

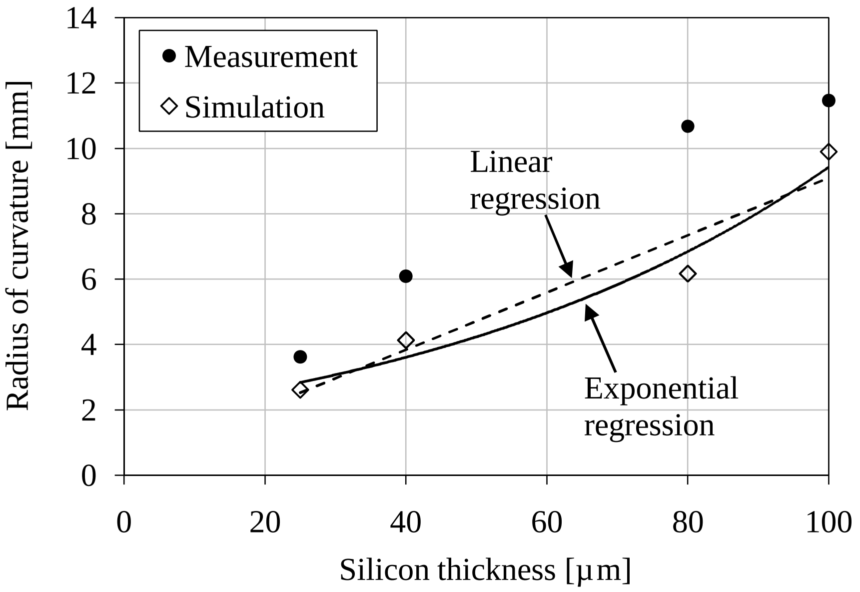

Figure 5.59: Simulated and measured radius of curvature at the breaking point of thin silicon chips as a function of silicon thickness (modified from [118]). The simulations were carried out for an assumed tensile strength of silicon of 1 GPa. An exponential and a linear regression curve were added for the simulation data.

The effect of stress reduction in stressed layers due to deformations resulting from this stress was systematically analyzed in [116]. The relative stress reduction was inversely proportional to the silicon thickness and could lead to a remaining stress of only 27% compared to the initial stress for the thinnest chip analyzed ($t_{Si} = 4$ µm, stressed layer thickness = 1 µm, initial stress level of 500 MPa in the undeformed state). Similar characteristics should be present in the metal layers in this thesis. These were, however, not investigated. Apart from the deformation-based stress reduction effect, the dependence of the radius of curvature on the silicon thickness for a predefined stressed layer (stressed layer thickness = 1 µm, initial stress = 500 MPa) was also analyzed in [116]. A linear dependence was found, which is much simpler than the correlation found in this thesis, which shows a decreasing radius of curvature for smaller silicon thickness, and then, for extremely thin silicon an increasing (or even inverted) radius of curvature. This was attributed to the superposition of a sequence of deformations during the reflow process, which are based on different geometries and naturally are oriented in opposite directions because one is caused by the rising temperature during ramp-up and the other by the falling temperature during ramp-down.

The different sections of the reflow process were considered separately in the FEM simulations and combined results were generated by superposition. With this approach the model could be optimized independently for various temperature ranges. Silicon is a well known and understood material, whereas electroplated materials often exhibit a large variation of properties, depending on the exact deposition parameters. Also these properties usually change with temperature. The temperature dependent deformation data generated in this thesis is well suitable to serve as input for improved material models. However, because material models were not at the focus of the investigation, the observed dependence of deformations on temperature were not translated into new material models, but rather incorporated into the model by a simple adjustment of the CTEs and/or Young's moduli. A good demonstration of the capabilities of the temperature dependent in-process bow measurements in this thesis with regard to the analysis of material properties is the confirmation of the known transition point for gold/tin solder at about 190°C.

With very few exceptions soldering of ultrathin silicon chips has not been carried out before. This thesis represents the first systematic analysis of reflow and soldering induced deformations in individual ultrathin chips or ultrathin-chip-on-flex-polyimide systems.

In this thesis, it was observed that deformations from ramp-up were partly compensated (and in some cases even overcompensated) during ramp-down. The amount of bending per degree Celsius temperature change (referred to as the *bending angle gradient*) depended on the geometry of the metal structures (which was different for the ramp-up and the ramp-down simulations) and on the bulk silicon thickness. It was found, that by adapting the silicon thickness it should be possible to obtain samples which deform similarly during ramp-up and ramp-down, essentially yielding chips that are nearly undeformed after ramp-down. To illustrate these findings, the simulated bending angle gradients for the ramp-up and ramp-down geometries of chips with gold/tin lines are compared in Figure 5.60. For a silicon thickness above 9 µm the gradient for Au-Ni-Au/Sn (that is for the ramp-down) is larger, leading to a convex bow after ramp-down. For a silicon thickness below 9 µm, however, the situation is different: the Au-Ni structures have a larger bending angle gradient. This means that samples with $t_{Si} < 9\ \mu m$ deform more during ramp-up than during ramp-down, resulting in a concave deformation after ramp-down. This effect of silicon thickness dependent bow orientation after ramp-down was also observed in reflowed samples (see Figure 5.11 for an example of a chip with a local silicon thickness of 2 µm, having a locally concave bow after reflow). In another sample with $t_{Si} \approx 8\ \mu m$ the bow was nearly zero after ramp-down, as predicted by the simulation. The observed effect is mostly of geometric nature. This can be seen in Figure 5.61, where simulation results are presented for the bending angle gradient of thin chips with gold lines, having the same thicknesses as in the previous simulation. The curves qualitatively look the same as for the Au-Ni vs. Au-Ni-Au/Sn simulation in Figure 5.60, even though the metal lines on the thin silicon chips were made only of gold.

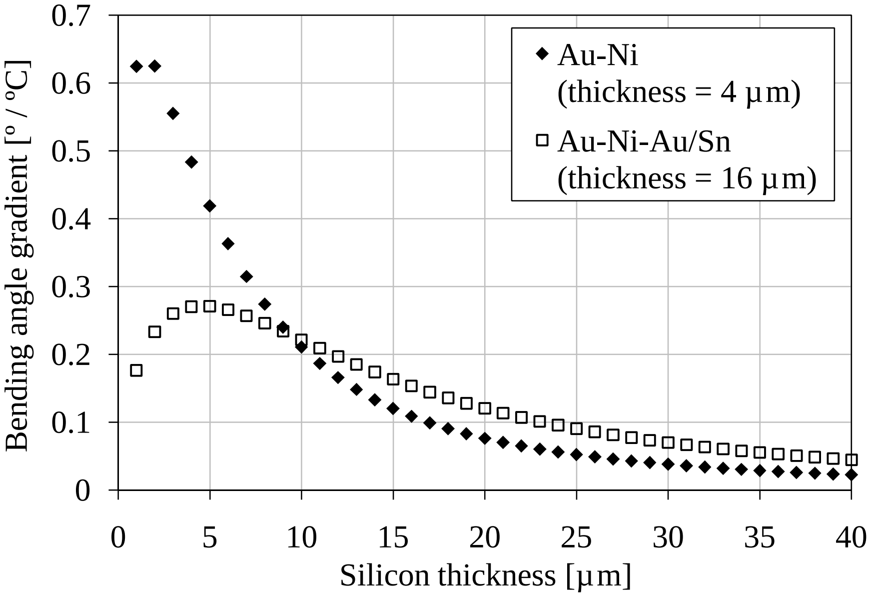

Figure 5.60: Simulation results for the bending angle gradient for 2 different metal layer sequences in thin chips with 100 µm wide gold/tin lines.

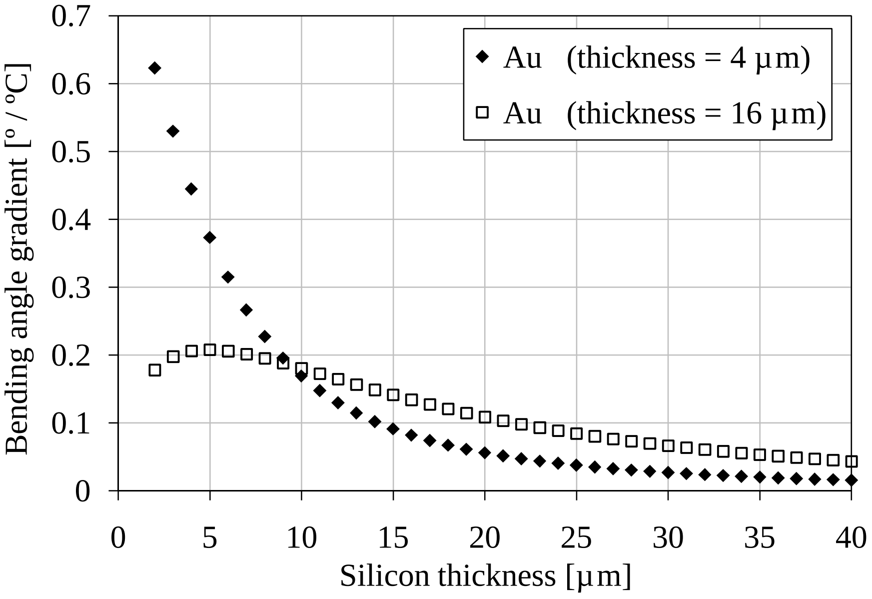

Figure 5.61: Simulation results for the bending angle gradient for 2 different gold layer thicknesses in thin chips with 100 µm wide gold lines as a function of silicon thickness.

The simulation model adaptations in this thesis were mainly necessary because of imperfect material parameters that were taken from the literature and not determined experimentally for the fabricated layers. In thin film applications the material properties not only depend on the deposition parameters, but also on the type of machine that was used for sample fabrication. The same holds true for electroplated layers, which is why material properties are preferably determined from special samples deposited with the material of interest. The temperature dependence of these properties is of special importance and is well known for many bulk materials. Electroplated layers, however, often show different behavior, like crystal reorganization during heating, resulting in a more complex deformation behavior.

Apart from these material-based influences, there are also many technological and design-related factors that have an impact on the amount and exact shape of deformations. These are often difficult to control and may be responsible for some of the observed differences between simulations and measurements. Major processing factors with an influence on chip bow are:

- silicon thickness distribution (planarity per chip is usually around ± 1 µm after the lapping process, but typically is increased to around ± 3 µm after wet etching),
- defect density on the silicon surfaces (i.e. on the chip backside and on the chip edge; see [185] for details on defect-based stress in mechanically thinned silicon),
- solder thickness and composition variation due to the electroplating process (caused by the inhomogeneity of the current density across the wafer, turbulences in the agitated electrolyte, and the dependence of the ion replenishment on structure size),
- variations of the mean grain size in the gold tracks and nickel diffusion barrier, also caused by an inhomogeneous current density during plating,
- local stress differences due to varying relaxation during ramp-down (caused by inhomogeneous temperature distribution in the soldering oven).

Major design-related factors with an influence on chip bow are:

- bump geometry: if the bumps are longer than wide, the contractive force in the direction of larger bump extension is greater leading to cylindrical deformations,
- bump size: generally speaking, the larger the bump diameter or the solder thickness, the larger the deformation. Therefore, bending is usually most pronounced in the chip area with bump diameters of $Ø_{bump} = 200\,\mu m$, followed by the area with $Ø_{bump} = 150\,\mu m$ and then the area with $Ø_{bump} = 100\,\mu m$,
- orientation of design features (like gold tracks) relative to the crystal lattice of silicon: because the elastic modulus for silicon in the <100> direction is about 23% smaller than in the <110> direction, bending preferably occurs around the <100> direction (resulting in the stretching or compression of the <100> direction). This feature can be observed in chips with a more or less homogeneous bump distribution or in extremely thin silicon ($t_{Si} \approx 2\,\mu m$) chips with a 300 nm thermal silicon dioxide layer (see Figure 5.16 b)). The structure of the SiO_2 layer depends on processing parameters and may be crystalline or amorphous [186]. SiO_2 has a much smaller coefficient of thermal expansion than silicon (about 0.5 ppm/°C for SiO_2 vs. 3.12 ppm/°C for Si). During

> ramp-down, after the thermal oxidation at about 1200°C the silicon contracts 6 times more rapidly than the silicon dioxide resulting in large tensile stress in the silicon.

In this thesis, thin chips were fabricated and soldered to polyimide tapes with gold tracks using a gold/tin solder. The process sequence, the choice of materials, and the sample design were defined in order to be compatible with the production process of the EPI-RET retina implant. In order to transfer the suggested processes to the EPI-RET implant, the aluminum bond pads of the CMOS chips need to be treated with the zincate process. The feasibility of this process in combination with ultrathin silicon CMOS chips was demonstrated in [9]. With their inherent flexibility and low weight, the ultrathin chips offer a unique opportunity to improve the retina implant. As a medical product for permanent implantation the requirements for regulatory approval are very strict for the EPI-RET implant. By using gold/tin solder a significant step toward a permanent, biocompatible, and reliable solder connection was made.

In this thesis, the minimum silicon thickness of the fabricated chips was 3.5 µm for chips with an edge length of 10 mm, resulting in a maximum length-to-thickness ratio for the bulk silicon of about 3000, which is much larger than for any other group working on flip chip solder connections for thin chips. For the first time a simulation model was tested against measured deformation data for extremely thin silicon chips. The anisotropy of silicon was taken into account and shown to have an influence on the direction of bending. More precisely, the relative thickness of the bulk silicon compared to the metal structures was found to be the decisive factor. For the first time it was proven, that for extremely thin silicon, the direction of the bending axis is dominated by the design of the metal structure, whereas for less thin silicon samples the anisotropic properties of the bulk silicon dominate the bending axis direction.

By using a special chip design based on lines of gold/tin solder, global deformations of ultrathin silicon samples could be measured for the first time during gold/tin reflow. The inverted bending behavior during ramp-up and ramp-down could therefore be measured separately. Optimized FEM models were derived from this in-process data, that showed good agreement between simulation and measurement. The model was based on the superposition of results from three separate simulations (for deformations due to intrinsic plating stress and thermal mismatch during ramp-up and ramp-down). Without further adjustments the devised three-step model was used to correctly predict local silicon deformations perpendicular to the gold/tin lines. In another set of chips with circular gold/tin bumps, the local silicon deformation around the reflowed bumps could not be correctly reproduced using the previously successful 3-step simulation. However, by using only the ramp-down simulation model in combination with a reduction of the CTE for gold/tin to 70% of the literature value, good quantitative and qualitative agreement was again achieved. In this case the reduction of the gold/tin CTE seems to be equivalent to the superposition of inversely oriented deformations from the ramp-up and ramp-down in the previous simulations. Similarly, a single step simulation was used to correctly simulate deformations found after flip chip soldering of thin silicon chips onto polyimide tapes. Here, however, the parameters for gold/tin were left unchanged and only the Young's modulus for nickel and gold were reduced to about 60% of the literature values. Finally, it was shown that the thermal mismatch between polyimide tapes and thin silicon chips is strongly influenced by the added metal structures (gold tracks, gold/tin solder, etc…). In simulations it was shown, that by reducing the silicon thickness to about 2 µm the silicon chip expanded about the same as the polyimide substrate during reflow, reducing thermal misalignment significantly.

6 Testing

In this chapter mechanical and electrical tests carried out on individual silicon chips or silicon chips soldered to polyimide tapes are presented.

6.1 Shear tests

The shear tests were carried out to qualify the tin plating process with respect to gold/tin solderability. Therefore, the bump design and dimensions were not yet optimized. Instead of Au-Ni-Au/Sn bumps, much simpler bumps composed of only 30 µm of gold + 10 µm of tin were used. To carry out the shear tests, 20 silicon chips (edge length = 5 mm, t_{Si} = 525 µm) were prepared with 192 cylindrical gold/tin bumps. Each chip contained 88 bumps with Ø = 100 µm, 60 bumps with Ø = 150 µm, and 44 bumps with Ø = 200 µm. For each shear testing sample two chips were flipped onto each other and soldered. The resulting 10 samples were attached to a carrier and tested in a Dage shear tester (type 2400PC). The main settings were:

- land speed: 50 µm/s,
- shear speed: 12.5 µm/s,
- shear height: 700 µm.

The maximum force was recorded and the samples were analyzed under the microscope. In Figure 6.1 the maximum force per bump for the 10 soldered samples is given.

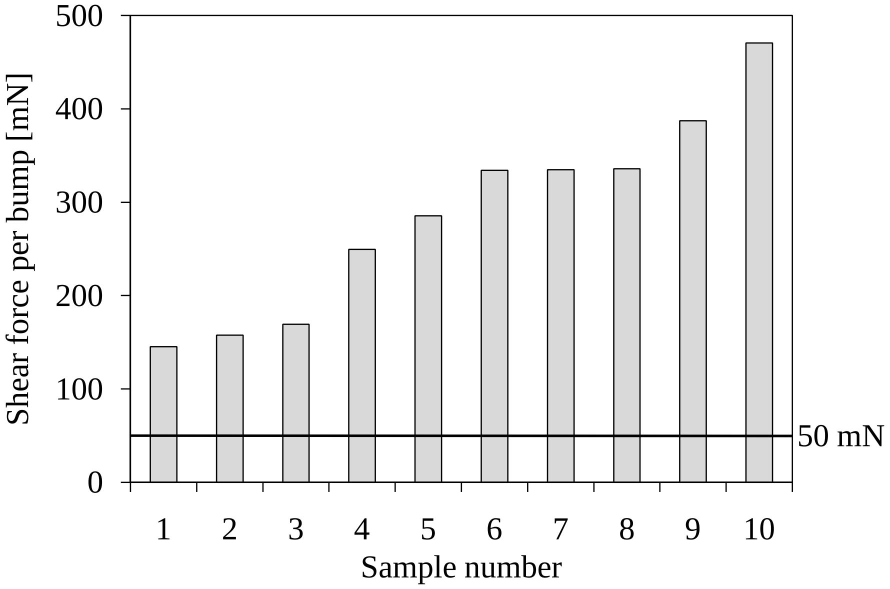

Figure 6.1: Measured shear force per bump in 10 soldered samples (samples sorted by shear force), each having 192 bumps with diameters of 100 µm, 150 µm, and 200 µm. The minimum value of 50 mN required for good solder connects is also depicted.

All values are larger than the value of 50 mN per bump required for good flip chip connections according to DIN EN60749 22:2003 Method D [187]. With shear forces between 27.8 N and 90.4 N the chips also passed MIL-STD-883E: Method 2019.5 [188], which requires a shear force of at least 25 N for chips with a die area ≥ 4 mm². In summary all chips had a sufficient bond strength.

The microscopic analysis yielded that all connections broke at the plating-base-to-silicon dioxide interface or within the plating base. An explanation for the large spread of measured shear forces may be the varying adhesion of the plating base to the silicon dioxide surface.

The solder material itself is generally much stronger than could be shown in these tests, which were limited by the plating base adhesion. In [189], for example, electroplated gold-tin bumps consisting of 25 µm of gold + 5 µm of tin were flip chip soldered, typically resulting in a shear strength between 1.4 N and 2.4 N per 120 µm diameter bump.

6.2 Mechanical bending of bare silicon chips

In order to analyze the flexibility and mechanical stability of thin silicon chips fabricated using the DbyT process, bending tests were carried out in a 4-point bending unit, shown in Figure 6.2.

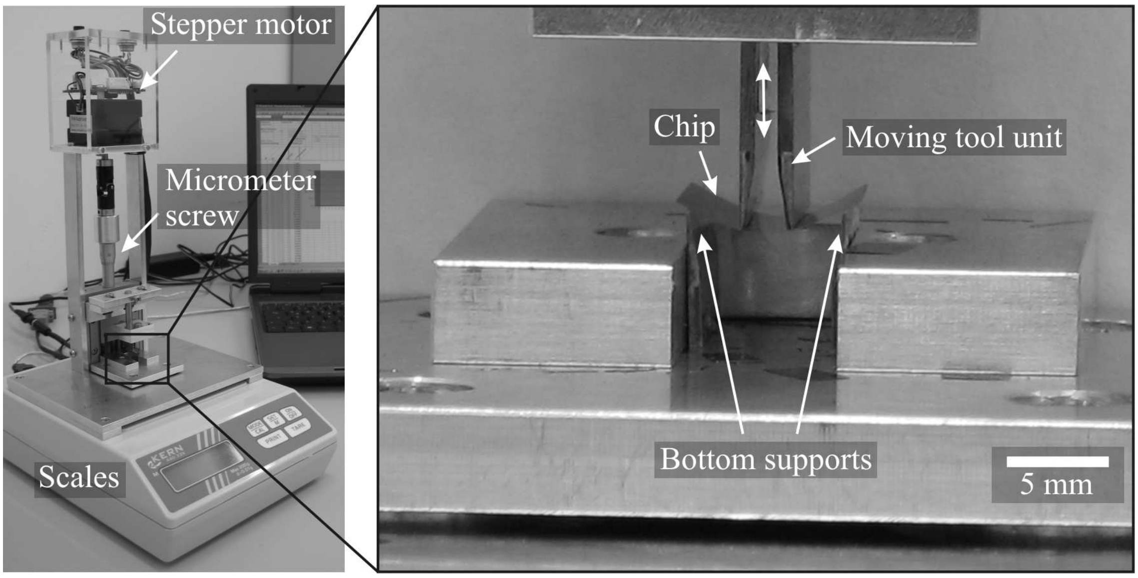

Figure 6.2: 4-point bending unit consisting of commercial scales for force measurement, a stepper motor and micrometer screw driving the moving tool, and a computer for stepper motor control and data read-out.

The chips were fabricated from bare silicon wafers (without electroplated layers) using the DbyT process. The chips with an edge length of 10 mm were placed onto 2 mechanical supports with a distance of ≈ 8 mm. The two top supports with a distance of ≈ 2.5 mm were slowly moved downward, driven by a stepper-motor. The chip middle was forced downward, resulting in a constant moment in the chip between the two top supports. The commercial scales, supporting the bottom part of the tool, continuously recorded the resulting force. The top tool was moved downward until chip breaking occurs (or until the chip slid through the two bottom supports). In Table 6.1 the results (maximum force F_{max} and dislocation y_{max}, radius at breaking R_{break}, and stress in the silicon surface at breaking σ_{break}) for 4 chips with different silicon thicknesses are given. The maximum stress in the silicon surface reached a value above 300 MPa and bending radii as small as 4 mm were observed before breaking.

Table 6.1: 4-point bending results for 4 thin silicon chips without electroplated layers.

Chip	t_{Si} [µm]	F_{max} [mN]	y_{max} [µm]	R_{break} [mm]	σ_{break} [MPa]
1	31	380	939	6.46	312
2	25	206	1218	6.25	260
3	22	86	455	10.20	140
4	16	85	1207	3.97	262

Cyclic bending tests were also carried out. In Figure 6.3 the measured force required to repeatedly bend a silicon chip with a thickness of 16 mm by dislocating the center by 0.6 mm is shown.

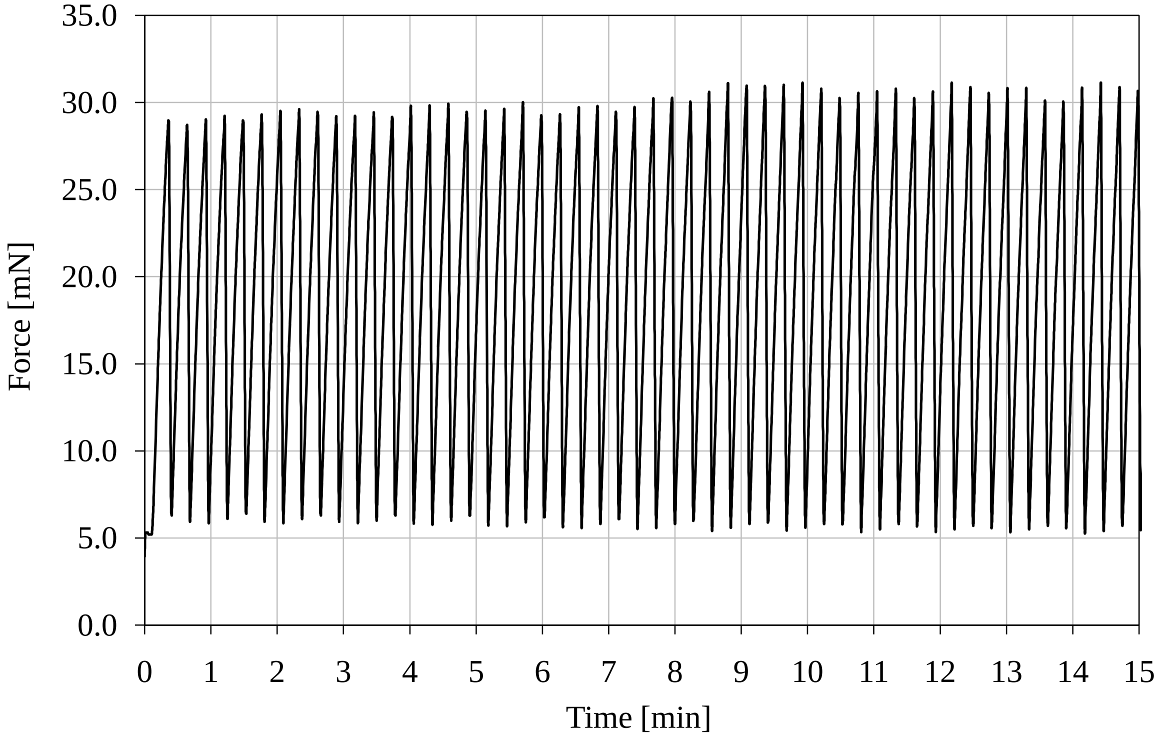

Figure 6.3: Cyclic bending test of a silicon chip with a thickness of 16 µm and an edge length of 10 mm. The diagram shows the force measured by the bending unit during the cyclic bending test.

The chip could be bent at least fifty times without breaking. The number of cycles was limited by the chips tendency to move sideways during the test, finally sliding off the support.

6.3 Mechanical bending of soldered polyimide-silicon systems

The mechanical bending results in the previous section are very promising, demonstrating the general robustness of bare silicon chips in single and multiple bending tests. However, the mechanical stability of thin chips soldered to flexible substrates may be much weaker and therefore must be analyzed experimentally.

Several soldered samples were mechanically bent by forcing them into cylindrical metal forms of decreasing radius. This test setup was chosen, because it more closely resembled the situation, when such a soldered flex chip system is placed into the human eyeball. After all, it is the aim of this thesis to explore and demonstrate possible strategies for thin chip integration into the EPI-RET eye implant using biocompatible flip chip soldering. The bending test was carried out using forms with radii of 40 mm, 30 mm, 20 mm, 15 mm, 10 mm, and 7.5 mm. An example of a flex system bent by forcing it into a cylindrical metal form is shown in Figure 6.4. Most samples broke at a forced bending radius of 7.5 mm. The human eye has an average inner radius of 12 mm [36]. Therefore, most chips would survive adaptation to the radius of curvature of the human eye.

Using an underfill, the mechanical stability and minimal bending radius may further be improved. For example, in [151] it was shown that in standard silicon chips, soldered to polyimide tapes using gold/tin solder, the mean-time-to-failure in thermal cycling tests is increased by a factor of about 30 when an underfill filled with silica particles is used. The underfill reduces the stress on the interconnects when the substrate and chip contract differently during cool-down.

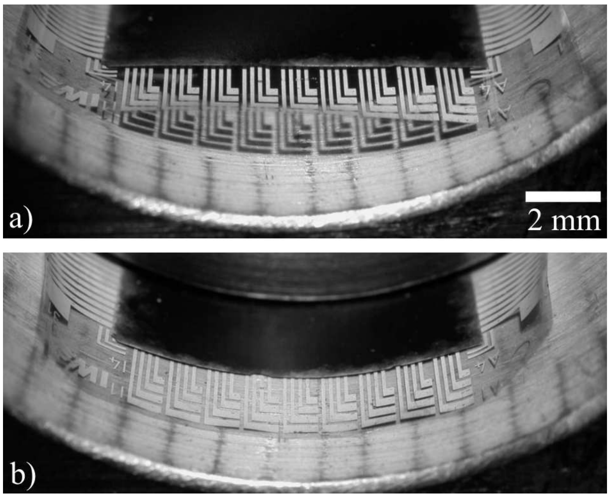

Figure 6.4: Bending of a complete system consisting of silicon chip with a thickness of approximately 20 µm soldered to a polyimide tape; a) unbent system in a cylindrical form b) the same system pushed down by another metal cylinder and forced to adapt to the 10 mm radius of the cylindrical form.

Also, the silicon chip was not protected when it was forced into the cylindrical form. Introducing a mechanical polymer buffer may further improve the chips resistance to breaking during bending.

One flex system was bent such that the complete silicon chip was in contact with the cylindrical form as shown in Figure 6.5. This way a smaller bending radius of 7.5 mm could be achieved without breaking the chip.

Figure 6.5: Bending of a complete system consisting of a silicon chip with a thickness of approximately 20 µm, soldered to a polyimide tape. The image shows mechanical bending with a radius of 7.5 mm.

6.4 Electrical measurements

In order to verify the electrical functionality of soldered silicon-polyimide flex systems, the Daisy chain connectivity and Kelvin contact resistance were measured. After soldering process optimization, a bump yield close to 100% was achieved with average contact resistances as low as 0.63 Ω. Details on bump yield and contact resistance measurements for an optimized soldering process are given in Table 6.2.

Table 6.2: Yield and electrical properties of a flex chip soldered to a polyimide substrate with Au/Sn.

Bump diameter [µm]	Functional Daisy chains	Bump yield [%]	Avg. contact resistance [mΩ]
100	8 / 10	99	7.3
150	10 / 10	100	1.2
200	10 / 10	100	0.63

For comparison, these contact resistances are lower than other connection techniques that are typically used for thin chip attachment, like gold-filled ACA ($R \approx 20$ mΩ for 100 µm contacts) and NCA ($R \approx 10$ mΩ for 100 µm contacts) [3].

7 Conclusion

7.1 Summary

In this thesis, a significant contribution was made to the understanding of local and global deformations during reflow of ultrathin chips and during subsequent soldering of these chips to polyimide tapes using eutectic gold/tin (80Au20Sn) solder. The feasibility of a gold/tin solder flip chip process for attachment of ultrathin silicon chips to flexible polyimide substrates, which survive forced mechanical bending down to a radius of 10 mm, was demonstrated for the first time.

Silicon test chips with an edge length of up to 10 mm and a silicon thickness between about 4 µm and 45 µm were fabricated using the Dicing-by-Thinning process. The process started with a dry etching step using a resist mask to create trenches around the designated chip areas on the wafer's front, followed by a backside thinning sequence that was based on consecutive lapping, CMP, and wet etching steps. Upon reaching the front side trenches, the thinning process was stopped and individual chips with a thickness similar to the initial trench depth were obtained. A typical chip strength of 250 MPa was reached with this process and bare silicon chips survived cyclic bending tests with at least 50 cycles.

Prior to the DbyT process the chips were prepared with different electroplated and lithographically patterned metallic structures like gold tracks, lines of gold-tin, or gold-tin bumps with diameters of 100 µm, 150 µm, or 200 µm. Some of these structures were part of the Daisy chains and Kelvin contacts used to measure connectivities and contact resistances in flip chip soldered samples. Underneath the gold/tin solder a nickel diffusion barrier with a thickness of 2 µm was introduced to prevent intermetallic reactions between the tin in the solder and the gold tracks. Also, by using the nickel barrier, the composition of the reflowed solder was defined by the deposited amounts of gold and tin alone. This facilitated the ensuing soldering process control, because according to [189] the structure and composition of thick gold + tin bumps after reflow are sensitive to the heating rate (for example slow heating rates (of about 10°C/s) lead to increased ζ-phase formation, resulting in poor wetting behavior).

Similar to the silicon test chips, polyimide tapes were fabricated with corresponding metal structures. The pure gold and tin layers were reflowed to form eutectic gold/tin solder caps. This bump reflow and the subsequent soldering were generally carried out in a nitrogen atmosphere with 5% formic acid and in some cases by adding a drop of 2% adipic acid dissolved in isopropanol directly to the sample. The reflow was necessary to reduce the number of voids in the solder connects, but had the disadvantage of substantial bowing, which impeded handling and alignment during the subsequent soldering process. After soldering, functional samples with a total height below 50 µm were obtained (with the following approximate layer thicknesses: polyimide substrate: 5 µm, gold tracks: 2*2 µm, gold bases 2*2 µm, nickel diffusion barriers: 2*2 µm, gold/tin solder: 14 µm, bulk silicon: 16 µm).

Individual soldered systems showed bump connectivities above 99% (as shown in Daisy chain measurements) and average contact resistances below 1 mΩ (according to Kelvin contact measurements on interconnects with a diameter of 200 µm). Minimum forced bending radii of 8 mm and 10 mm (depending on the bending setup) were achieved without chip breakage.

The magnitude of deformations resulting from the bump reflow should be minimized for handling purposes. A finite element model was devised using the ANSYS 11.0 simulation software. The model, which was based on only a small section of each sample in order to reduce simulation time, predicted chip and substrate deformations during reflow. The simulated bows were compared to in-process measurements of the chip bow during reflow in air. Also, local deformations (for example around individual bumps) were simulated and measured using complex data processing to obtain relevant deformation data from the surface profiling measurements. Because the geometry and material properties of the solder bump changed during reflow, at least two different geometries were simulated for each reflow process, and material parameters, like the Young's modulus and coefficient of thermal expansion, were adapted to achieve good agreement between the simulation and measurement. Because chips with different thicknesses were available, the simulations and measurements were correlated as a function of silicon thickness. Using the results from these model optimizations, the local deformation underneath and around individual solder bumps were also simulated and the model was adapted, where necessary.

The in-process deformation measurement was based on optical image processing, allowing dynamic processes, like the melting of tin, phase transitions, and relaxation in the solder and metal structures, to be observed. Although the silicon thickness was the most important factor with respect to chip bow after reflow, other factors also had a significant influence on bowing. For example, as an anisotropic material, silicon preferably bent around the <100> direction. However, the thinner the chips, the more dominant the geometry of the metal structures with respect to the bending direction. As shown in simulations and confirmed by measurements, the transition from silicon-dominated to metal-structure-dominated bending behavior is at about $t_{Si} = 13\ \mu m$ for the Daisy chain structures analyzed in this thesis. Although thinner chips generally showed larger deformation after reflow, it was found in the simulations that for very thin silicon (typically in the order of a few microns) the bow during ramp-up would just be compensated by the inversely oriented bow during ramp-down, resulting in a flat chip after reflow. The plausibility of this simulation result was demonstrated with a sample that had a thickness of only 2 µm along one chip edge and was indeed concavely bent along this edge after ramp-down (that is it was inversely bent compared to thicker samples). It was also found that the intrinsic stress from electroplating had to be taken into account for good agreement between simulation and measurement. To further improve the agreement between measurement and simulation, the bow measurements and simulation results were averaged in one group of chips with an inhomogeneous silicon thickness.

7.2 Outlook

The material properties of the different materials were taken from the literature and adapted to achieve a good fit between simulation and measurement. By determining the mechanical properties (like the CTE, Young's modulus, and Poisson's ratio) of the actually used materials, an even better FE model may be devised. In this scope, the temperature dependence of material parameters, as well as structural reorganization processes at elevated temperatures, would be of special interest and should also be determined. Material parameters could, for example, be determined by electroplating one material at a time onto thin chips and observing global deformations as a function of temperature for various heating rates or for long dwell times at certain temperatures. By varying the silicon thickness and/or the thickness of the plated material additional data could be collected.

The fabrication and soldering processes for ultrathin silicon chips described in this thesis were shown to work well with test chips and substrates. Extending the application of these processes to the EPI-RET implant is the next natural step. In [9] the successful thinning of CMOS chips using a slightly different DbyT process compared to this thesis was shown (the major difference being a dry etching step at the end of the thinning sequence). The chips were connected to polyimide substrates by isotropically conductive adhesive and therefore required treatment of the non-conductive aluminum oxide surface of the bond pads of the chips prior to flip chip attachment. This was achieved by means of the zincate process, which replaces the non-conductive aluminum oxide surface layer by a sequence of electroless zinc, nickel, and gold. By using a gold/tin soldering process instead of ICA, but also in combination with the same zincate process, a more reliable flux-less (and hence biocompatible) bonding process may be achieved. Although the functionality of gold/tin soldered EPI-RET chips has yet to be demonstrated, CMOS chips should generally not be damaged by typical gold/tin soldering temperatures of 330°C. Other processes like the silicone encapsulation and forced implant bending also need to be tested with respect to their influence on chip functionality. Although CMOS chips generally remain functional, when exposed to bending-induced stress, the transistor characteristics are known to change. However, this may be compensated by using an adapted CMOS design. The mechanical stability of the implant may be improved by application of an underfill after the soldering process. The total implant thickness may be further reduced by optimizing the individual metal layer thicknesses. Especially the thickness of the gold/tin solder may be reduced from its present thickness of 14 µm after soldering.

Glossary

Symbols

A	Area
a	First of three lattice parameters
b	Second of three lattice parameters
b	Bow
b_{Si}	Edge length
c	Third of three lattice parameters
c	Concentration
c_{xy}	Elastic constants or stiffness coefficients of <100> silicon (1< x,y <6)
c'_{xy}	Elastic constants or stiffness coefficients of <110> silicon (1< x,y <6)
d	Distance
D	Diffusion coefficient; height difference (bow)
D_{xy}	Diameter of object xy
d_{sb}	Distance between support beams
D_0	Maximum diffusion coefficient
E	Young's modulus
E_A	Activation energy for diffusion
F	Force
G	Shear modulus
h	Height
I_{in}, I_{out}	Electric current flowing into or out of a device under test
J	Geometrical moment of inertia
k	Boltzmann constant (= $8.617 \cdot 10^{-5}$ eV/K) or CTE reduction factor for Au and Ni
[K]	Solution matrix in the ANSYS software
L	Liquid material in the gold/tin phase diagram; length
[L]	Lower triangular matrix
l_{xy}	Length of object xy
l_1	Direction cosine (of angle of rotated x-axis and original x-axis)
l_2	Direction cosine (of angle of rotated x-axis and original y-axis)
m	Mass
M	Moment
n-type	Doped with acceptor atoms
p-type	Doped with donor atoms
Q	Activation energy
R	Radius of curvature; molar gas constant (R = 8.314 J/mol·K)
t_{xy}	Thickness of object xy
T	Temperature
T_{melt}	Melting temperature
U_{meas}	Measured voltage
[U]	Upper triangular matrix
w_{xy}	Width of object xy
x, y, z	The three perpendicular spatial directions
α-Sn	Allotropic form of tin for T < 13°C (crystal structure: diamond cubic)
(α-Sn)	α-tin where some tin atoms were substituted by gold atoms
β	$Au_{10}Sn$
β-Sn	Allotropic form of tin for T > 13°C (crystal structure: bct lattice)

(β-Sn)	β-Sn where some tin atoms were substituted by gold atoms
α	Bending angle; coefficient of thermal expansion
α_{CTE}	Coefficient of thermal expansion
ζ	Gold/tin phase with 10 to 18.5 wt.% tin, exists between 190°C and 519°C
ζ'	Au_5Sn
δ	AuSn
Δm	Misalignment
ε	$AuSn_2$; direction strain
η	$AuSn_4$
γ	Gradient of bending angle vs. temperature; shear strain
γ_1	Gradient for Au-Ni-Au-Sn on Si for low temperatures (around 85°C)
γ_2	Gradient for Au-Ni-Au-Sn on Si for high temperatures (around 200°C)
γ_3	Gradient for Au-Ni on Si between 280°C and 330°C
γ_4	Gradient for Au-Ni-Au/Sn on Si between 190°C and 280°C
γ_5	Gradient for Au-Ni-Au/Sn on Si between 20°C and 190°C
$\gamma_{4,5}$	Gradient for Au-Ni-Au/Sn on Si between 20°C and 280°C
γ_1'	Gradient for Au-Ni-Au-Sn on PI for low temperatures (around 85°C)
γ_2'	Gradient for Au-Ni-Au-Sn on PI for high temperatures (around 140°C)
γ_{3A}'	Gradient for Au-Ni on PI for ramp-up between 280°C and 330°C
γ_{3B}'	Gradient for Au-Ni on PI for ramp-down between 330°C and 280°C
γ_4'	Average gradient for Au-Ni-Au/Sn on PI between about 190°C and 20°C
σ	Electrical conductivity; direction stress
τ	Shear stress
ρ	Electrical resistivity
ν	Poisson's ratio
θ	Angle between bending axis and chip edge
Ø	Diameter
$\parallel$	Parallel
$\perp$	Perpendicular

Abbreviations and Acronyms

ACA	Anisotropically conductive adhesive
Ag	Silver
Al	Aluminum
Al_2O_3	Aluminum oxide, alumina
at.%	Atomic percent
Au	Gold
(Au)	Gold with a small amount of tin
$(Au_{0.5}Ni_{0.5})Sn_4$	$AuSn_4$, where half of all gold atoms are replaced by nickel atoms
AuSn	An intermetallic compound of gold and tin
$AuSn_2$	An intermetallic compound of gold and tin
$AuSn_4$	An intermetallic compound of gold and tin
Au_5Sn	An intermetallic compound of gold and tin
$Au_{10}Sn$	An intermetallic compound of gold and tin
avg.	Average
80Au20Sn	Gold-rich eutectic with 80 wt.% gold and 20 wt.% tin
10Au90Sn	Tin-rich eutectic with 10 wt.% gold and 90 wt.% tin
bct	Body centered tetragonal
BCB	Benzocyclobutene
BPDA-PDA	poly(biphenylene tetracarboxylic dianhydride p-phenylene diamine)
CF_2Cl_2	Dichlorodifluoromethane

CF_4	Carbon tetrafluoride
CMOS	Complementary metal oxide semiconductor
CMP	Chemical-mechanical polishing, chemical-mechanical planarization
Cr	Chromium
CSP	Chip scale package
CTE	Coefficient of thermal expansion
Cu	Copper
DBG	Dicing before grinding
DbyT	Dicing-by-Thinning
Dev.	Deviation
DI	Deionized
DNP	Distance to neutral point
e-Ni	Electroless nickel
FEA	Finite element analysis
FEM	Finite element model
FR4	A flame retardant material often used as a substrate material
fcc	Face-centered cubic
GaAs	Gallium arsenide
GPS	Global positioning system
hcp	Hexagonal close-packed
hex	Hexagonal
HF	Hydrofluoric acid
HNO_3	Nitric acid
H_3PO_4	Phosphoric acid
HNA	Hydrofluoric acid + nitric acid + acetic acid etching solution
IC	Integrated circuit
ICA	Isotropically conductive adhesive
ICP	Inductively coupled plasma
IMB	Integrated module board
IMC	Intermetallic compound
I/O	Input/output
IPA	Isopropanol
IPC	In-process control
IRC	International Roadmap Committee
ISB	Immersion solder bumping
ITRS	International technology roadmap for semiconductors
J_2	Iodine
KJ	Potassium iodine
KOH	Potassium hydroxide
$K_3[Fe(CN_6)]$	Potassium ferricyanide
LCD	Liquid crystal display
MCM	Multi chip module
MEMS	Microelectromechanical systems
MIL-STD	Military standard
MOSFET	Metal oxide semiconductor field effect transistor
NCA	Non-conductive adhesive
Ni	Nickel
Ni_3Sn_2	Chemical compound consisting of 3 nickel and 2 tin atoms
N/A	Not available, not applicable
orth	Orthorhombic
O_2	Molecular oxygen

PCB	Printed circuit board
PCSB	Plastic core solder bump
Pd	Palladium
PDA	Personal digital assistant
PDMS	Polydimethylsiloxane
PET	Poly ethylene terephthalate
PI	Polyimide
PLD	Pulsed laser deposition
pn	Positively/negatively doped
ppm	Parts per million
Pt	Platinum
PTE	Periodic table of elements
PVD	Physical vapor deposition
RCC	Resin coated copper
RF	Radio frequency
RFID	Radio frequency identification
RoHS	Restriction of hazardous substances
RPM	Revolutions per minute
RT	Room temperature
RTP	Rapid thermal processing
R2R	Role-to-role, reel-to-reel
SEM	Scanning electron microscope
SF_6	Sulfur hexafluoride
Si	Silicon
SiC	Silicon carbide
Sim.	Simulation
SiO_2	Silicon dioxide
SiP	System in Package
Sn	Tin
Sn^{4+}	Stannic tin
Sn^{2+}	Stannous tin
(Sn)	Tin with a small amount of gold
SOI	Silicon on insulator
TAB	Tape automated bonding
TFTC	Thin film transistor circuits
Ti	Titanium
TiW	Titanium tungsten
T_{melt}	Melting temperature
TTV	Total thickness variation
UBM	Under bump metallization
UV	Ultraviolet (radiation)
vol.%	Volume percent
wt.%	Weight percent
Zn	Zinc
2D	Two-dimensional
3D	Three-dimensional

References

[1] K.-D. Lange, "Hochentwickelte Integrations- und Packagingtechnologien für miniaturisierte Systeme", Mikrosystemtechnikkongress, Dresden, VDE Verlag GmbH, Berlin-Offenbach, ISBN 978-3-8007-3061-2 (2007)

[2] "Flexible Substrate Magazine", Publisher and Editor-in-Chief: Mark Fihn, available at www.veritasetvisus.com, Vol. 3, No. 5 (2007)

[3] J. W. Blade, "Foldable Flex and Thinned Silicon Multichip Packaging Technology", Kluwer Academic Publishers Boston, ISBN 0-7923-7676-5 (2003)

[4] Anon, "International Technology Roadmap for Semiconductors, 2007 Edition, Assembly and Packaging", available at www.itrs.net (2007)

[5] M. Feil, C. Adler, G. Klink, M. König, C. Landesberger, S. Scherbaum, G. Schwinn, and H. Spöhrle, "Ultra thin ICs and MEMS elements: techniques for wafer thinning, stress-free separation, assembly and interconnection", Journal of Microsystem Technology, Vol. 9, pp. 176-82, DOI: 10.1007/S00542-002-0223-5 (2003)

[6] J. Haberland and C. Kallmayer, "Ultra Thin Flip Chip Interconnects", Frequenz - Berlin, Vol. 58, pp. 65-9 (2004)

[7] W.J. Kröninger, "Flexibility and Strength of Backside-treated Silicon-Chips", APiA Seminar (2002)

[8] T. Onishi, "SiP Technology Development in Japan", 11th Annual International KGD Packaging and Test Workshop, Napa, California (2004)

[9] G. Kizilirmak, ”Frei applizierbare MOSFET-Sensorfolie zur Dehnungsmessung”, dissertation, RWTH Aachen University (2007)

[10] G. Klink, M. Feil, F. Ansorge, R. Aschenbrenner, and H. Reichl, "Innovative Packaging Concepts for Ultra Thin Integrated Circuits", 51st Electronic Components and Technology Conference, pp. 1034-9, DOI 10.1109/ECTC.2001.927940 (2001)

[11] Fraunhofer ISIT, Annual report (2002)

[12] C. Landesberger, "Innovative Fertigungsverfahren für neue Produkte auf der Basis dünner Silizium-Bauelemente, Ergebnisbericht des BMBF Verbundprojektes "InnoSi", Fraunhofer IRB Verlag, ISBN 3-8167-6461-4 (2004)

[13] Website: http://www.epiret.de (2008)

[14] C. Koch, W. Mokwa, M. Goertz, and P. Walter, "First Results of a Study on a Completely Implanted Retinal Prosthesis in Blind Humans", 7th IEEE conference on Sensors (2008)

[15] T. Stieglitz, H. Beutel, M. Schuettler, and J.-U. Meyer, "Micromachined, Polyimide-Based Devices for Flexible Neural Interfaces", Biomedical Microdevices, Vol. 2, No. 4, pp. 283-94, DOI 10.1023/A:1009955222114 (2000)

[16] C. Koch, H. Fassbender, U. Nolten, M. Görtz, and W. Mokwa, "Fabrication and Assembly Techniques for a 3rd Generation Wireless Epiretinal Prosthesis", 19th MicroMechanics Europe Workshop (2008)

[17] K. Hungar, M. Görtz, E. Slavcheva, G. Spanier, C. Weidig, and W. Mokwa, "Production processes for a flexible retina implant", Journal of Sensors and Actuators A - Physical, Vol. 123-24, pp. 172-8, doi:10.1016/j.sna.2005.03.061 (2005)

[18] K. Hungar, W. Mokwa, "Gold/tin soldering of flexible silicon chips onto polymer tapes", Journal of Micromechanical Engineering, Vol. 18, No. 6, DOI: 10.1088/0960-1317/18/6/064002 (2008)

[19] M. Reiche and G. Wagner, "Wafer Thinning: Techniques for Ultra-thin Wafers", Advanced Packaging Magazine, Vol. 12, No. 3 (2003),

[20] L. Böttcher, A. Neumann, T. Löher, A. Ostmann, and H. Reichl, "Integration of Active Components Into Rigid and Flex Substrates", presentation by IZM at Productronica (2005)

[21] Unaxis Wafer Processing, "Metal Deposition by PVD for UBM/ RDL and Backside of Thin Wafers", Die Products Consortium's KGD Packaging&Test Symposium - APiA Wafer Level Packaging Seminar (2004)

[22] R. Plieninger, "SiP Experience and Future Directions", presentation at 11th Annual International KGD Packaging and Test Workshop, Napa, California (2004)

[23] Project FLIBUSI, Results Bulletin, No. 1 (2003)

[24] K. Arita, "Plasma Stress Relief Technology for Wafer Thinning Process", 11th Annual International KGD Packaging and Test Workshop, Napa, California (2004)

[25] H.H. Jiun, I. Ahmad, A. Jalar, and G. Omar, "Effect of wafer thinning methods towards fracture strength and topography of silicon die", Journal of Microelectronics and Reliability, Vol. 46, No. 5-6, pp. 836-45 (2006)

[26] C. McHatton and C.M. Gumbert, "Eliminating backgrind defects with wet chemical etching", Journal of Solid State Technology, Vol. 41, No. 11, pp. 85-90 (1998)

[27] M. Kulawski, K. Henttinen, I. Suni, F. Weimar, and J. Mäkinen, "A novel CMP Process on Fixed Abrasive Pads for the Manufacturing of highly planar thick film SOI Substrates", Proceeding of Materials Research Society Symposium, Vol. 767, pp. 133–9 (2003)

[28] F. Henley, A. Lamm, S. Kang, Z. Liu, and L. Tian, "Direct Film Transfer (DFT) Technology for Kerf-Free Silicon Wafering", European Photovoltaics Conference and Exhibition, Valencia (2008)

[29] C. Landesberger, S. Scherbaum, and K. Bock, "Carrier techniques for thin wafer processing", ManTech Conference, pp. 33-6 (2007)

[30] K. Bock, M. Bleier, O. Köthe, and C. Landesberger, "New Manufacturing Concepts for Ultra-Thin Silicon and Gallium Arsenide Substrates", International Conference on Compound Semiconductor Manufacturing Technology (2003)

[31] G. Kizilirmak, W. Mokwa, and U. Schnakenberg, "Performance Studies on Inversion Channels of nMOSFETs under Extreme Mechanical Load", Proc. of IEEE Sensors, pp. 1585-8, DOI 10.1109/ICSENS.2004.1426494 (2004)

[32] W. Christiaens, B. Vandevelde, E. Bosman, and J. Vanfleteren, "UTCP : 60 µm Thick Bendable Chip Package", International Wafer-Level Packaging Conference (2006)

[33] W. Christiaens, E. Bosman, W. Huwel, W. Perdu, and J. Vanfleteren, "Flexible Polyimide Based Ultra-Thin Chip Package (UTCP)", International Workshop on Flexible & Stretchable Electronics, at IMEC, Leuven, Belgium (2007)

[34] M. Cauwe, W. Christiaens, J. Vanfleteren, J. De Baets, and A. Ostmann, "Embedding Active Components as a 3D Packaging Solution", Advanced Microelectronics Magazine (May/June), pp. 15-9 (2006)

[35] N. Sillon, J.-C. Souriau, and J. Brun, "Innovative Flip Chip Solution for System-On-Wafer Concept", Workshop 3S by LTPY, Atlanta (2005)

[36] H.K. Trieu, L. Ewe, W. Mokwa, M. Schwarz, and B.J. Hosticka, "Flexible Silicon Structures for a Retina Implant", Proceedings of MEMS 98, pp. 515-9 (1998)

[37] S.P. Lacour, J. Jones, S. Wagner, T. LI, and Z. Suo, "Stretchable Interconnects for Elastic Electronic Surfaces", Proceedings of the IEEE, Vol. 93, No. 8, pp. 1459-67 (2005)

[38] J.-Y. Chen, L.-S. Huang, C.-H. Chu, and C. Peizen, "A new transferred ultra-thin silicon micropackaging", Journal of Micromechanics and Microengineering, Vol. 12, pp. 406-9 (2002)

[39] W. M. Choi, J. Song, D.-Y. Khang, H. Jiang, Y.Y. Huang, and J.A. Rogers, "Biaxially Stretchable 'Wavy' Silicon Nanomembranes", Nano Letters, Vol. 7, No. 6, pp. 1655-63 (2007)

[40] Y. Sun, W.M. Choi, H. Jiang, Yonggang Y. Huang, and J.A. Rogers, "Controlled buckling of semiconductor nanoribbons for stretchable electronics", Nature Nanotechnology, Vol. 1, pp. 201-7, DOI 10.1038/nnano.2006.131 (2006)

[41] E. Gaulhofer and H. Oyrer, "Wafer Thinning and Strength Enhancement to Meet Emerging Packaging Requirements", IEMT Europe Symposium (2000)

[42] M. Töpper et al., "Biocompatible Hybrid Flip Chip Microsystem Integration for Next Generation Wireless Neural Interfaces", IEEE Electronic Components and Technology Conference, pp. 705-8 (2006)

[43] Anon, "Hitachi Develops a New RFID with Embedded Antenna μ-Chip -- Makes Possible Wireless Links that Work Using Nothing More Than a 0.4mm X 0.4mm Chip, One of the World's Smallest ICs--", press release by Hitachi Ltd. (2003)

[44] Anon, "World's Smallest and Thinnest 0.15 x 0.15 mm 7.5 µm Thick RFID IC Chip", press release by Hitachi Ltd. (2006)

[45] T. Gray, "Hitachi Creates Paper-Thin RFID Tags", CRM Buyer Magazine (February) (2007)

[46] R. Das, "Printed Electronics - Technology Overview and Applications", white paper, IDTechEx Ltd. (2004)

[47] K. Bock, "Reel-to-Reel Application Center", white paper, IZM (2006)

[48] W.J. Kröninger, F. Hecht, G. Lang, F. Mariani, S. Geyer, and L. Schneider, "Time for Change in Pre-Assembly? The Challenge of Thin Chips", 51st Electronic Components and Technology Conference, pp. 1029-33, DOI 10.1109/ECTC.2001.927939 (2001)

[49] K.-H. Bock, "Heterointegration Technologies for Polytronic Systems", Mikrosystemtechnikkongress, Dresden, VDE Verlag GmbH, Berlin-Offenbach, ISBN 978-3-8007-3061-2 (2007)

[50] F. Kriebel, "Smart Active Label – Low Cost Mikrosysteme für Massenanwendungen", Mikrosystemtechnikkongress, Dresden, VDE Verlag GmbH, Berlin-Offenbach, ISBN 978-3-8007-3061-2 (2007)

[51] T. Someya, T. Sekitani, S. Iba, Y. Kato, H. Kawaguchi, and T. Sakurai, "A large-area, flexible pressure sensor matrix with organic field-effect transistors for artificial skin applications", Proceedings of the National Academy of Sciences of the United States of America, Vol. 99, No. 27, pp. 9966-70, DOI 10.1073_pnas.0401918101 (2004)

[52] S. Gimpel, U. Möhring, and D. Zschenderlein, "Mikrosystemtechnik in Textilien und deren Anwendung", Mikrosystemtechnikkongress, Dresden, VDE Verlag GmbH, Berlin-Offenbach, ISBN 978-3-8007-3061-2 (2007)

[53] J. Fjelstad, "Flexible Circuit Technology 3rd Edition", BR Publishing Inc. (Seaside, Oregon, USA), ISBN 978-0-9796189-0-1 (2007)

[54] A. Schubert et al., "Numerical and Experimental Investigations of Large IC Flip Chip Attach", Electronic Components and Technology Conference, pp. 1338-46 (2000)

[55] A. Schubert, R. Dudek, H. Walter, E. Jung, A. Gollhardt, B. Michel, and H. Reichl, "Reliability Assessment of Flip-Chip Assemblies with Lead-free Solder Joints", Electronic Components and Technology Conference, pp. 1246-1255 (2002)

[56] D.N. Pascual and S. Callender, "Flip Chip Bonding of Flexible Circuit Devices - Designed for Biomedical Applications", Advanced Packaging Magazine, Vol. 13, No. 10 (2004)

[57] C. Kallmayer, H. Oppermann, S. Anhock, R. Azadeh, R. Aschenbrenner, and H. Reichl, "Reliability investigations of flip-chip on flex using different solder materials", Electronic Components and Technology Conference, pp. 3003-10 (1998)

[58] D. Katze, "No Flow-Fluxing Underfill Material and Process Considerations", HDI Conference (2000)

[59] H.G. Song, "Microstructural Evolution of Eutectic Au-Sn Solder Joints", dissertation, Berkeley (2002)

[60] H. Oppermann, "Entwicklung von Lötprozessen mit Au/Sn für optoelektronische und höchstfrequente Anwendungen", presentation at Berlin-Brandenburger Akademie der Wissenschaften 21 December (2004)

[61] F. Hohnke, "Untersuchungen an galvanisch erzeugten AuSn-Lotbumps", diploma thesis, FHTW Berlin - University of Applied Sciences (2003)

[62] S. Akhlaghi and D.G. Ivey, "Effect of Processing Parameters on the Electroplating of Au-Sn Solders", Journal of Plating and Surface Finishing, Vol. 90, No. 7, pp. 36-9 (2003)

[63] A.F.J. Baggerman and M.J. Batenburg, "Reliable Au-Sn Flip-Chip Bonding on Flexible Prints", Transactions on Components, Packaging, and Manufacturing Technology B, Vol. 18, No. 2, pp. 257-63 (1995)

[64] G.S. Matijasevic, C.Y. Wang, and C.C. Lee, "Void Free Bonding of Large Silicon Dice Using Gold-Tin Alloys", IEEE Transactions on Components, Hybrids, and Manufacturing Technology, Vol. 13, No. 4, pp. 1128-34 (1990)

[65] R.S. Forman and G. Minogue, "The Basics of Wafer-Level AuSn Soldering", Chip-Scale Review Magazine, Vol. 8, No. 7, pp. 55-9 (2004)

[66] S. Weiß, E. Kaulfersch, M. Topfer, E. Zakel, B. Michel, and H. Reichl, "Design and technology for fluxless die bonding of high power laser bars using Au(80)Sn(20)-solder", 10th Annual Meeting of IEEE Lasers and Electro-Optics Society, pp. 281-2 (1997)

[67] M. Ishikawa, H. Sasaki, S. Ogawa, M. Kohinata, A. Mishima, and H. Yoshida, "Application of Gold-Tin Solder Paste for Fine Parts and Devices", Proceedings of the 55th Electronic Components and Technology Conference, pp. 701-9, DOI 10.1109/ECTC.2005.1441346 (2005)

[68] M. Ishikawa, S. Nonaka, H. Sasaki, M. Kohinata, A. Mishima, and H. Yoshida, "Application Trends of Gold-Tin Solder Paste for Fine Pitch Devices", white paper, Mitsubishi Materials Corporation (2005)

[69] Website: http://emat.eng.hmc.edu/matlsprops/alloys/80au_20sn.htm (2008)

[70] C. Kallmayer, H. Oppermann, J. Klöser, E. Zakel, and H. Reichl, "Experimental results on the self-alignment process using Au/Sn metallurgy and the growth of the xi-phase during the reflow", International Flip Chip, Ball Grid Array, TAB and Advanced Packaging Symposium, pp. 225-36 (1995)

[71] Fraunhofer ISIT, Annual report (2003)

[72] Zribi, R.R. Chromik, R. Preshus, J. Clum, K. Teed, L. Zavalij, J. DeVita, J. Tova, and E.J. Cotts, "Solder Metalization Interdiffusion in Microelectronic Interconnects", Transactions on Components and Packaging Technologies A, Vol. 23, No. 2, pp. 383-7, DOI: 10.1109/6144.846778 (2000)

[73] C.Y. Wang and C.C. Lee, "A Bonding Technique for Thin GaAs Dice with Via Holes Using Gold-Tin Composites", Transactions on Component, Hybrids, and Manufacturing Technology, Vol. 14, No. 4, pp. 874-8 (1991)

[74] Website: http://www.flipchips.com/tutorial46.html (2008)

[75] G. Minogue and R. Mullapudi, "A Novel Approach For Hermetic Wafer Scale MEMS RF and GaAs Packaging", GaAs ManTech Conference (2005)

[76] W. Pittroff et al., "Mounting of High Power Laser Diodes on Boron Nitride Heat Sinks Using an Optimized Au/Sn Metallurgy", IEEE Transactions on Advanced Packaging, Vol. 24, No. 4, pp. 434-41 (2001)

[77] J. Cai, Q. Wang, X. Li, W. Kim, S. Wang, J. Hwang, and C. Moon, "Microstructure of AuSn Wafer Bonding for RF-MEMS Packaging", 6th International Conference on Electronic Packaging Technology, pp. 537-41, DOI 10.1109/ICEPT.2005.1564660 (2005)

[78] M. Hutter, H. Oppermann, G. Engelmann, J. Wolf, O. Ehrmann, R. Aschenbrenner, and H. Reichl, "Calculation of Shape and Experimental Creation of AuSn Solder Bumps for Flip Chip Applications", Electronic Components and Technology Conference, pp. 282-8 (2002)

[79] M. Hutter, F. Hohnke, H. Oppermann, M. Klein, and G. Engelmann, "Assembly and Reliability of Flip Chip Solder Joints Using Miniaturized Au/Sn Bumps", IEEE Electronic Components and Technology Conference, pp. 49-57 (2004)

[80] T. Ishii and S. Aoyama, "Fluxless Fabrications of Sn-Au Solder Microbumps by a Hydrogen Plasma Reflow Technique", Journal of Electronic Materials, Vol. 34, No. 5, pp. 630-4 (2005)

[81] B. Djurfors and D.G. Ivey, "Pulsed Electrodeposition of the Eutectic Au/Sn Solder for Optoelectronic Packaging", Journal of Electronic Materials, Vol. 30, No. 9, pp. 1249-54 (2001)

[82] Indium Corporation of America, "Application Notes - Stencil Design", white paper, downloaded from www.indium.com in (2004)

[83] S. Chen, M. Chen, X. Yi, H. Zhang, and S. Liu, "Research on Wafer Scale Bonding Method Based on Gold-Tin Eutectic Solders", Proceedings of International Conference on Electronic Packaging Technology, pp. 187-9 (2003)

[84] G. Greitmann, H. Burkard, and J. Link, "AuSn Thin Film Solder Layers for Assembly of Optoelectronic Devices", 14th European Microelectronics and Packaging Conference (2003)

[85] C. Belouet, C. Villard, C. Fages, and D. Keller, "Achievement of Homogeneous AuSn Solder by Pulsed Laser-Assisted Deposition", Journal of Electronic Materials, Vol. 28, No. 10, pp. 1123-26 (1999)

[86] D. Kim, J. Kim, G.L. Wang, and C.C. Lee, "Nucleation and growth of intermetallics and gold clusters on thick tin layers in electroplating process", Materials Science and Engineering A, Vol. 393, pp. 315–9 (2005)

[87] W. Sun and D.G. Ivey, "Microstructural study of co-electroplated Au/Sn alloys", Journal of Materials Science, Vol. 36, pp. 757-66 (2001)

[88] Y. Zhang and D.G. Ivey, "Phase Formation in Gold-Tin Alloys Electroplated from a Non-cyanide Bath", GaAs ManTech Conference (2003)

[89] A. He, B. Djurfors, S. Akhlaghi, and D.G. Ivey, "Pulse Plating of Gold-Tin Alloys for Microelectronic and Optoelectronic Applications", white paper, Micralyne Inc. (2003)

[90] J. Gauci, A.-M. Borg, R. Caruana, K. Whitlaw, and J. Crosby, "The Qualification of a Pure Tin Plating Process as a Lead Free Finish for I.C. Packaging", IPC/JEDEC 7th International Conference on Lead Free Electronic Components and Assemblies (2004)

[91] D. Schuhmacher, H. Scheithauer, M. Alavi, R. Niekrawietz, B. de Heij, R. Zengerle, and P. Koltay, "Erzeugung von Mikrotropfen aus flüssigem Lötzinn mittels einer hochparallelen und kontaktlosen Drucktechnik", Mikrosystemtechnikkongress, Dresden, VDE Verlag GmbH, Berlin-Offenbach, ISBN 978-3-8007-3061-2 (2007)

[92] E. Beckert, R. Eberhardt, A. Tünnermann, G. Azdasht, and E. Zakel, "Solder Bumping – eine neue Technologie zur Montage optoelektronischer Mikrosysteme", Mikrosystemtechnikkongress, Dresden, VDE Verlag GmbH, Berlin-Offenbach, ISBN 978-3-8007-3061-2 (2007)

[93] C.Y. Wang and C.C. Lee, "A Eutectic Bonding Technology at a Temperature Below the Eutectic Point", Proceedings of IEEE Electronic Components and Technology Conference, pp. 502-7 (1992)

[94] C.E. Ho and C.R. Kao, "Interfacial Reaction Between Solders and Surface Finishes", Taiwan Printed Circuit Association Forum, pp. 238-46 (2004)

[95] S. Anhöck, H. Oppermann, C. Kallmayer, R. Aschenbrenner, H. Reichl, and L. Thomas, "Investigations of Au-Sn alloys on different end-metallizations for high temperature applications", International Electronics Manufacturing Technology Symposium IEEE/CPMT, Berlin, pp. 156-62, (1998)

[96] H.G. Song, J.P. Ahn, and J.W. Morris, Jr, "The Microstructure of Eutectic Au-Sn Solder Bumps on Cu/Electroless Ni/Au", Journal of Electronic Materials, Vol. 30, No. 9, pp. 1083-7 (2001)

[97] J. Tjandra, C.L. Wong, J. How, and S. Peana, "Au-Sn Microsoldering on Flexible Circuit", Electronic Packaging Technology Conference, pp. 52-7 (1997)

[98] M. Hutter, G. Engelmann, and H. Oppermann, "Betrachtung von galvanisch erzeugten AuSn-Bumps unter geometrischen Gesichtspunkten und die Auswirkungen auf den Bumpingprozess und die Flip-Chip-Montage", Fraunhofer IZM, white paper (2003)

[99] C. Kallmayer, H. Oppermann, G. Engelmann, E. Zakel, and H. Reichl, "Self-Aligning Flip-Chip Assembly Using Eutectic Gold/Tin Solder in Different Atmospheres", IEEE/CPMT International Electronics Manufacturing Technology Symposium, pp. 18-25 (1996)

[100] Indium Corporation of America, "Application Notes - Fluxless Soldering", white paper, downloaded from www.indium.com in (2004)

[101] J.F. Kuhmann, A. Preuss, B. Adolphi, K. Maly, T. Wirth, W. Oesterle, W. Pittroff, G. Weyer, and M. Fanciulli, "Oxidation and Reduction Kinetics of Eutectic SnPb, InSn, and AuSn: A Knowledge Base for Fluxless Solder Bonding Applications", Transactions on Components, Packaging, and Manufacturing Technology, Part C, Vol. 21, No. 2 (1998)

[102] T. Ishii and S. Aoyama, "Novel Micro-Bump Fabrication for Flip-Chip Bonding", Journal of Electronic Materials, Vol. 33, No. 11, pp. 22-4 (2004)

[103] W. Lin and Y.C. Lee, "Study of Fluxless Soldering Using Formic Acid Vapor", Transactions on Advanced Packaging, Vol. 22, No. 4, pp. 592-601, DOI: 10.1109/6040.803451 (1999)

[104] G.S. Matijasevic and C.C. Lee, "A Reliability Study of Au-Sn Eutectic Bonding with GaAs Dice", Proceedings of the 27th IEEE Reliability Physics Symposium, pp. 137-40 (1989)

[105] C.C. Lee and J. Kim, "Wafer Bonding Using Fluxless Process with Sn-Rich Sn-Au Dual-layer Structure", Proceedings of International Symposium on Advanced Packaging Materials: Processes, Properties and Interfaces, pp. 105-9 (2005)

[106] J. Sasaki, M. Itoh, T. Tamanuki, H. Hatakeyama, S. Kitamura, T. Shimoda, and T. Kato, "Multiple-Chip Precise Self-Aligned Assembly for Hybrid Integrated Optical Modules Using Au–Sn Solder Bumps", IEEE Transactions on Advanced Packaging, Vol. 24, No. 4, pp. 569-75 (2001)

[107] N. Okinaga, H. Kuroda, and Y. Nagai, "Excellent Reliability of Solder Ball Made of a Compliant Plastic Core", IEEE Electronic Components and Technology Conference, pp. 1345-9 (2001)

[108] S. Kröhnert, "Reliability, Temperature Cycling, Mechanical Robustness", presentation at 18th MicroMechanics Europe Workshop (2007)

[109] D. Kim and C.C. Lee, "Fluxless flip–chip Sn Au solder interconnect on thin Si wafers and Cu laminated polyimide films", Materials Science and Engineering A, Vol. 416, pp. 74–79, DOI 10.1016/j.msea.2005.09.084 (2005)

[110] J. Kim, D. Kim., G.L. Wang, J. Park, and C.C. Lee, "Electroplated Sn-Au structures for fabricating fluxless flip-chip Sn-rich solder joints", 54th Electronic Components and Technology Conference, pp. 1642-6, DOI 10.1109/ECTC.2004.1320335 (2004)

[111] R.W. Chuang, D. Kim, J. Park, and C.C. Lee, "A Fluxless Process of Producing Tin-Rich Gold-Tin Joints in Air", IEEE Transactions on Components and Packaging Technologies, Vol. 27, No. 1, pp. 177-81 (2004)

[112] S.-W.R. Lee and J.H. Lau, "Design for Plastic Ball Grid Array Solder Joint Reliability", Circuit World Magazine, Vol. 23, No. 2, pp. 11-3 (1997)

[113] Anon, "International Technology Roadmap for Semiconductors, 2007 Edition, Simulation and Modeling", available at www.itrs.net (2007)

[114] J.-P. Sommer, R. Dudek, B. Michel, E. Noack, and B. Seiler, "Zuverlässigkeit für elektronische Bauelemente mit hoher Packungsdichte", Mikrosystemtechnikkongress, Dresden, VDE Verlag GmbH, Berlin-Offenbach, ISBN 978-3-8007-3061-2 (2007)

[115] J.-P. Sommer, B. Michel, and A. Ostmann, "Numerical Characterisation of Electronic Packaging Solutions based on Hidden Dies", IEEE 6th International Conference on Electronic Packaging Technology, pp. 300-6, DOI 10.1109/ICEPT.2005.1564694 (2005)

[116] Anon, "Scientific Report 2006 - Packaging Reliability", white paper, IMEC (2006)

[117] S. Schönfelder, J. Bagdahn, M. Ebert, M. Petzold, K. Bock, and C. Landesberger, "Investigations of Strength Properties of Ultra-Thin Silicon", 6th EuroSimE Conference, pp. 105-11 (2005)

[118] K.-H. Shin, C.-R. Moon, T.-H. Lee, C.-H. Lim, and Y.-J. Kim, "Flexible wireless pressure sensor module", Sensors and Actuators A, Vol. 123–124, pp. 30–35, DOI 10.1016/j.sna.2005.01.008 (2005)

[119] O. Wittler, H. Walter, R. Dudek, W. Faust, W. Jun, and B. Michel, "Deformation and Fatigue Behaviour of AuSn Interconnects", 8th Electronics Packaging Conference, pp. 297-301, Singapore (2006)

[120] A. Schubert, B. Michel, and H. Reichl, "Materials Mechanics Issues of Smart Card Manufacturing", IEEE/CPMT International Electronics Manufacturing Technology Symposium, pp. 176-81 (1996)

[121] A. Schubert, R. Dudek, B. Michel, and H. Reichl, "Materials Mechanics and Mechanical Reliability of Flip Chip Assemblies on Organic Substrates", International Symposium on Advanced Packaging Materials, pp. 106-9 (1997)

[122] H. Okamoto and T.B. Massalski, "Binary Alloy Phase Diagrams", ASM International, Metals Park, Ohio (1987)

[123] C.C. Lee and R.W. Chuang, "Fluxless Non-Eutectic Joints Fabricated Using Gold–Tin Multilayer Composite", IEEE Transactions on Components and Packaging Technologies, Vol. 26, No. 2 (2003)

[124] H. Okamoto, "Au-Sn (Gold-Tin)", Journal of Phase Equilibria and Diffusion, Supplemental Literature Review: Section III, 1547-7037 ASM International, DOI: 10.1007/s11669-007-9147-1 (2003)

[125] M. Hutter, H. Oppermann, G. Engelmann, and H. Reichl, "High Precision Passive Alignment Flip Chip Assembly Using Self-alignment and Micromechanical Stops", Electronics Packaging Technology Conference, pp. 385-8, DOI: 10.1109/EPTC.2004.1396639 (2004)

[126] L. Coppola, D. Huff, F. Wang, R. Burgos, and D. Boroyevich, "Survey on High-Temperature Packaging Materials for SiC-Based Power Electronics Modules", Procedins of Power Electronics Specialist Conference, pp. 2234-40, DOI: 10.1109/PESC.2007.4342356 (2007)

[127] Indium Corporation, product data sheet for eutectic gold/tin (2008)

[128] B. Predel, "Landolt-Börnstein - Group IV Physical Chemistry (New Series IV/5)", Springer Verlag, Berlin, O. Madelung, ISBN 978-3-540-15516-4 (1996)

[129] J.S. Hwang, "Modern Solder Technology for Competitive Electronics Manufacturing", McGraw-Hill Publishing Co., ISBN 978-0070317499 (1996)

[130] B. Hugsted, L. Buene, T. Finstad, O. Lonsjo, and T. Olsen, "Interdiffusion and Phase Formation in Au/Sn Thin Film Couples with special emphasis on substrate Temperature During Condensation", Thin Solid Films 98, pp. 81-94 (1982)

[131] S. Nakahara, R.J. McCoy, L. Buene, and J.M. Vanderberg, "Room Temperature Interdiffusion Studies of Au/Sn Thin Film Couples", Thin Solid Films, pp. 185 (1981)

[132] G. Matijasevic, "Bonding Technology of Semiconductor Devices and its Characterization using Scanning Acoustic Microscope", dissertation, University of California (1991)

[133] L. Buene, "Interdiffusion and Phase Formation at Room Temperature in Evaporated Gold-Tin Films", Thin Solid Films 47, pp. 159 (1977)

[134] D. Gregersen, L. Buene, T. Finstad, O. Lonsjo, and T. Olsen, "A Diffusion Marker in Au/Sn Films", Thin Solid Films No. 78, p. 95 (1981)

[135] M.-G. Barthès and C. Pariset, "A low energy electron diffraction - Auger electron spectroscopy study of tin on (100) and (111) Au", Thin Solid Films, vol. 77, pp. 305-312 (1981)

[136] P. De Padova, M. Fanfoni, R. Larciprete, M. Mangiantini, and S. Priori, "A synchrotron radiation photoemission study of the oxidation of tin", Journal of Surface Science, No. 313, pp. 379-91 (1994)

[137] J.Y. Tsai, C.W. Chang, Y.C. Shieh, Y.C. Hu, and C.R. Kao, "Controlling the Microstructures from the Gold-Tin Reaction", Journal of Electronic Materials, Vol. 34, No. 2, pp. 182-7 (2005)

[138] G. Gottstein, "Physikalische Grundlagen der Materialkunde", 2. Auflage, Springer Verlag (2001)

[139] J.-H. Park, J.-H. Lee, Y.-H. Lee, and Y.-S. Kim, "Reaction Characteristics of the Au-Sn Solder with Under-Bump Metallurgy Layers in Optoelectronic Packages", Journal of Electronic Materials, Vol. 31, No. 11, pp. 1175-80 (2002)

[140] S. Anhöck, A. Ostmann, H. Oppermann, R. Aschenbrenner, and H. Reichl, "Electroless metal deposition used as underbump metallization (UBM) for high temperature application - a comparison to other bumping techniques", chapter in "Functional Materials" (ISBN 9783527302543), DOI 10.1002/3527607420 (2006)

[141] D. Roundy and M.L. Cohen, "Ideal strength of diamond, Si, and Ge", Physics Review B, Vol. 64, No. 21, DOI 10.1103/PhysRevB.64.212103 (2001)

[142] E. Ventsel and T. Krauthammer, "Thin Plates and Shells: Theory, Analysis, and Applications", published by CRC Press Inc. (2001)

[143] N. Schwarzer, "About the theory of thin coated plates", white paper, Technical University of Chemnitz (ca. 2004)

[144] Website: http://www.asmec.de (2008)

[145] K.-S. Chen, T.Y.-F. Chen, C.-C. Chuang, and I.-K. Lin, "Full-Field Wafer Level Thin Film Stress Measurement by Phase-Stepping Shadow Moiré", IEEE Transactions on Components and Packaging Technologies, Vol. 27, No. 3, pp. 594-601 (2004)

[146] K.E. Petersen, “Silicon as a mechanical material“, Proc. of the IEEE, Vol. 70, No. 5, pp. 420-57 (1982)

[147] T. Randall, "Characterizing the Ductile Response of Brittle Semiconductor Materials to Dynamic Contact Processes", dissertation, North Carolina State University (2004)

[148] A. Haapalinna, S. Nevas, and D. Pähler, "Rotational grinding of silicon wafers—sub-surface damage inspection", Materials Science and Engineering B, Vol. 107, pp. 321–31, DOI 10.1016/j.mseb.2003.12.008 (2004)

[149] C. Landesberger, G. Klink, G. Schwinn, and R. Aschenbrenner, "New Dicing and Thinning Concept Improves Mechanical Reliability of Ultra Thin Silicon", Proceedings of Advanced Packaging Materials International Symposium, pp. 92-7, ISBN 0-930815-64-5 (2001)

[150] L. Kiesewetter, "Mechanische Belastbarkeit von Mikrobauteilen", Exercise, Technical University of Cottbus (2007)

[151] A.F.J. Baggerman, "Technology and Reliability of TAB and Flip-Chip Interconnections", dissertation, Technical University of Delft (1998)

[152] R.R. Chromik, D-N. Wang, A. Shugar, L. Limata, M.R. Notis, and R.P. Vinci, “Mechanical properties of intermetallic compounds in the Au–Sn system”, Journal of Materials Research, Vol. 20, No. 8, pp. 2161-72 (2005)

[153] Y. Cao, S. Allameh, D. Nankivil, S. Sethiaray, T. Otiti, and W. Soboyejo, "Nanoindentation measurements of the mechanical properties of polycrystalline Au and Ag thin films on silicon substrate: Effects of grain size and film thickness", Journal of Materials Science and Engineering A, Vol. 427, pp. 232-40 (2006)

[154] A. Keil, “Werkstoffe für elektrische Kontakte”, Springer Verlag, Berlin (1960)

[155] L.T. Novelty, "Metals handbook - Properties and selection of metals, 8th Edition, Vol. 1", American Society for Metals, Metals Park (1961)

[156] B. Freund, S. Suresh, and L.B. Freund “Thin Film Materials: Stress, Defect Formation, and Surface Evolution”, Cambridge University Press, ISBN 978-0521822817 (2004)

[157] C.-W. Baek, Y.-K. Kim, Y. Ahn, and Y.-H. Kim, "Measurement of the mechanical properties of electroplated gold thin films using micromachined beam structures", Sensors and Actuators A, Vol. 117, pp. 17-27 (2005)

[158] H.D. Espinosa, B.C. Prorok, and M. Fischer, "A methodology for determining mechanical properties of freestanding thin films and MEMS materials", Journal of Mechanics and Physics of Solids, Vol. 51, pp. 47-67 (2003)

[159] T.P. Weihs, S. Hong, J.C. Bravman, and W.D. Nix, "Mechanical deflection of cantilever microbeams: A new technique for testing the mechanical properties of thin films", Journal of Material Research Vol. 3, No. 5, 931-42 (1988)

[160] P. Várguez, F. Avilés, and A.I. Oliva, "Mechanical properties of gold nanometric films onto a polymeric substrate", Journal of Surface and Coatings Technology, Vol. 202, pp. 1556-63 (2008)

[161] A.K. Chinthakindi, D. Bhusari, B.P. Dusch, and P.A. Kohl, "Recrystallization and Intermetallic Formation in Au/Al and Au/Zn Bimetallic Films", Journal of Electronic Materials, Vol. 31, No. 10, pp. 1080-9 (2002)

[162] B. Huang and T.-Y. Zhang, "Microbridge tests: II. On buckled thin gold films", Journal of Micromechanical Engineering, Vol. 16, pp. 134-42, DOI: 10.1088/0960-1317/16/1/018 (2006)

[163] W.H. Safranek, “The properties of electrodeposited metals and alloys”, 2nd edition, American Electroplaters and Surface Finishers Society, Orlando, Florida (1986)

[164] T. Fritz, M. Griepentrog, W. Mokwa, and U. Schnakenberg, “Determination of Young’s modulus of electroplated nickel”, Journal of Electrochimica Acta, Vol. 48, pp. 3029-35 (2003)

[165] Z.-H. Xu and D Rowcliffe, "Deriving mechanical properties of solft coatings using nanoindentation: an application of mechanism-based strain gradient plasticity", Journal of Surface and Coatings Technology, Vol. 157, pp. 231-7 (2002)

[166] F.C. Nix and D. MacNair, The Thermal Expansion of Pure Metals: Copper, Gold, Aluminum, Nickel, and Iron, Physical Review, Vol. 60 (1941)

[167] C.J. Smithells “Metals Reference Book”, Butterworth and Co Ltd, London, ISBN 978-0408706278 (1976)

[168] B. Kebabi, C.K. Malek, and F.R. Ladan, "Stress and microstructure relationships in gold thin films", Journal of Vacuum Science Technology B, Vol. 41, No. 4-6, pp. 1353-55 (1990)

[169] J.W. Dini and H.R. Johnson, “Coefficient of thermal expansion of sulphamate nickel electrodeposits”, Journal of Materials Science 10, pp. 1253-54 (1976)

[170] F.G. Yost, M.M. Karnowsky, W.D. Drotning, and J.H. Gieske, "Thermal Expansion and Elastic Properties of High Gold-Tin Alloys", Metallurgical Transactions A, Vol. 21, pp. 1885-9 (1990)

[171] J. Kim, D. Cho, and R.S. Muller, "Why is (111) Silicon a better mechanical Material for MEMS?", Proceedings of Transducers, Vol. 1, pp. 662-65 (2001)

[172] J.J.Wortman and R.A. Evans, “Young’s modulus, shear modulus, and Poisson’s ratio in silicon and germanium”, J. of. Appl. Physics, Vol. 36, No. 1 (1965)

[173] D.R. Lide, ”CRC Handbook of Chemistry and Physics 84th Edition”, CRC Press, New York (2003)

[174] A.P. Zhernov, ”Lattice constant and coefficient of linear thermal expansion of the silicon crystal. Influence of isotopic composition”, Journal of Low Temperature Physics, Vol. 26, No. 12 (2000)

[175] "Landolt Börnstein - Semiconductors, Volume 41", Springer Verlag (Berlin), ISBN 978-3-540-64070-7 (2001)

[176] V.M. Glazov and A.S. Pashinkin, ”The Thermophysical Properties (Heat Capacity and Thermal Expansion) of Single-Crystal Silicon”, High Temperature, Vol. 39, No. 3, pp. 413–19 (2001)

[177] HD Microsystems, product data sheet for PI2611 (2008)

[178] M. Joodaki, G. Kompa, and H. Hillmer “Thermo-mechanical Stress Analysis and Measurement in Quasi-Monolithic Integration Technology (QMIT)”, IEEE Trans. on Device and Materials Reliability Vol. 5, No. 3, pp. 581-94 (2005)

[179] M. Fukuhara and A. Sampei, "Temperature Dependence of Elastic Moduli and Internal Dilational and Shear Frictions of Polyimide", Journal of Polymer Science B, Vol. 34, pp. 1579-82 (1996)

[180] H.-C. Liou, P.S. Ho, and R. Stierman, "Thickness dependence of the anisotropy in thermal expansion of PMDA-ODA and BPDA-PDA thin films", Journal of Thin Solid Films, No. 339, pp. 68-73 (1999)

[181] ANSYS 11.0 User Manual and Help Guide (2008)

[182] C. Krüger, “Entwicklung temporärer Kontaktelemente für die Mikrosystemtechnik”, dissertation, RWTH Aachen (2003)

[183] A. Buch, "Pure Metals Properties“, ASM International and Freund Publishing House Ltd., Materials Park, Ohio (1999)

[184] A. Schubert, R. Dudek, R. Leutenbauer, R. Daring, J. Kloeser, H. Oppermann, B. Michel, and H. Reichl, "Do Chip Size Limits Exist for DCA?", International Symposium on Advanced Packaging Materials, p. 150-57 (1999)

[185] D. Houdeau, "Bruchmechanische Störtiefenbewertung rückseiten-gedünnter Scheiben aus einkristallinem Silizium", dissertation, Technical University Berlin (1994)

[186] N.V. Rumak, V.V. Khatko, and V.N. Plotnikov, "Structure and Properties of Silicon Dioxide Thermal Films", chapter in N.V. Rumak et al.: "Structure and Properties of Silicon Dioxide Films", pp. 93-100 (1984)

[187] DIN EN 60749-22_2003-12, Method D, "Mechanische und klimatische Prüfverfahren, Teil 22: Kontaktfestigkeit (Bond-strength)" (2003)

[188] 12 MIL-STD-883E, Method 2019.5, "Die Shear Strength" (1987)

[189] C. Kallmayer, D. Lin, J. Kloeser, H. Oppermann, E. Zakel, and H. Reichl, "Fluxless Flip-Chip Attachment Techniques Using the Au/Sn Metallurgy", IEEE International Electronics Manufacturing Technology Symposium, pp. 20-8 (1995)

Appendix

Appendix A: Conventions

Gold/tin phases

In the literature there are many different ways to refer to the gold/tin intermetallic system and its compounds. Typical acronyms are Au/Sn, Au-Sn and AuSn. In this thesis, the term Au/Sn is used to refer to gold/tin alloys and Au-Sn is used to refer to a stack composed of a gold and tin layer. Finally, AuSn is used to refer to the intermetallic compound consisting of 50 at.% Au and 50 at.% Sn.

When looking at certain compositions of gold/tin alloys, the multitude of descriptive terms found in the literature is even greater. All of the following terms refer to the 80 wt.% gold-rich eutectic composition: 80Au20Sn, Au80Sn20, 80Au-20Sn, Au20Sn, Au-20Sn, 80/20 Au-Sn. In this thesis, the term 80Au20Sn is used. Also, if not otherwise specified, the numbers before the chemical symbols refer to weight percentages as opposed to volumetric or atomic percentages.

Gold/tin eutectics

The gold/tin system has 2 eutectic points:

- 80 wt.% Au + 20 wt.% Sn (T_{melt} = 280°C),
- 10 wt.% Au + 90 wt.% Sn (T_{melt} = 218°C).

If not otherwise specified the term *eutectic* refers to the gold-rich eutectic with 80 wt.% Au and a melting temperature of 280°C.

Soldering and bump reflow

When referring to the heating of the bumps of an individual chip to produce a solder cap the term *bump reflow* is used. To describe the joining of a chip and substrate after flip chip, the term *soldering* is used.

Bending direction and bending axis

The term *bending direction* refers to the direction in which the four edges of a chip move during bending, whereas the term *bending axis* refers to the axis around which bending occurs (or in other words: the center axis of the cylinder that the bent chip forms a part of). For example in Figure 0.1 the chip is bent in the z-direction, the axis of bending, however, is the y-axis.

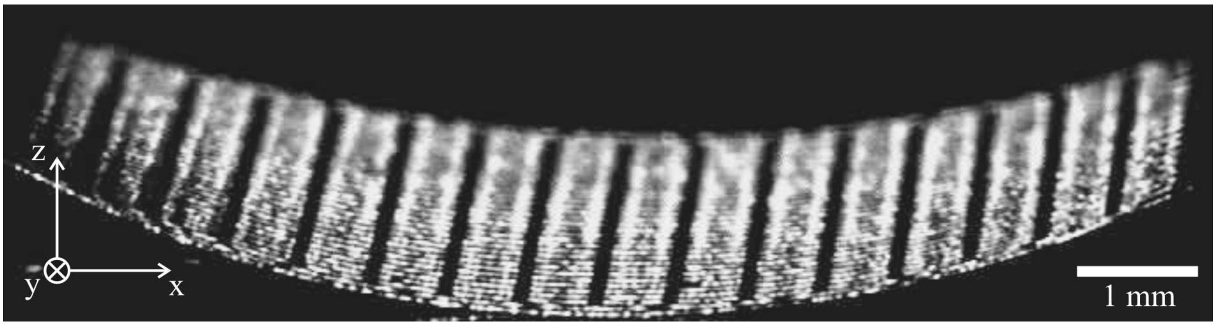

Figure 0.1: Definition of the terms bending direction *and* bending axis. *The chip in this image is bent around the y-axis or, in other words, it is bent in the z-direction.*

Convex and concave deformations

The terms *convex* and *concave* describe the direction of deformation as observed from the backside of the silicon chips. Convex deformations (with the silicon side bulging out) have a negative bending angle, while concave deformations have a positive bending angle.

Miller convention for crystallographic planes and directions

In this thesis, the Miller notation described in Table 0.1 is used.

Table 0.1: Bracket types of the Miller convention used to identify specific directions or planes or equivalent families thereof in crystals.

Bracket type	Used to identify
[ ]	a specific direction
< >	a family of equivalent directions
( )	a specific plane
{ }	a family of equivalent planes

Appendix B: Derivation of the trilateration formula

When comparing the bow of reflowed chips of completely soldered systems, the bending radius is an objective means to quantitatively compare the chip bows. Using trilateration the bending radius of a cylindrically shaped sample can be determined from the coordinates of 3 points on a circle of the cylinder's surface. The coordinates of the circle's center are determined by trilateration and, using these center coordinates, the radius is determined as the distance between the center point and any of the three points of the circle. The coordinates of the three known points of the circle are (x_1, z_1), (x_2, z_2), and (x_3, z_3) and their distance d to the circle center (M_x, M_z) is:

$$\begin{aligned} R &= d\big((M_x, M_z), (x_1, z_1)\big) = \sqrt{(M_x - x_1)^2 + (M_z - z_1)^2} \\ &= d\big((M_x, M_z), (x_2, z_2)\big) = \sqrt{(M_x - x_2)^2 + (M_z - z_2)^2} \qquad (0.1) \\ &= d\big((M_x, M_z), (x_3, z_3)\big) = \sqrt{(M_x - x_3)^2 + (M_z - z_3)^2}\,. \end{aligned}$$

This yields two equations with the two unknowns M_x and M_z:

$$R = \sqrt{(M_x - x_1)^2 + (M_z - z_1)^2} = \sqrt{(M_x - x_2)^2 + (M_z - z_2)^2} \qquad (0.2)$$

and

$$R = \sqrt{(M_x - x_2)^2 + (M_z - z_2)^2} = \sqrt{(M_x - x_3)^2 + (M_z - z_3)^2}\,. \qquad (0.3)$$

These equations can be simplified to

$$M_x = \frac{{x_1}^2 - 2 \cdot M_z \cdot z_1 + {z_1}^2 - {x_2}^2 + 2 \cdot M_z \cdot z_2 - {z_2}^2}{2 \cdot (x_1 - x_2)} \qquad (0.4)$$

and

$$M_x = \frac{{x_2}^2 - 2 \cdot M_z \cdot z_2 + {z_2}^2 - {x_3}^2 + 2 \cdot M_z \cdot z_3 - {z_3}^2}{2 \cdot (x_2 - x_3)}\,. \qquad (0.5)$$

Equating these formulas yields the value for M_z:

$$M_z = \frac{(x_2 - x_3) \cdot ({x_1}^2 + {z_1}^2 - {x_2}^2 - {z_2}^2) - (x_1 - x_2) \cdot ({x_2}^2 + {z_2}^2 - {x_3}^2 - {z_3}^2)}{(2 \cdot z_3 - 2 \cdot z_2) \cdot (x_1 - x_2) + (2 \cdot z_1 - 2 \cdot z_2) \cdot (x_2 - x_3)}\,. \qquad (0.6)$$

Inserting this value for M_z in formula (0.5) yields a value for M_x. From these values for M_x and M_z and the coordinates of one of the originally given points the bending radius R can be calculated using formula (0.1).

Appendix C: Correction factor for bow measurements in samples with a rotated deformation cylinder

The initial optical bow measurement assumes that bending occurs around the y-axis. The actual deformation cylinder is oriented at an angle γ relative to the y-axis in the original chip plane. In Figure 0.2 the deduction of corrected x coordinates is shown for an example with θ = 45°. Using these corrected x values in the trilateration formulas in Appendix B returns the correct radius of curvature for the chip bow.

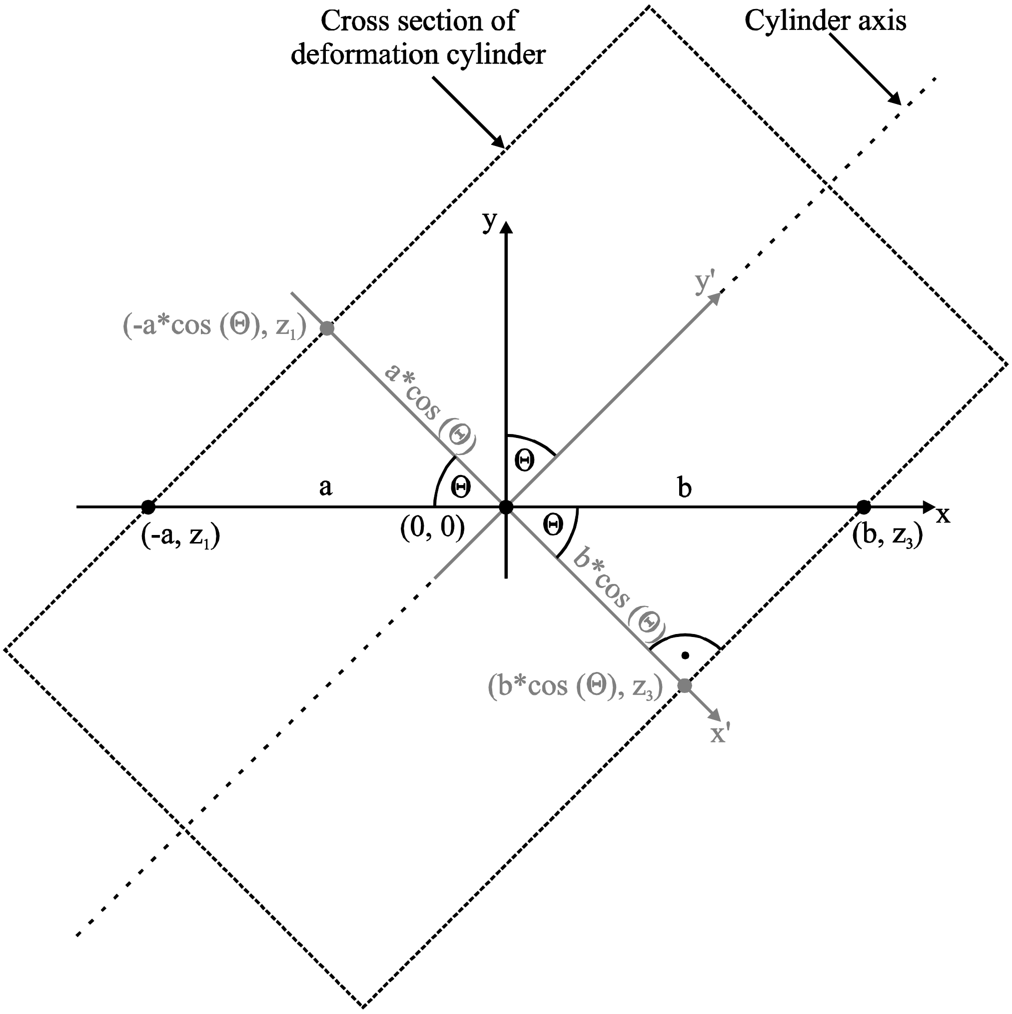

Figure 0.2: Deduction of corrected x coordinates in cylindrically deformed chips for correct bow calculation. The diagram shows a cross section of the deformation cylinder with a θ = 45° angle relative to the chip edges (which are parallel to the x and y axes). The assumed x coordinates of the right and left point are -a and b. The x coordinates in the 45° rotated coordinate system x'-y' are -a·cos(θ) and b·cos(θ).

Lebenslauf

Kaspar Hungar, geboren am 4. Juli 1977 in Köln

1983 - 1987	Grundschule in Köln
1987 - 1996	Humboldt-Gymnasium in Köln, Abschluss Abitur
1996 - 1997	Wehrdienst
1997 - 2003	Studium der Elektrotechnik an der RWTH Aachen, Abschluss Diplom
2003 - 2008	Wissenschaftlicher Angestellter am Institut für Werkstoffe der Elektrotechnik, Lehrstuhl I der RWTH Aachen
Seit 2008	Entwicklungsingenieur bei der Schott Solar AG