Development of Shape Memory Alloy - Si Bimorph Nanoactuators

Zur Erlangung des akademischen Grades
DOKTOR DER INGENIEURWISSENSCHAFTEN
der Fakultät für Maschinenbau
Karlsruher Institut für Technologie (KIT)

genehmigte

DISSERTATION

von

Dipl. Phys. MPhys. Franziska Jasmin Lambrecht
aus Freiburg im Breisgau

Tag der mündlichen Prüfung: 16. Dezember 2016
Referent: Prof. Dr. Manfred Kohl
Korreferent: Prof. Dr. Volodymyr Chernenko

I hereby testify, that the here presented work has been written by myself, has not been used for other assessments, that no other sources than the cited were used, and that citations are identified as such.

Karlsruhe, 17th of October 2016

Acknowledgments

I would like to thank my supervisor Prof. Dr. Manfred Kohl for his support, trust, and encouraging discussions throughout my work. Above all, I am thankful for introducing me to the diverse field of shape memory alloys. I want to thank for providing me the opportunity to present my research on international conferences and exchange experiences with other scientists all over the world. He was always there to guide, when problems occurred and left enough freedom to develop new ideas and concepts.

I am very grateful, that Prof. Dr. Volodymyr Chernenko has kindly agreed to review my thesis and for the discussion and support during the project. Moreover, I want to thank for the collaborative work, including the deposition of the NiMnGa thin films. Without the material of exceptional quality, this work would not have been successful. Further, I want to thank Iván Rodriguez Aseguinolaza for the fabrication of the films, his visit to Germany and his overall support of my research work.

I would like to extend my thanks to the past and present members of our group, who helped me during my daily work regarding concepts and simulations as well as solving minor or major problems in the lab. Further, I would like to emphasize Hinnerk Ossmer and Marcel Gültig, who besides supporting me in any macroscale problem, e.g. soldering or wire-bonding, always had time for interesting discussions. Many thanks go to Sascha Mühlbrandt and Kira Köhnle for their support concerning the optical aspects of my work, in terms of simulation and fabrication. Besides the before listed colleagues, I would like to emphasize the thoroughly enjoyable experience working together with the FuE1-SMD research group.

The technical staff helped throughout countless discussions and supplied immediate support if problems occurred. Especially Alban Muslija, who supported me in developing new dry etching processes, Andreas Bacher and Marie-Christin Gamer without their assistance successful structuring by EBL would not have been possible. I am thankful for all other IMT staff members, helping me through paperwork and all kinds of obstacles.

Finally, I would like to thank my family and friends, in particular Tiemo Bückmann, my parents Helga and Armin, and my siblings Alexander and Rebekka, for their ongoing support.

Abstract

The ongoing trend of miniaturization poses a significant challenge on the integration of functional units on a limited space. Belonging to the group of *Smart Materials*, Shape Memory Alloys (SMAs) generate large work outputs per volume density, especially in small dimensions. Furthermore, they exhibit multifunctional performance and provide biocompatibility. Hence, shape memory alloys are predestined for applications on the nanoscale. This work presents the fabrication, characterization of NiMnGa/Si bimorph double beam nanoactuators and their integration within a Si-photonics on-chip waveguide structure. The design of bimorph structures enables reversible actuation, whereas the double beam layout allows for Joule heating of single devices. A fabrication process is developed based on electron-beam lithography and reactive ion etching using hydrogen bromide. Thereby, it is possible to fabricate NiMnGa/Si devices with critical dimensions down to 50 nm.

Due to the size of the structures the characterization has to be done in a Scanning Electron Microscope (SEM). Therefore, an in-situ measurement setup using nanoactuators is designed and set up inside a SEM, enabling Joule heating of individual double beam structures. Furthermore the setup is extended by the option to homogeneously heat the whole sample. By deflection and electrical resistance measurement, the thermo-mechanical properties of the NiMnGa/Si nanoactuators are investigated within this work.

Furthermore, a FEM model is set up to approximate the behavior of NiMnGa/Si bimorph structures with respect to individual material properties. The model is successfully utilized to model the device performance upon Joule heating as well as homogeneous heating. In addition, the dependency of device characteristics on the thickness ratio of NiMnGa and Si is evaluated.

The presented results show the Shape Memory Effect in the mechanical response of nanostructured bimorph devices for the first time. Furthermore, a vanishing hysteresis in the electrical resistance behavior is observed for decreasing Si layer thickness. The characteristic electrical resistance drop associated with the phase transformation increases for decreasing beam width, which indicates a reduction of non-transforming material. All observed phenomena show the high impact of the substrate, in particular its dimensions, on the device characteristics. By comparison with the simulation, practicable approaches for future actuator designs are proposed such as an optimum thickness ratio of the NiMnGa layer compared to the Si layer. Finally, a design, a simulation model and first measurements on a SMA-based nanoactuator for optical switching are presented. This demonstrates the suitability of SMA devices in optical on-chip applications.

Kurzzusammenfassung

Die Verkleinerung funktionaler Bauteile bietet die Möglichkeit deren Leistungsdichte, sowie deren Geschwindigkeit signifikant zu erhöhen. Zudem kann durch die parallele Herstellung vieler Bauelemente mittels Lithographie-Verfahren die Packungsdichte erhöht, und somit die Herstellungskosten pro Bauteil reduziert werden. Die Errungenschaften, die beispielsweise in der Mikroelektronik erzielt wurden, haben unsere heutige Informationsgesellschaft geprägt, bzw. erst möglich gemacht. Der stetige Trend hinzu immer kleineren Bauteilgrößen ist daher nach wie vor ungebrochen, jedoch stößt die Miniaturisierung auf Grund der steigenden Komplexität langsam an ihre Grenzen. Eine Option der Komplexitätssteigerung entgegen zu wirken, ist die Verwendung von intelligenten Werkstoffen, sog. Smart Materials. Der Begriff Smart Materials bezeichnet multifunktionale Materialien, deren Eigenschaften sich durch die Veränderung von äußeren Einflüssen wie Wärme, Druck, einem Magnetfeld oder elektrischer Spannung steuern lassen. Im Bereich der Aktorik wird dadurch der Einsatz von komplexen Stellmotoren überflüssig - das Material ist die Maschine. Dies ermöglicht das Vordringen in nanoskalige Bauteilgrößen. Auf Grund ihrer hohen Energiedichten eignen sich insbesondere Formgedächtnislegierungen für Anwendungen in der Nanoaktorik, da sie auf Grund ihrer Kristallstruktur die außergewöhnliche Eigenschaft der Gestalterinnerung aufweisen.

Ziel dieser Arbeit ist die Entwicklung von Nanoaktoren auf Basis der Formgedächtnislegierung (FGL) NiMnGa und die Untersuchung ihrer thermomechanischen Eigenschaften. Des Weiteren soll durch die Kombination von FGL-Bauteilen mit optischen Komponenten ein Mikrosystem entworfen werden. Um reversiblen Aktuierung in Kombination mit der Möglichkeit elektrische Widerstandmessung einzelner Bauteile durchzuführen, wird eine Bimorph-Doppelbalken Geometrie gewählt. Die Verwendung von Si als Basismaterial eröffnet ein möglichst breites Anwendungfeld, insbesondere die Integration mit optischen Bauteilen ist vereinfacht. Mit dem entwickelten Prozessablauf, bestehend aus Elektronenstrahllithographie und reaktiven Ionen Ätzen, werden NiMnGa/Si Strukturen mit unterschiedlichen Abmessungen hergestellt, bis hinzu kleinsten Abmessungen von 50 nm bei Balkenlängen von bis zu 9 µm. Zur Charakterisierung wird ein In-Situ-Messsystem entwickelt, welches in der Lage ist einzelne Doppelbalken mittels direkten Ohm'schen Heizen oder durch homogene Temperaturerhöhung des gesamten Substrates zu aktuieren. Die mechanische Reaktion der NiMnGa/Si Strukturen wird mittels Rasterelektronenmikroskopie beobachtet. Um Rückschlüsse auf intrinsische Materialeigenschaften zu ziehen, wird ein gekoppeltes Finite-Elemente-Simulationsmodell erstellt. Dies ist in der Lage das Verhalten der Formgedächtnis-Bimorphe bei ansteigender Heizleistung als auch bei homogener Temperaturvariation vorherzusagen. Hierbei wird ein konkurrierendes Verhalten des Bimetall-Effekts, basierend auf den unterschiedlichen ther-

mischen Ausdehnungskoeffizienten von NiMnGa und Si, und des Formgedächtniseffekts identifiziert.

Die präsentierten Ergebnisse zeigen erstmals den Formgedächtniseffekt im mechanischen Verhalten nanostrukturierter Bimorph Strukturen. Desweiten wird ein nahezu vollständiges Verschwinden der Hysterese gefunden. Insbesondere für Anwendungen in der Sensorik stellt die hysteretische Eigenschaft von Formgedächtnismaterialien einen großen Nachteil da. Eine Reduktion dieser Eigenschaft ab einer gewissen Bauteilgröße würde daher insbesondere in diesem Bereich eine Vielzahl von neuen Anwendungsmöglichkeiten eröffnen. Der Vergleich der Widerstandkennlinien kleiner werdender Aktorstrukturen zeigt Anzeichen für eine Verringerung des Restaustenitsanteils mit reduzierter Baugröße. Der Restaustenitanteil beschreibt den Anteil an Formgedächtnismaterial der nicht an der Phasenumwandlung teilnimmt, sondern in der Hochtemperturphase verharrt. Eine Verringerung des Restaustenitanteils geht somit mit einer Zunahme an umwandelndem Material einher, was wiederum die Effektgröße und somit die Leistung des Bauteils positiv beeinflusst. Neben grundlagenwissenschaftlichen Aspekten dieser Arbeit wird die Eignung der entwickelten Nanoaktoren als optisches Schaltelement gezeigt.

Publications

Parts of this thesis have already been published:

In scientific journals

- H. Ossmer, F. Lambrecht, M. Gültig, C. Chluba, E. Quandt, and M. Kohl, "Evolution of Temperature profiles in TiNi films for elstocaloric cooling", *Acta Materialia* **81**, 9-20 (2014)
DOI: 10.1016/j.actamat.2014.08.006
- F. Lambrecht, I.R. Aseguinolaza, V. Chernenko, and M. Kohl„ "Integrated SMA-based NEMS actuator for optical switching", *Proceedings IEEE International Conference of Micro Electro Mechanical Systems (MEMS)* **29**, 79-82 (2016)
DOI: 10.1109/MEMSYS.2016.7421562
- H. Ossmer, F. Wendler, M. Gültig, F. Lambrecht, S. Miyazaki, and M. Kohl, "Energy-efficient miniature-scale heat pumping based on shape memory alloys", *Smart Materials and Structures* **25**, 085037 (2016)
DOI: 10.1088/0964-1726/25/8/085037
- F. Lambrecht, C. Lay, I.R. Aseguinolaza, V. Chernenko, and M. Kohl, "Ni-Mn-Ga/Si Shape Memory Bimorph Nanoactuation", *Shape Memory and Superelasticity* **2** (4), 347-359 (2016)
DOI: 10.1007/s40830-016-0080-1

On scientific conferences (presented by myself)

- F. Lambrecht, C. Lay, I.R. Aseguinolaza, V. Chernenko, M. Kohl, 41th Micro and Nanoengineering, The Hague, *"Fabrication of Shape Memory / Silicon Bimorph Nanoactuators"*, (2015)
- F. Lambrecht, I.R. Aseguinolaza, V. Chernenko, M. Kohl, IEEE 29th International Conference of Micro Electro Mechanical Systems (MEMS), Shanghai,*"Integrated SMA-based NEMS actuator for optical switching"*, (2016)

Contents

Acknowledgments v

Abstract vii

Kurzzusammenfassung ix

Publications xi

Introduction 1

1 Materials & Technology 5
1.1 Shape Memory Alloys 5
1.1.1 Phenomenology 5
1.1.2 NiMnGa 9
1.1.3 Scaling and Size effects 11
1.1.4 Constitutive Models 15
1.2 Nanoactuator for Optical Switching 16
1.3 Fabrication 19
1.3.1 Patterning Techniques 19
1.3.2 Thin Film Deposition 28

2 Nanofabrication 33
2.1 Nanomachining after film deposition 35
2.2 Nanomachining prior to film deposition 37
2.3 Fabrication of SMA-based Optical Switch 42

3 Measurement Setup 47
3.1 Temperature Gradient 47
3.2 Homogeneous Temperature 49
3.3 Optical Setup 53

4 Simulation 55
4.1 Structural Mechanics 55
4.2 Electromagnetics 66
4.3 Wave Optics 75

5 Experimental Results 81
5.1 Electrical Performance 81
5.1.1 Temperature Gradient 81

5.1.2 Homogeneous Temperature 91
5.2 Mechanical Performance 95
5.3 Optical Performance 104

6 Conclusion & Outlook **107**
6.1 Conclusion 107
6.2 Outlook 108

Appendix **113**

Bibliography **115**

List of Abbreviations

λ	Wavelength
ξ	Proportion of martensite
A_F	Austenite finish temperature
A_S	Austenite start temperature
CCP	Capacitive Coupled Plasma
EBL	Electron Beam Lithography
EBPVD	Electron Beam Physical Vapor Deposition
F	Size effect correction factor
FSMA	Ferromagnetic Shape Memory Alloy
HF	Hydrofluoric acid
HBr	Hydrogen bromide
IBE	Ion Beam Etching
ICP	Inductively Coupled Plasma
Lo	Lorentznumber
MEMS	Micro Electro-Mechanical System
M_F	Martensite finish temperature
M_S	Martensite start temperature
NiMnGa	Alloy composed of the rare metals Nickel, Manganese and Gallium
NEMS	Nano Electro-Mechanical System
RIE	Reactive Ion Etching
SEM	Scanning Electron Microscope
SF_6	Sulfur hexafluorid
SMA	Shape Memory Alloy
SME	Shape Memory Effect
SOI	Silicon-on-Insulator
TE	Transverse Electric
TM	Transverse Magnetic

Introduction

The potential of miniaturization has been demonstrated by the achievements in microelectronics of smaller and smaller devices providing higher functionality, higher efficiency at yet lower fabrication costs. As a consequence, the aim of further miniaturization towards the nanoscale is unabated. The concepts to approach nanometer-sized devices can be categorized into two groups: One the one hand, downscaling, and on the other hand exploring completely new concepts. The first provides the advantage, that the schedule is quite simple: making every single part of the device smaller results in a smaller final product. Even though the implementation of this schedule in practice is quite challenging anyway, it is also limited due to the increase in complexity. For example the manufacturing of an overall nanometer-sized electromagnetic motor would obviously demand for below nanometer-sized components. Going off the beaten path can mean to utilize biological or chemical processes. Just few days ago, 10/05/2016, the Nobel Prize in chemistry was given to three scientists *for the design and synthesis of molecular machines.* Their research included the development of an electrically driven nanocar consisting of a molecule with four functional units which can be moved in a preferred and linear direction across a surface [1]. As extraordinary as these results are, they are quite far from actual application, since for instance the experimental setup demands for scanning tunneling microscopy at a temperature of 7 K.

An alternative is to use intrinsic material properties. This world provides several smart materials, which exhibit a strong coupling between thermal, electric, magnetic, or mechanical properties. The direct coupling of these properties enables a highly simplified setup compared to external coupling of conventional materials, since the fundamental mechanisms of smart materials take place at atomic length scales. A common quantity to compare different material classes in terms of their suitability as actuators is the energy density, defined by the work output per unit volume. Figure 0.1 presents typical ranges of actuation stress, strain and energy density values for materials exhibiting direct coupling. Here, the energy density is defined as the product of actuation strain and stress, assuming that the material is operated under constant stress [2].

It can be clearly seen that Shape Memory Alloys (SMAs) provide the highest energy densities in the order of $10^7\,\mathrm{J/m^3}$. Moreover theoretical studies and preliminary experiments suggest that SMAs can act as machines themselves, or as stated by Bhattacharya and James "*The Material Is the Machine*" [3]. Combined with their favorable scaling behavior, this makes SMAs one of the most promising candidates for small-scale actuation applications. In the micrometer regime, SMA-based microelectromechanical systems (MEMS) have already been established [4]. Further size reduction broadens the range of applications, but also sets new challenges, that arise from the complexity of fabrication and device design. Besides these *experimental* challenges,

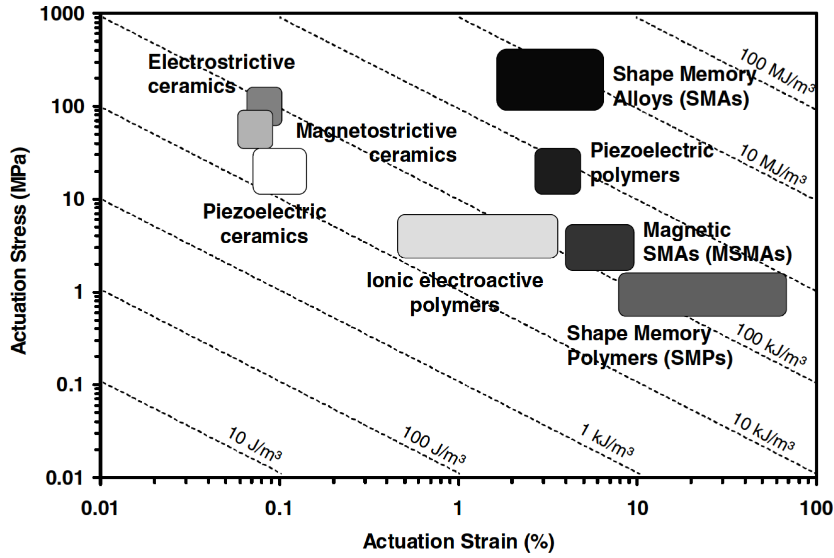

Figure 0.1: Actuation energy density diagram of different active materials that exhibit direct coupling [2].

the question is, whether fundamental physical constraints limit the downscaling of SMA-based devices. The general motivation and the starting point of this thesis is, to find the answers to this question.

The goal of this work is the fabrication of SMA-based actuators with decreasing size down to nanometer dimensions. Furthermore, their functionality needs to be investigated in terms of their mechanical, thermal and electrical properties. For this purpose, an in-situ measurement setup has to be developed. To examine substrate constraints and to identify most important design parameters, a simulation model is needed. Moreover, the effect of downscaling the critical dimensions of the fabricated actuator devices will be investigated. Finally, the developed actuator will be integrated in an optical chip layout to investigate its functionality in a state-of-the-art application.

Outline of this thesis

In Chapter 1, the underlying principle of the Shape Memory Effect (SME) and the basics of the used fabrication techniques are introduced. Thereby, the phenomenology of SMAs is described with the focus on NiMnGa. The effects of downscaling are discussed and the term *size effect* is defined in reference to the term *scaling*. Furthermore, the numerical methods used to model SMAs are introduced. After that, actuation concepts for optical switching are introduced. In the third part, the experimental methods used throughout this work are depicted in short.

In Chapter 2, the fabrication of NiMnGa/Si nanoactuators is presented. This includes the choice of an actuator layout and the presentation of two alternative fabrication techniques, before the fabrication of the developed SMA-based optical switch is illustrated.

The in-situ measurement setup is presented in Chapter 3. As the SME refers to the thermal actuation principles, the realisation of applying heat to the devices is a key challenge. In this work, two alternative concepts are introduced and compared. In the last section, the setup for measuring the optical performance is described.

A finite element model is set up to approximate the mechanical, electrical and optical performance of the NiMnGa/Si actuators. The details are described in Chapter 4, including the influence of changed geometry and materials parameter of both constituents.

In Chapter 5, the experimental results are presented and compared with the FEM model prescriptions.

Finally, the work is concluded and future applications are discussed in Chapter 6.

1 Materials & Technology

1.1 Shape Memory Alloys

Shape Memory Alloys (SMAs) belong to the group of active multifunctional materials, which exhibit a mechanical response to a non-mechanical field one or more orders of magnitude larger compared to standard material behavior. Within this section, the phenomenology of phase transformations in shape memory alloys is discussed at first with regard to the crystallography of the transformation. Thereafter, important terms and definitions are introduced and finally, the key features of the material focused within this work are presented.

1.1.1 Phenomenology

All SMAs have in common that they have (at least) two stable solid phases, a high temperature phase called austenite and a low temperature phase called martensite which differ in crystal structure. The thermally-induced first-order phase transformation between austenite and martensite is enabled by shear lattice distortion rather than diffusion. In SMAs the dominant mechanism for the martensitic transformation is twinning, i.e. lattice-invariant reversible self-accommodation by a small cooperative movement of the atoms [2, 5]. The crystal lattice of the austenite is higher in symmetry, generally having a cubic structure, whereas the martensite can have, e.g., tetragonal, orthorhombic or monoclinic structure. Due to the less symmetric structure of the martensite, for instance a reflection across a crystal plane does not restore the structure, but results in another orientation of the initial structure. Consequently, there are multiple configurations of the martensite lattice, which are called variants. In a macroscopic specimen this leads to a domain structure, in which each domain consists of several unit cells of one martensitic variant.

To estimate the size change associated with the phase transformation, a single two-dimensional unit cell is considered, as shown in Figure 1.1. The austenite unit cell is illustrated by a unit square with the lattice parameter, e.g. the side length, a. After the transformation to martensite, the unit cell is changed, which is illustrated by a parallelogram with the side length a′. Associated with the change from a $\rightarrow$ a′ is a length change ΔL_1. This size change ΔL_1 is due to the difference in lattice parameters. Neglecting its direction, ΔL_1 is equal for the different orientations of the martensite crystal lattice, e.g. equal for the different martensite variants. By forming an assembly of four unit cells, the austenite structure is given by default, whereas the martensite structure is not

exactly defined. Due to the different possible orientations of the martensite unit cell, an additional degree of freedom arises concerning the assembly of the variants. Assuming no additional boundary conditions, the variants will self-accommodate resulting in a so called *twinned* assembly. The twinned assembly is characterized by the amount of martensitic crystals having a certain orientation direction being equally averaged over the whole sample. In the picture of a single crystal consisting of four unit cells, this, for instance, refers to two unit cells of variant 1 and two unit cells of variant 2. As a consequence of this self-assembly mechanism, the resulting length change is not four times ΔL_1 but in the simplest case $\Delta L_4 = \Delta L_1$.

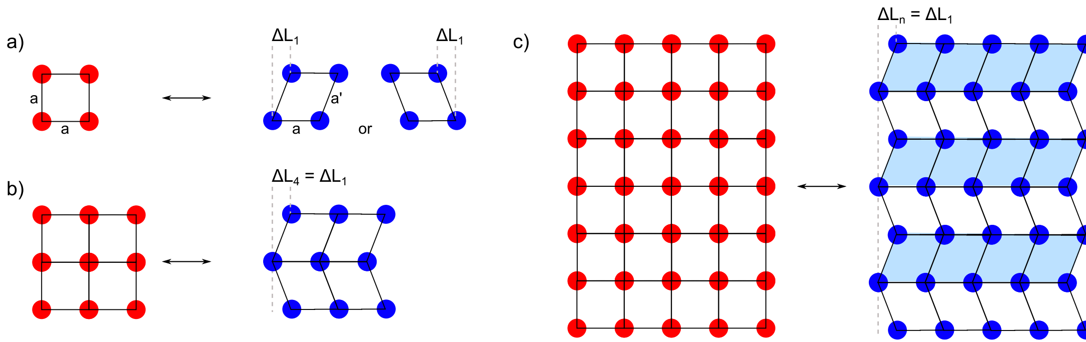

Figure 1.1: Schematic illustration of the shape change associated with the phase transformation for different dimensions. a) single unit cell, single crystal with b) four unit cells and c) n unit cells.

Considering a macroscopic single crystal with n unit cells, the same mechanism will lead to the formation of domains with martensite variant 1 next to domains with martensite variant 2. The boundary between the arising domains is called *twin boundary*. The resulting shape change due to the phase transformation from austenite to twinned martensite is limited by the domain size. As illustrated in 1.1 c), the occurring length change ΔL_n would be equal to ΔL_1 even for a macroscopic specimen, if the domain size is as small as one unit cell. Since ΔL_1 is directly correlated to the lattice parameter mismatch a' to a, it is negligible. To induce a significant shape change, the self-assembled martensite structure has to be aligned. This can be achieved by applying a mechanical load, which makes one variant more favorable than the others. As a result, the self-assembled martensitic structure is manipulated, resulting in a structure with one dominant variant. This state is called *de-twinned* martensite and is characterized by a macroscopic shape change. Indeed, a number of studies suggest, that the maximum shape change possible is reached by a good matching of the unit cell edge lengths [6, 7].

If a block of de-twinned martensite is now again heated up, it will transform into austenite and undergo a complete shape recovery. This process is referred to as the Shape Memory Effect (SME). More precisely, the described behavior is denoted as the **one-way SME**. The whole process is shown schematically in Figure 1.2 including the corresponding shape change.

In detail, the process can be divided into four parts: 1. Cooling, 2. loading, 3. unloading and 4. heating. Starting from a sample being completely in the austenitic state (A) the structure is cooled down into a twinned martensitic state (B). During this phase

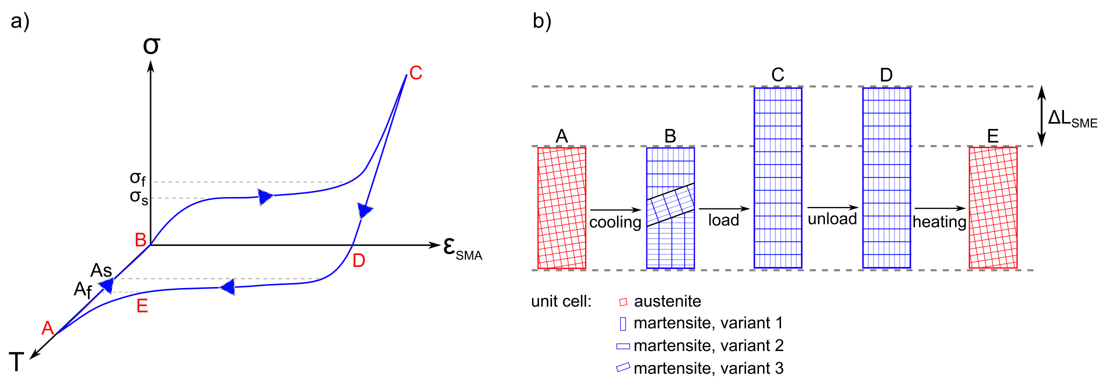

Figure 1.2: Representative thermomechanical loading path for the one-way shape memory effect. a) Combined stress-strain-temperature diagram. σ is the uniaxial stress with the corresponding strain ϵ_{SME}, being the change in length $\Delta L/L$ of the specimen along the direction of applied load normalized by its overall length L. b) Illustration of the shape change induced by de-twinning.

change no significant size change occurs. By applying a mechanical load, the sample is exposed to an increasing value of stress, being elastically deformed. Upon a critical stress level σ_s, the de-twinning of the martensite variants starts, resulting in a large size change ΔL_{SME}. Because of the analogy to plastic deformation in terms of the emerging size change in conventional metals, the behavior just described is often referred to pseudo-plastic deformation. The term pseudo is used, because the deformation is reversible. Beyond σ_f the whole sample is in its de-twinned state (C) and continues with elastic behavior. In the third section, the mechanical load is released, which is associated with no crystallographic reorientation - the sample stays de-twinned martensite (D). Finally, the structure is heated up. Corresponding to the critical stress value, a critical temperature has to be reached to start the transformation from martensite to austenite. This temperature is denoted as the austenite start temperature A_s. From this temperature on, the unit cells start to deform into the crystal structure of the austenite. At the austenite finish temperature A_f all unit cells are transformed back in the austenitic state and the whole deformation ΔL_{SME} has been resetted. The strain recovered due to the phase transformation from de-twinned martensite to austenite is called transformation strain ϵ_t. By subsequent cooling, the austenite will again transform into twinned martensite, therefore the whole cycle can be repeated up to 100 % reversibility. According to the already mentioned transformation temperatures A_s and A_f associated with the transformation from martensite to austenite, there are also two characteristic temperatures associated with the reverse transformation. The transformation from austenite to martensite begins at the martensite start temperature M_s and completes at the martensite finish temperature M_f [8]. In general, the martensitic transformation is associated with a hysteresis, hence four characteristic transition temperatures, $A_{S,F}$ and $M_{S,F}$. However, latest results show near-zero hysteresis for tailored lattice parameters of quaternary SMAs [9].

Up to this point, only heating and cooling under zero applied load have been considered. In fact, these four characteristic transformation temperatures strongly depend on the

applied load. With increasing stress values, irrespective of whether loading is tensile or compressive, the transformation temperatures will increase as well. Thus, the transformation from austenite to martensite cannot only be thermally induced but also by applying a sufficiently high mechanical load. The shape recovery upon stress-induced phase transformation is called **pseudo-elastic effect**. If the operation temperature is below the austenite start temperature, the shape is only partly resetted. On the other hand, if the temperature is too high,$T \geq T_{yield}$, the critical stress value for the pseudo-plastic effect can exceed the yield stress value, resulting in permanent and non-reversible deformation. Figure 1.3 shows the loading path including the corresponding shape change.

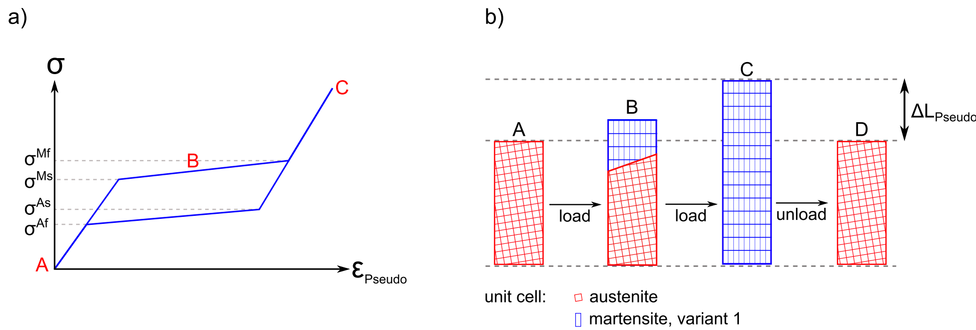

Figure 1.3: a) Representative stress-strain diagram of a pseudo-elastic material for temperatures $A_f < T < T_{yield}$ and b) corresponding crystallographic variations causing the shape change.

If the temperature is above A_f and below T_{yield}, the crystal structure of the specimen is completely austenitic (A). By applying a mechanical load, the sample will at first deform elastically up to a critical stress value σ_{Ms}. Once this value is reached, the transformation into martensite, more precisely into de-twinned martensite, is initiated. In the arising de-twinned martensite phase, the variant, which is oriented along the direction of the applied load will be dominant. Therefore, small load changes will lead to large deformations during the reorientation period, e.g. large inelastic strains. The boundary between material in the austenitic state and material in the martensitic state is called *phase boundary*. At σ_{Mf} the whole material is martensitic and will deform elastically upon any further loading (B). The reverse transformation upon unloading is characterized by σ_{As}, the stress value at which the transformation to austenite starts, and σ_{Af} at which the transformation is complete.

For actuation applications, a cyclic behavior with no externally applied mechanical load is of major interest. In SMAs this behavior is called **two-way SME** and can be enabled by repeated thermo-mechanical cycling, i.e. training [10]. The repetition for a large number of cycles can cause lattice defects. These changes in the microstructure can lead to macroscopic effects in the material behavior. If an untrained specimen is repeatedly loaded following a defined loading path, the hysteretic response will change from cycle to cycle, since a certain amount of plastic strain is generated, which leads only to partial recovery. When the inelastic strain saturates, the difference between two cycles vanishes. The introduced defects during the training favor a certain orientation of the martensitic unit

cell. As a consequence, the SMA will not transform into a self-accommodated martensitic structure after cooling but in an already aligned, e.g. twinned state. The shape change between twinned martensite and austenitic structure is therefore enabled during pure thermal cycling. However, the defects also disturb the lattice distortion needed for the pseudo-plastic deformation, resulting in significantly lower strain values for the two-way SME.

1.1.2 NiMnGa

NiMnGa belongs to the group of Heusler alloys named after F. Heusler, who discovered ferromagnetic properties in Cu_2MnSn compounds. In these alloys, the ferromagnetism is a result of the double-exchange mechanism of neighbouring magnetic atoms, and not directly derivable from the individual properties of the constituents [11]. After their discovery in 1903, several Heusler alloys showing temperature-dependent ferromagnetic spin-alignment were identified. In NiMnGa, both martensite and austenite phase are ferromagnetic, and a ferro- to paramagnetic transition takes place at the Curie temperature T_C. In austenitic state, NiMnGa exhibits the face-centered-cubic L21 crystal structure, as illustrated in Figure 1.4 a), which would be simply a body-centered-cubic lattice, if all atoms were identical. The martensitic phase can be described as a tetragonal distortion of the parent phase with an additional superstructure with long periodicity along the c axis. The periodicity can be equal to five atomic layers, commonly abbreviated as 10M, but also seven-layered orthorhombic (14M) or non-modulated tetragonal martensite (NM) have been found [12–14].

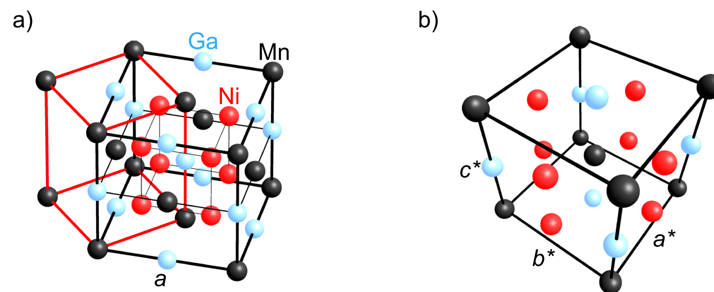

Figure 1.4: Crystal structure of the Heusler alloy NiMnGa. a) Austenitic phase having face-centered-cubic L21 crystal structure. b) Illustration of the tetragonal configuration of the martensitic phase. (compare Ref. [15])

In Section 1.1.1 the crystallographic changes associated with the martensitic phase transformation and their impact on the size of the structure have been discussed. Besides the shape of the specimen, nearly all material properties are linked to the crystal lattice. Therefore, parameters like the electrical conductivity are an indicator for the course of phase transformation. Figure 1.5 a) shows the representative change in electrical resistance of a common SMA film as a function of temperature. In the beginning, the material behaves like a common metal, e.g. the resistance increases with increasing temperature. After the start of the transformation from martensite to austenite, the

slope changes significantly and the resistance drops, until the transformation is completed. From this point on, the material is in its austenitic state and again the resistance increases linearly with temperature. The reverse transformation, e.g. from austenite to martensite upon cooling, is characterized by the martensite start and finish temperature. The characteristic transformation temperatures are determined by the tangential method, indicated in Figure 1.5 a), making the temperature-dependent electrical resistance a standard and easy to handle method to characterize SMAs.

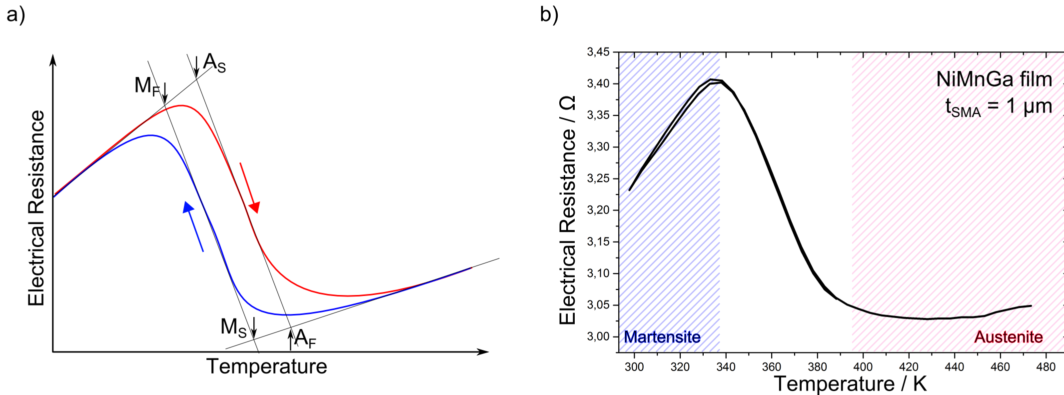

Figure 1.5: a) Schematic of the electrical resistance course upon heating (red) and upon cooling (blue) of a SMA. With the use of the tangential method the characteristic transition temperatures can be determined. b) Electrical resistance versus temperature of a NiMnGa film with the nominal layer thickness of 1 µm and the stoichiometric composition $Ni_{53.4}Mn_{26.4}Ga_{20.2}$. The austenitic phase (red) and the martensitic phase (blue) are highlighted.

By comparing the representative resistance curve, shown in Figure 1.5 a), with the corresponding data for an 1 µm thick NiMnGa film, shown in Figure 1.5 b), the most remarkable feature is the almost vanishing difference between the heating and the cooling behavior. The hysteresis, defined by the difference in transformation temperatures upon cooling compared to heating, is almost zero, making the NiMnGa film a highly promising candidate for actuation applications.

In addition to the electrical properties, NiMnGa exhibits also magnetic properties. Below the Curie temperature T_C, which is 376 K for the stoichiometric composition [16], NiMnGa is ferromagnetic and above T_C it is paramagnetic. The transition temperature is well below the Curie temperature for all presented morphologies, enabling spontaneous magnetization of the martensitic variants at zero applied field, since the local magnetization vector in each variant is oriented along one preferred crystallographic direction, referred to as the magnetic easy axis. By applying an external field, the local magnetization vectors align with the direction of the applied field. This alignment can be achieved by magnetic domain wall motion, magnetization vector rotation and by magnetic field-driven reorientation of martensitic variants. The latter is unique for ferromagnetic SMAs (FSMAs) and results in a macroscopic shape change. The process of magnetic field-induced shape memory effect is comparable to the pseudo-elastic effect and is illustrated in Figure 1.6. Instead of loading the sample mechanically, a *magnetic load* is applied. As already mentioned, the magnetization of a martensitic FSMA specimen is not equal to zero, because the individual

magnetic moments orient along each other and sum up to a macroscopic magnetization value (A). If an external magnetic field is applied, the magnetic moments align with the direction of the external field. If the applied magnetic field is perpendicular to the easy axis of the initial martensitic variant, the individual magnetic moments in the unit cells slightly rotate, which is the normal behavior of any ferromagnet (B). Upon a critical magnetic field strength, denoted as H_2 in Figure 1.6, the force is high enough to enable the reorientation by twin boundary motion. The initial variant with the easy axis pointing perpendicular to the direction of the external field transforms into a variant with its easy axes oriented parallel to the applied field (C) until the whole sample is in a new martensitic state (D). The reorientation of variants changes the shape of the specimen. This length change, denoted as $\Delta L_{magnetic}$ in Figure 1.6, is known as the magnetic field-induced strain (MFIS).

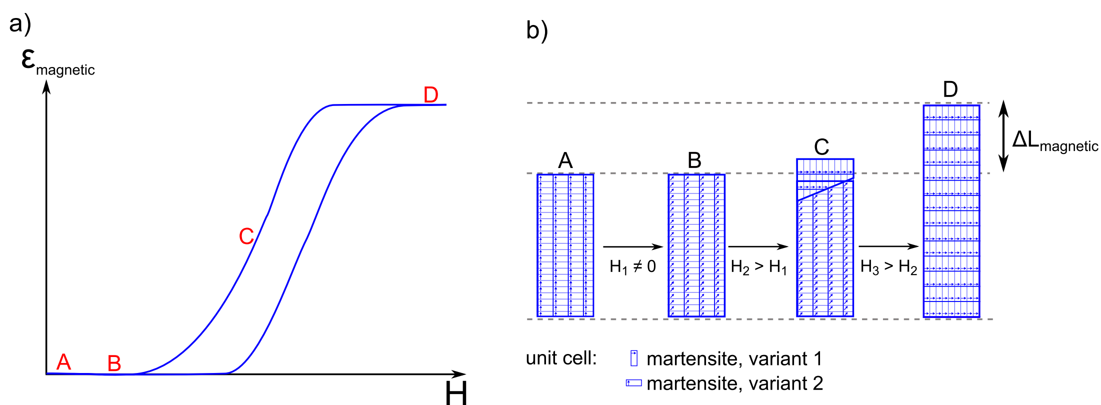

Figure 1.6: a) Representative magnetic loading path and b) schematic of the corresponding crystallographic transformations. The magnetic-field induced length change $\Delta L_{magnetic}$ occurs by applying a magnetic field H perpendicular to the magnetic easy axes. The local magnetization is represented by the blue arrows. After reaching a sufficient magnetic field, here H_2, a new variant grows at the expense of the initial variant with its easy axis along the magnetic field direction.

Even though FSMAs also show the conventional temperature or stress-induced SME, the magnetic field-induced SME is not limited by heat transfer times, enabling much higher frequency ranging up to 1 kHz [17]. Additionally, values up to 12 % for the magnetic field-induced strain have been reported [18], which exceeds the recoverable strain reported in magnetostrictive materials [19, 20] and also the electric field induced-strains observed in piezoelectric materials [21] by far.

1.1.3 Scaling and Size effects

Reducing the characteristic length of an actuator device results in well-known changes of the performance and other changes, which are a matter of present-day research. The first describe the proportionality between an input value and the resulting output quantity. This simple relation is summed up under the term **Scaling**. To give an example, the area of a circle A is given by the equation $A_{circle} = \pi r^2$. Reducing the input value,

here the radius r, by a factor of two, results in a change in area A (output value) by a factor of 4. By introducing a scaling variable Λ describing the specimen size, the corresponding scaling law can be derived. For the area of a circle, it is equal to $A \sim \Lambda^2$. According to this simple example, a scaling law can be defined for each actuation principle, which helps to identify actuation principles more and less suitable for applications in the nano regime. In Table 1.1 the scaling behaviors of the forces of most common actuation principles are presented [22]. Here, the term 'thermal' sums up actuation by thermal expansion. Commonly, heat is generated from electrical power input, hence in an active manner, whereas cooling takes place passively by conduction or convection. Depending on the materials thermal conductivities and the surface of the device, the temperature is transferred. Optical actuation does not denote indirect thermal actuation by optical heating, but the effects of radiation pressure and transverse gradient force. The first acts along the propagation direction, whereas the second acts transversely to the propagation direction. For both, actuation has been shown, however, the dimension of actuated specimen is normally in the nanoscopic regime.

Table 1.1: Scaling behavior of most common actuation principles, with F being the generated force and Λ being the scaling factor.

Principle	Scaling behavior
electrostatic	$F \propto \Lambda^2$ [23]
magnetostatic (Lorentz force)	$F \propto \Lambda^4$ [24]
magnetostatic (Magnetic field gradient)	$F \propto \Lambda^3$ [25]
thermal	$F \propto \Lambda^2$ [26–28]
inverse piezoeffect	$F \propto \Lambda^2$ [29]
optical	$F \propto \Lambda^2$ [30, 31]

As the SME follows direct thermal principles, its scaling behavior is indeed favorable, especially since the reduced thermal mass enables higher frequencies compared to macroscopic applications [32].

Besides these known scaling issues, the probability of arising **Size effects** increases with decreasing dimensions. The term size effects sums up all effects which occur only beneath a critical size and can be again divided into two groups: extrinsic and intrinsic size effects [33]. Extrinsic size effects arise due to geometrical constraints, like a high surface-to-volume ratio. The geometry leads to differences in the behavior of the specimen compared to bulk materials, which do not follow the scaling law dependency. For example, uniaxial compression and tension experiments on cylindrical metal nanopillars have shown a size effect in their plastic deformation behavior, which can be related to the increase of surface-to-volume ratio. With decreasing diameter, the flow stress increases. The data of recent single-crystalline nanopillar experiments is shown in Figure 1.7 and reveals the significance of the trend.

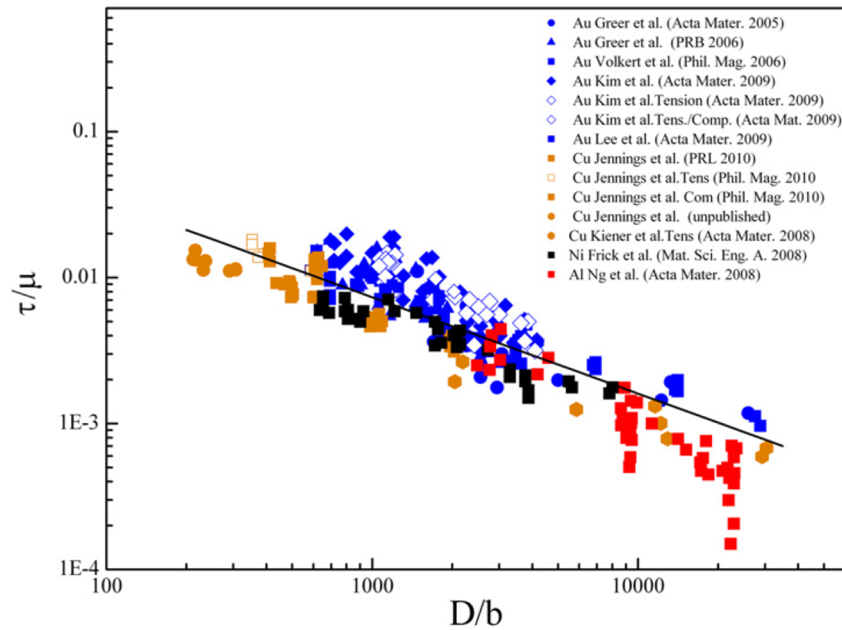

Figure 1.7: Shear flow stress normalized by shear modulus on appropriate slip system for metallic micro- and nano-pillars tested in compression and tension [33].

Up to now, in SMAs no solid correlation between pillar diameter and strength has been found [34]. Frick et al. reported the complete suppression of pseudo-elasticity in NiTi nanopillars with less than 200 nm diameter size, but in parallel the martensitic transformation prevailed [35–38]. In contrast, reversible stress-induced martensite transformation was observed in NiTi nanopillars with 200 nm pillar diameter [39] and in CuAlNi pillars [40, 41]. However, significant changes in the transformation temperatures were observed in CuAlNi wires [42] and in thin SMA films [43, 44].

Besides the macrostructure-based size effects just mentioned, also intrinsic, e.g. microstructure size effects may arise. The main issue concerning microstructural effects is the grain size of the material. Metals and alloys occur most commonly in their polycrystalline form. Therefore, the investigation of the influence of grain size, grain boundary spacing, dislocation density and their interaction is of particular interest. In bulk materials, grain boundaries limit the dislocation motion, leading to an increase in yield strength with decreasing grain size, described by the Hall-Petch relation. For grains below 40 nm grain size, a significant change in the grain size dependence is observed. The dependency obtained in bulk materials is reversed, which is denoted as the inverse Hall-Petch relation. The yield strength as a function of grain size is shown in Figure 1.8.

In this work, the focus is on polycrystalline NiMnGa as the active material, thus consisting of austenitic grains above the austenite finish temperature. As the material is cooled below martensite finish temperature, it consists of martensitic grains, with each grain having a self-accommodated, e.g. twinned structure of variants. By applying a mechanical load or a magnetic field, the reorientation of martensitic variants in each grain starts. Since the grain boundaries constrain the twin boundary motion, the induced strain is significantly reduced [45]. By shrinking the characteristic length scale of the specimen down to the

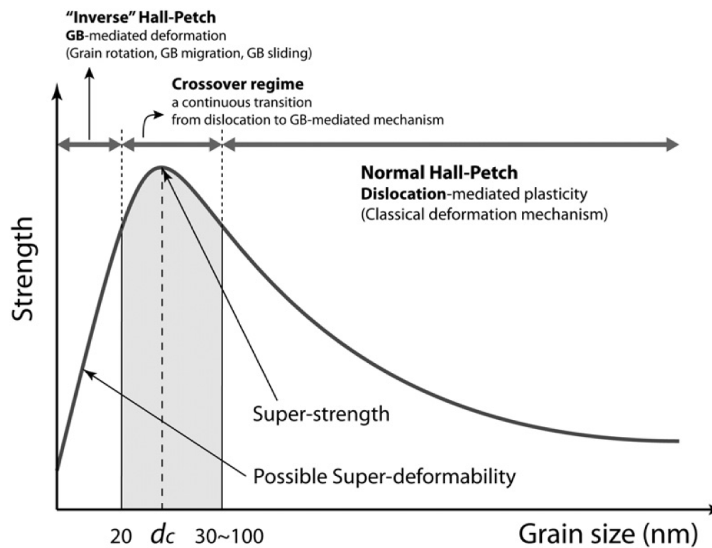

Figure 1.8: Strength of polycrystalline materials as a function of grain size: Hall-Petch relation and transition to inverse Hall-Petch [33].

grain size, these constrains may be reduced. In previous studies on NiMnGa foams, the MFIS strain could be increased from almost zero to up to 8.7 % [46]. Moreover, a change in critical de-twinning stress compared to bulk has been observed in micrometer-sized NiMnGa structures [47, 48].

A major challenge concerning both, intrinsic and extrinsic size effects, is the clear correlation between cause and effect. To extract bulk material properties such as plasticity from nanoindentation data, for instance, suffers from the so-called indentation size-effect [49]. Experimental artifacts influencing the measurement of size effects can be caused by fabrication issues, which gain importance at small scales or be related to the detection method. For instance, if the fabrication includes the combination of two thick material layers, each individual material behavior is defined by its individual material properties. If though one layer is very thin, the interface region starts to play an important role on the overall behavior of the specimen. Due to lattice mismatch between the crystal structure of the base material compared to the crystal structure of the top layer material, the crystal structure is modified in the interface region. Also, the interface between top layer and air can cause a size effect, since the amount of surface oxide compared to bulk material increases. Moreover, the material deposition technique can produce another chemical composition of the film in the first seconds of the process, before the system is in its equilibrium state. Also, the post-processing, like the focused ion beam milling of nanopillars can cause a damaged layer at the surface. In all these cases, the thin film or pillar properties would show significant difference in performance compared to bulk material behavior, but not caused by an extrinsic size effect but just because the bulk material and the nanostructured material are not exactly the same material anymore. Concerning experimental artifacts, also the measurement method plays an important role, for instance, the difficulties in comparing nanoindentation experiments with standard tensile tests. Even by performing a tensile test on a bulk specimen, the type of loading, load or displacement control, is known to play an important role. Distinguishing these

types in nanointentation experiments demands for additional simulation, since the arising load states are more complex. The necessity of simulation to extrapolate the results is associated with uncertainties up to systematic errors.

The investigation of nanoactuators is therefore always associated with the investigation of size effects. This includes not only the measurement of deviations in the performance compared to bulk material behavior, but also a deeper understanding what could be the reason for it - whether it is an artifact or new physics.

1.1.4 Constitutive Models

Since the discovery of the SME, models have been developed to understand the underlying physical mechanism of the martensitic transformation and to use this knowledge to facilitate the design and fabrication of SMA devices. This section gives an overview of the basic approaches and provides the framework of the SMA model developed during this work. The state of a SMA depends on the temperature, the stress but also the history of the state. Various constitutive models for SMA behavior have been established, which can be divided into three classes: Microscopic , micro-macro and the macro models [50, 51]. The first intend to describe the microstructural features of SMA behavior, like phase nucleation, interface motion and martensite twin growth. Using molecular dynamics or the Ginzburg-Landau theory, these models are capable of describing the SMA behavior on a lattice or grain-crystal level. This fundamental approach is very powerful, but results in large complex models and impractical computational costs when applied to model the behavior of macroscopic specimens. The category of micro-macro models relies on micromechanics, but average the behavior of single grains using scale transition techniques to describe macroscopic effects. For example, Patoor et al. developed a constitutive law based on energetic considerations for all 24 martensitic variants and the austenitic phase. With this model it is possible to predict SMA behavior on a single crystal level [52] and by assemblage of grains it is even possible to describe polycrystalline behavior [53]. The class of macro models involves phenomenological considerations, simplified thermodynamic descriptions and direct experimental data fitting. Tanaka and Nagaki were the first who coupled a phenomenological equation for the martensitic phase fraction with a kinetic law, which describes the evolution of the phase fraction as a function of stress and temperature [54, 55]. The corresponding parameters are engineering based and can be determined by relatively simple experiments. This presents the main advantage of phenomenological models - the simulation input parameters can be experimentally verified and are not abstract theoretical values. However, the association of physical causes is impeded, since the constitutive law is not based on a fundamental microscopic model. Belonging to the category of marco models, thermodynamically based models are often denoted as mesoscale models to highlight their theoretical origin compared to phenomenological approaches. For example, the Seelecke-Müller-Achenbach model uses statistical and thermodynamic concepts to describe the evolution of martensitic factions based on thermally activated processes [56].

In Figure 1.9, the Gibbs free energy is shown for a SMA with three possible states: austenite and two martensite variants. The austenitic cubic unit cell is illustrated in 2d

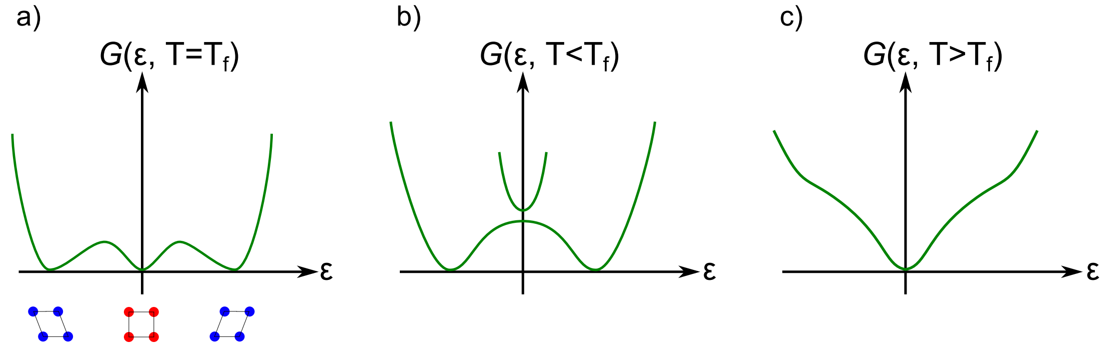

Figure 1.9: Schematic illustration of the Gibbs free energy G depending on temperature T and strain ε. a) Temperature equal to the transformation temperature T_f, b) $T < T_f$ and c) $T > T_f$.

by a unit square. According to the unit square representing the cubic structure of the austenite lattice, the less symmetric martensite lattice is illustrated by a parallelogram with different orientations indicating the two variants. For simplification, only a single transition temperature, denoted as T_f, is considered, instead of start and finish temperatures. By further assuming zero hysteresis and zero external load, the Gibbs free energy exhibits three local minima at $T = T_f$, illustrated in 1.9 a). For decreasing temperature, the martensitic minima decrease below the austenitic minimum, until the latter disappears, as shown in 1.9 b). Accordingly, the austenite phase becomes the energetically favorable state for temperatures well above the transformation temperature, illustrated in 1.9 c).

In general, the transition rate from phase α, e.g. austenite, to phase β, e.g. martensite variant 1, depends on the barrier between the different phases $\Delta G_{\alpha\beta}$ and strain. In a thermally activated process this transition rate is proportional to the Boltzmann factor given by $e^{-\frac{\Delta G_{\alpha\beta}}{k_B T}}$ with the Boltzmann constant k_B and temperature T.

1.2 Nanoactuator for Optical Switching

In fact, the suitability of an actuation principle for nanoscale applications highly depends on the individual objective. A look on energy density and the scaling laws helps to identify more or less promising principles, but the exact requirements can tip the balance. Does the application demand for static or dynamic operation? Is the frequency crucial or should the single travel range be as large as possible? Is intrinsic sensing required? Does the environment limit the range of actuation principles? For example, piezo-resistive actuators work near the breakdown voltage, therefore should not be applied if easy-inflammable gases or high humidity are present. Also, the available input power or input voltage might be limited. Besides this, the long term stability could be an issue or energy consumption might be of prior interest. Moreover, the compatibility of the fabrication technology can be the decisive factor, since the major part of nanoactuator applications include the combination with other nano devices.

In this work, a SMA-based NEMS (Nano-Electro-Mechnical System) actuator is developed for optical switching applications. Due to the known limits of metal interconnections, optical interconnections gain more and more interest. Therefore, the demand for optical data transfer solutions ranging from rack-to-rack over chip-to-chip and finally to inter-chip connections increases steadily. Associated with the development of optical interconnections is the integration of optical devices with state-of-the art silicon microelectronics. This includes not only low-loss waveguides, light emitters, amplifiers and lasers but also modulators - or more simplified - optical switches [57].

Optical switches or, in general, the modification of the propagation of light in a waveguide structure can be enabled by changing the crucial transmission parameters. For common planar waveguides the guiding is dependent on geometry and material parameters. Figure 1.10 shows a schematic of the characteristic geometry parameters for planar waveguide structures, consisting of a rectangular waveguide structure of material with a higher refractive index than the surrounding top and base material. In the most simplest description, the light travels along the waveguide due to total internal reflection at the interfaces. Constructive interference is observed, if the phase shift due to the reflection at the interfaces $\varphi_{base-wg}$ and φ_{air-wg} and the propagation across the waveguide Φ are a multiple of 2π. This condition limits the propagation inside the waveguide to a certain set of angles, resulting in propagation modes with a specific light velocity and electromagnetic field distribution. In rectangular waveguides, it is common to distinguish between transverse electric (TE) modes, with the electrical field parallel to the substrate, and transverse magnetic (TM) modes with the magnetic field parallel to the substrate. For very small waveguide geometries only one mode exists, called single mode waveguides, which are of high interest especially for sensing or telecommunication applications.

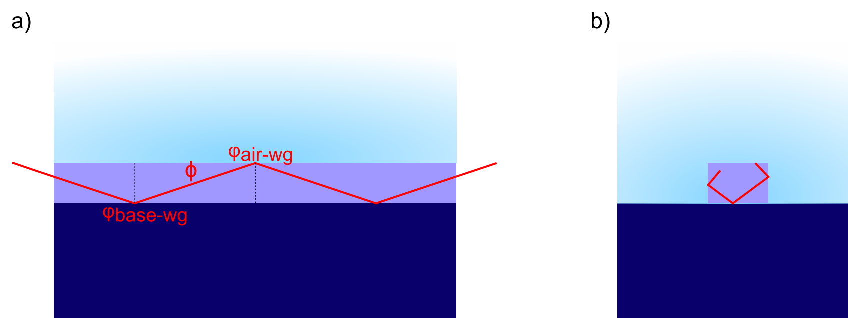

Figure 1.10: Propagation by total internal reflection in a planar rectangular waveguide. a) side-view and b) cross section.

The characteristic parameters for planar waveguide structures are the refractive index of the participating materials and the dimensions, like waveguide width and thickness. These are at the same time the parameters to adapt to modify the transmission. The intuitive way would be to choose electro-optic, acousto-optic, thermo-optic effects or even inject free carriers in a semi-conductor waveguide to tune the refractive index of the waveguide material. The most widely investigated concept is electro-optical modulation, however, in silicon the modulation efficiency is very weak due to its poor electro-optical properties [58]. As a result common structures have to be quite big to achieve a sufficient change of refractive index [59–61]. One way to fabricate compact electro-optical modulators is to

confine both the optical field and electrical carriers. Recent results have shown electro-optical modulators with very high quality factors using ring resonators with diameters in the order of 10 µm. As a consequence of the high confinement, the optical bandwidth of the modulator is narrowed significantly. Cascaded ring resonators can be used to overcome this limitation, although the individual resonance is critically affected by fabrications and temperature variations [62, 63]. Summing up, the use of material effects as electro-optical properties provides the great advantage of very fast switching operation, but suffers from their small effect size. Hence, either a not-compact modulator design or more complex structures are required, which most likely limit other modulator properties, like the optical bandwidth [64].

An alternative to the just mentioned concepts is the modification by mechanical actuation, which has been developed since the early 1990s in the micrometer regime [65–67]. The various concepts can be divided into three groups: First changing the coupling between one waveguide to another, 1.11 a). Second change the direction of the waveguide itself, 1.11 b), and last changing the propagation inside the waveguide. Concerning the latter the propagation can either be modified by inducing longitudinal strain to the waveguide and in this way change its refractive index, 1.11 c), or by modifying the effective mode index by interaction with the evanescent field, 1.11 d).

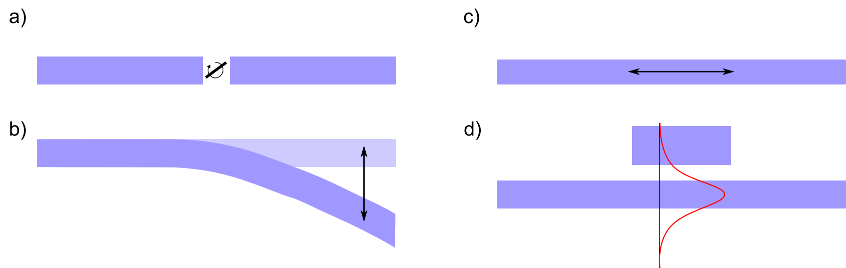

Figure 1.11: Mechanical actuation principles to modify the propagation though a common planar waveguide. a) coupling, b) redirection c) strain and d) evanescent field

In micrometer dimensions, the coupling between two fixed waveguides is, e.g., enabled by positioning a rotating mirror between them [65]. This concept is as simple as efficient but is not easy to miniaturize. The concept of redirecting the waveguide as well as the strain induction imply partly freestanding waveguides, which lead to an increase in fabrication complexity and also induce propagation losses. The positioning of a slab of material in close vicinity to a waveguide results in the conversion from its initial refractive index to an effective index of refraction, dependent on the material of the slab and its distance to the waveguide. Choosing equal materials for slab and waveguide makes these structures easy to fabricate. Moving the slab enables not only switching but also tuning the propagation though the waveguide. In micrometer dimensions this concept has already been successfully used as a light modulator within a microphone [68].

The key advantage of using a SMA as the active material for the development of a NEMS actuator for optical switching is the large travel range. Moreover, the actuator position, e.g., the resulting propagation mode, can be held without energy consumption. Although most SMAs show a hysteresis, their suitability as sensor elements has been shown

[69], eliminating the need for alignment of separate actuation and sensing components. Summing up, the integration of SMA-based NEMS actuators in photonics most likely enables significant performance improvements and new functionality.

1.3 Fabrication

When talking about nanoactuators one has to distinguish between devices, which are capable to generate displacements with nanometer accuracy and being itself two or more dimensions bigger, like x-y stages or nanoindenter-tools, and on the other hand devices, which are itself nanometer sized. The latter is the point of interest in the presented work here, e.g., nanoactuators, which generate nanometer displacements, while being itself in the nanometer regime. To realize nanoactuators according to the just mentioned definition, the development of appropriate fabrication techniques is a key element. Up front, there are two concepts, when trying to approach nanometer sized functional elements: On the one hand, the so called *bottom-up* approach and on the other hand, the so called *top-down* method. The first refers to the guided growth of nanostructures using self-assembly algorithms manipulated by adapted boundary conditions. Following the top-down approach, in contrast, the initial specimen is a bulk shape sample, which is partly removed by high-resolution structuring processes. The elaborate method in microelectronics and microsystems technology is the top-down approach, more precisely, the most common concept is photolithographic structuring. In the following, an overview of the used processes is given, the details and concrete examples of the adaptions and further developments acquired during this work are presented in Chapter 2.

1.3.1 Patterning Techniques

In general, photolithography uses monochromatic light, which exposes a photo sensitive polymer. Between the photo sensitive polymer, called the resist, and the light source a mask is positioned. The mask shields certain parts against the illumination producing locally exposed resist structures next to unexposed ones. By the exposure of the photo sensitive polymer, the polymer chains either get destroyed, which is called a positive (tone) resist, or the monomer chains get crosslinked, which is called a negative (tone) resist. In the first case, the exposed resist is more easy to resolve than the not exposed parts, compared to the second case, where it is the other way around. Using this technique, it is possible to produce several structures in parallel by one exposure step making the fabrication fast and economic. The maximum recoverable resolution of the manufacturable structures is theoretically limited by Abbé's law [70, 71]:

$$d = \frac{\lambda}{2 \cdot n \cdot sin(\alpha)} = \frac{\lambda}{2 \cdot NA} \tag{1.1}$$

with the wavelength λ the angle of incidence α and the refractive index n; the latter can be summarized by the term numerical aperture $NA = n \cdot sin(\alpha)$. With the use of an ArF-excimer-laser as a light source ($\lambda = 193\,\mathrm{nm}$) and the use of special immersion lenses

($NA = 1.35$), the resolution has been reduced to 22 nm [72, 73]. Recent developments focusing on extreme ultra violet (EUV) light sources, like plasma gas discharge, have the potential to further push the limit [74]. However, the presented photo lithography manufacturing always requires a mask, which has to be produced by a direct lithography method - typically Electron Beam Lithography (EBL).

Moreover, if one is not interested in a pure resist structure, the primary resist structure has to be transferred into the target material. This is done by an etching process using the pre-structured resist as a protective mask. In the following, the different patterning techniques used in this project are introduced. First, the exposure technique, EBL, is described, followed by the explanation of the two main etching procedures and the associated drying techniques.

1.3.1.1 E-beam Lithography

Electron Beam Lithography (EBL) is a mask-less lithography method. As all other direct-write methods it suffers under low throughput due to the serial exposure but provides maximum flexibility. Since no fixed masks are used, the layout can be changed after every single exposure step. Especially in research, the possibility to take the experiences of former actuator generations into account is highly wanted. Additionally, the layout can contain nanometer sized features as well as macrostructures in the range of millimeters, for instance needed as contact pads for electrical connections. Besides the mask-less procedure, the main difference between common photo lithography and EBL is the used light source. While photo lithography uses photons, EBL uses accelerated electrons. Although, the maximum recoverable resolution is still given by equation 1.1, this practically does not limit the resolution. The wavelength of an accelerated electron λ_e is defined by its velocity v_e and its mass m_e via the de-Broglie relation:

$$\lambda_e = h/m_e \cdot v_e \tag{1.2}$$

For common acceleration voltages of 100 kV the approximated velocity of an electron is of the order of 0.55 times the velocity of light leading to a wavelength of approximately $0.037 \cdot 10^{-10}$ m. Thus, the theoretical limit given by the Abbé's law plays only a minor role. In practice, the maximum accessible resolution is limited by the electron distribution, the quality of the electron source, the beam guidance, the focussing optics and the resist and developer properties.

The structures presented in this work are made with the *Vistec VB6-UHR-EWF* in the 100 kV operation mode. The main elements of a standard electron beam writer are shown in Figure 1.12.

As an electron source the VB6 electron beam uses a Schottky-field-emission cathode built out of a sharp single crystalline tungsten tip coated with zirconium oxide. The zirconium oxide and the tungsten form a inter-metallic phase at elaborate temperatures, so that the high tungsten work function is decreased and the electron emission is significantly increased. The working temperature of the cathode in the used electron beam systems is about 1800 K, which provides sufficient emission and relative long lifetime [75, 76].

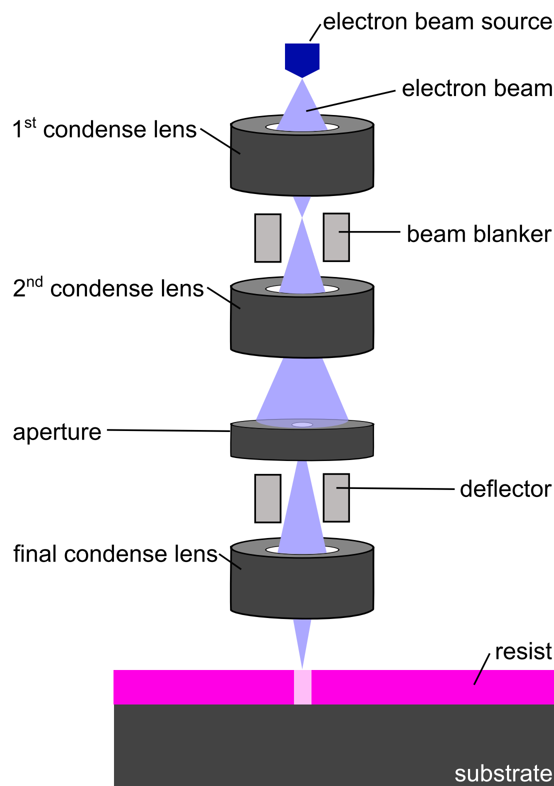

Figure 1.12: Schematic illustration of electron beam lithography.

Afterwards, the free electrons are accelerated by a voltage of a few kV at the extractor, whereas the negative suppressor is suppressing electrons with low velocity in the beam direction. The final velocity is given by the voltage at the anode - here 100 kV. To focus the beam as well as to adjust the beam current, the first capacitive lens is used. Next, the beam passes a beam blanker, which corrects the angle and the offset. Afterwards, a second magnetic lens focusses the beam onto the aperture. Hereby, it is possible to block the beam, for example between moving the beam from one writing field to another [77]. For the VB6 apertures with a diameter of 40, 70 and 100 µm are available. The smaller the aperture is chosen the better the depth of focus and the higher the resolution, but the beam current is reduced in parallel. Therefore, one has to balance these tradeoffs against writing time and resist thickness. After the aperture a system of deflector coils is used for a coarse alignment in a range of 100 to 1310 µm. Additionally, the substrate is mounted on a movable platform. Finally, a second magnetic lens focuses the beam onto the substrate. The fine alignment of the focus and the correction of astigmatism are enabled in real time dependent on the position relative to the surface of the resist by smaller coils inside the lens. The exact position of the resist surface is determined by the reflection signal of a laser beam on a CCD camera.

Besides the precision of the focusing and guidance optics also the interaction of the electron beam with the resist and the substrate material are limiting factors concerning the smallest feature size. The interaction can be divided into three effects: The forward scattering, the backward scattering and the generation of secondary electrons. All three effects lead to the deposition of energy in areas of the resits, which are not in the target area of the

electron beam exposure. Hence, the produced resist structures can be enlarged, rounded, being wrong positioned or small gaps could even vanish completely. In EBL, the forward and backward scattering, which are commonly summarized by the term *proximity effect*, matter the most. For deflection angels of 90 ° and more, inelastic collisions of the beam electrons with the electrons of the resist atoms lead to a backward scattering of the electrons. As a consequence, the electron beam diameter increases with increasing resist thickness. The relation between the resulting electron beam diameter d_f and the resist thickness has been observed empirically and can be described with the following equation:

$$d_f = 0.9 \cdot \left(\frac{t_R}{E}\right)^{1.5} \tag{1.3}$$

with the resist thickness t_R in nm, the acceleration energy E in keV [78].

At a deflection angle of less than 90 °, the electrons are forward scattered due to elastic collisions of electrons with resist atoms and substrate atoms. Besides, secondary electrons can be generated by inelastic collisions. Due to the layer thickness ratio and the lower atomic number of common resist materials compared to substrate materials, both effects are dominant in electron - substrate interaction [79]. The energy input decreases with the acceleration voltage, but in parallel the affected area increases, up to several microns round the focus [78].

To prevent or minimize this effect, the scatting behavior of the electrons dependent on the layer sequence can be simulated. The result is taken to adapt either the dose, summarized with the term dose correction [80], or the borders between exposed and not-exposed areas are adapted, called shape correction [81]. Another technique is the equalization of the backscattered dose received by all layout points. This is accomplished by exposing the reverse tone of the required pattern in a second exposure step [82]. The last is a quite simple procedure but accompanied with the double writing time and a lower contrast of the resist material.

1.3.1.2 Etching techniques

After the successful EBL step, the resist is developed and the exposed or unexposed parts are resolved, depending on whether the tone of the resist is negative or positive. The result is a resist pattern on a substrate material. To transfer the resist pattern into the substrate material, an etching procedure is used. The resulting width of the target substrate structures depends on the resist layer thickness, which is again limited by the selectivity of the used etching process. The selectivity is defined as the etch rate of the resist material versus the etch rate of the substrate material. If the ratio is too low, it might be necessary to first transfer the resist pattern into a mask material. The mask material is often a noble metal, which is chemically inactive and its hardness has a positive influence on the selectivity. In general, two main etching techniques are distinguished. On the one hand, wet-etching techniques and on the other hand, dry etching procedures. Regarding wet-etching procedures, the ions do react in an aqueous solution with the material to be etched and the product is dissolved in the liquid afterwards. The advantages of this technique is its simplicity regarding the required equipment and the rich experience

concerning materials and their etchants. Besides, the etch rate can be well controlled by adapting the buffer, the thinner and the temperature. The main disadvantage is the limitation to isotropic etch profiles due to the small mean free path of the active etch particles. However, some etching solutions form an exception, since their etch rate highly depends on the crystal plane of the material, allowing for anisotropic etch profiles in the case of single crystalline materials. An additional disadvantage are capillary forces, which might become an issue during the drying after the wet-etching of very small structures. In contrast to the wet-etching techniques, dry-etching procedures demand for a closed system where the etching products and the gaseous etchants are pumped out. This increase in technical expenditure allows to etch materials for whom no liquid etchant exists. Moreover, anisotropic etch profiles become possible for polycrystalline materials. In the following, common dry etching techniques are introduced at first. Afterwards a standard wet-etching procedure is presented including the corresponding drying options.

Dry Etching

Dry-etching techniques can itself be divided into two groups - first, processes based on pure physical processes and second a combination of physical and chemical processes. In this work, the *Plasmalab 100 ICP 380 (Oxford Instruments)* is used, which provides one process chamber for **Ion Beam Etching (IBE)** and one process chamber for **Reactive Ion Etching (RIE)** connected by a Load-Lock transfer chamber. The details of both process chambers are explained in the following.

The IBE system itself consists of two chambers, one chamber for the plasma generation and one chamber, where the actual etching takes place. This prevents the contamination of the plasma chamber with ablated particles of the sample, which could cause changes in the plasma ignition conditions. Besides, different samples do not get contaminated with materials from previous processes. By the use of a chemically inert gas, e.g. Argon, a pure physical process is established, which enables to structure every material, independent on its atomic composition. Figure 1.13 shows a schematic setup of an ion beam etching system.

In the plasma chamber, Argon gas is excited by a capacitive coupled plasma (CCP). This method implies the use of two condenser plates driven by a single radio-frequency (RF) power supply, at 13.56 MHz. The resulting alternating electric field cause the gas molecules or atoms to dissociate, creating ions and free electrons. The free electrons can in turn ionize remaining neutral gas atoms directly or indirectly by collisions. By adjusting the electric power of the alternating field, the kinetic energy of the ions and therefore the strength of the etch attack can be adjusted. Often, the generation of a CCP is accompanied by the emission of light from excited atoms or molecules in the gas. The resulting positive Argon ions are accelerated by a negatively charged extraction grid on the sample holder. At a second cathode, the neutralizer, electrons are evaporated to recombine with the Argon ions. Thereby, the Argon beam impinging on the substrate mainly consists of neutral Argon atoms, limiting the charging of the sample and enabling a continuous material removal. Since the sample holder does not work as the electrode in this setup, it can be tilted relative to the argon beam. This allows for different etching

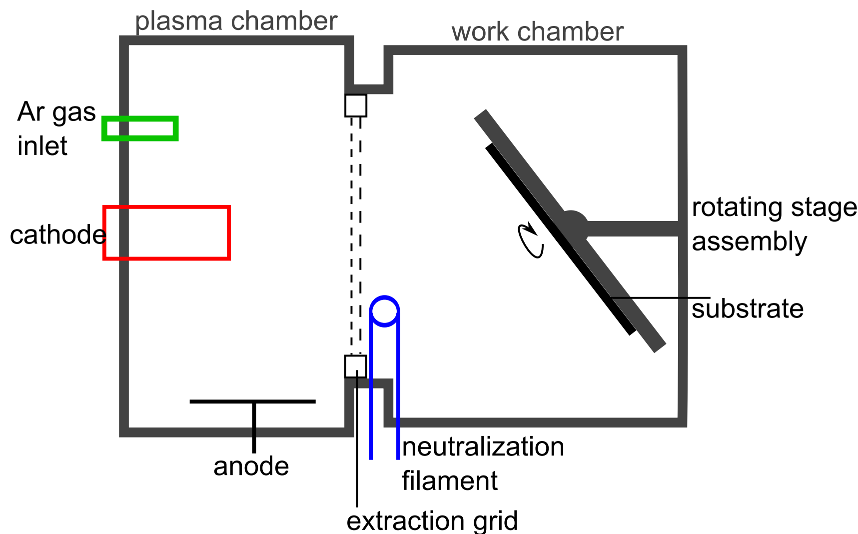

Figure 1.13: Schematic of an ion beam etching system.

angles also influencing the overall etch rate [83]. Since the material removal is solely due to the physical process of atom bombardment, the selectivity is quite low - depending primarily on the hardness of the material and not its chemical composition [84]. Besides, the aspect ratio is limited to values of the order of 1:1 due to redeposition. For increasing aspect ratio, the probability increases, that ablated substrate particles re-deposit on the sidewalls. This can be reduced by tilting and rotating the substrate holder during the etching process, but not be inhibited totally.

To allow for high selectivities and high aspect ratios **Reactive Ion Etching (RIE)** is used. The chemical ablation of ions is highly material specific and can be associated with very high etch rates. Compared to wet-etching techniques, the combination of chemical ablation and acceleration of ions towards the substrate allows for a high directionality. This enables anisotropic etch profiles of polycrystalline materials. Although, isotropic etch profiles can be generated using a certain set of parameters. Key setting parameters are the chamber pressure, determining the mean free path of the ions, the plasma power, determining the kinetic energy of the ions and the total amount of ions and substrate temperature, determining the etch rate. The latter can be adjusted between approx. 150 °C and approx. 300 °C by an integrated hot plate and the possibility to cool the substrate holder with liquid nitrogen. Therefore, the liquid nitrogen cools down gaseous helium which washes around the backside of the substrate. The position of the substrate itself is fixed by a glass holder. A schematic of the RIE chamber used throughout this work is shown in Figure 1.13.

The plasma generation for RIE is enabled in a similar manner than for IBE, by a CCP, but with some distinct different features. First, the plasma generation and the etching take place in the same chamber, with the second electrode being the sample holder. Thus, a rotation or tilting of the sample holder is not possible. To maximize the rate of chemical etching over the physical portion, the temperature of the CCP is normally chosen quite low. To generate a sufficient amount of ions and accelerate

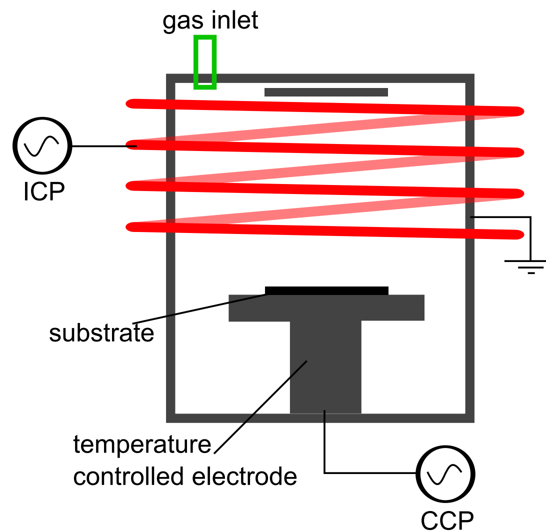

Figure 1.14: Schematic of a reactive ion etching system.

them towards the electrode, a shielding effect is tapped. Due to the higher amplitude of the electrons in the plasma, which is due to the lower mass compared to the ions, they impinge on the chamber side walls and any other interface and get absorbed. This leads to a leakage of electrons, thus a positively charged depletion region, which is called *sheath*. The resulting potential difference, between plasma and chamber side walls and also plasma and electrode, accelerates electrons more inside the plasma than to the electrode, which stabilizes the sheath and leads to further ionization. In parallel, the positively charged ions are accelerated to the interface region, e.g. the electrode. Nevertheless, the pure usage of CCP does not allow to isolate the generation of ions from their acceleration towards the substrate. Therefore, an increase in the chemical etch rate is always associated with an increase in the physical etch rate, e.g. mechanical removal of material. For most RIE applications, the portion of physical etching is an unwanted side effect, since the ablation rate is widely independent on the material, thus reducing the selectivity of the etch process. The overcome this correlation, the RIE chamber is equipped with the option to use an inductive coupled plasma (ICP). As illustrated in Figure 1.14, the inductive coupling is enabled by a coil driven by an RF power supply. The reversion of magnetic field generates electric fields, which in turn lead to dissociation and ionization. The potential difference of the accompanied sheath region is much smaller compared to a standard CCP one. The final acceleration of the gas ions towards the sample is enabled by a CCP-like setup. The combination allows to control the chemical and physical proportion of the etch rate widely independent. To monitor the stability of the plasma and get some indication about the physical proportion, only the counter electrode is grounded. During the etching process, a direct current voltage will arise between substrate holder and counter electrode, due to the acceleration voltage of the positive ions. This so called *Bias-Voltage* is easy to measure and simplifies the detection of variations during processing.

To establish SMA devices in standard microtechnology systems, the use of silicon as a substrate material is mandatory. Therefore, one of the main etching processes used during this work is the anisotropic etch of silicon with the main aims of high aspect ratios and rectangular, smooth sidewalls. Whereas the Bosch process [85] is well known to generate high aspect ratios with alternating etching and passivation steps, the resulting wave profile is way to rough, for example, for optical applications. A smoother profile is enabled by controlled and continuous construction and destruction of a passivation layer. One way to realize this, is to use a mixture of sulfur hexafluorid (SF6) and oxygen and a substrate temperature of below - 80 °C. At these temperatures the generated SiOxFy compounds are not able to become gaseous anymore and stay at the Silicon surface, passivating it against further etching [86]. By a relatively low chamber pressure, the mean free path is increased, leading to an anisotropic physical etch rate, which removes the passivation layer normal to the accelerated ions and keeping it at surfaces lateral to it [87]. This so-called cryo-process is an established anisotropic silicon etch process with very low surface roughness, but it has one major disadvantage regarding nanotechnology. The etch rate is of the order of less than 100 nm per 10 s and therefore below one minute for structures of a few hundreds of nanometers. With a suitable stopping layer and an appropriate process monitoring, the cryo-process can be used also for thin silicon layers, but the short process time has negative influence on the reproducibility; particularly if considering a high power striking step to start the ignition of the plasma and therefore the etching process. A good alternative is the use of hydrogen bromide (HBr) as the process gas. At temperatures of approx. 60 °C, the etch rate is approx. 100 nm per minute and therefore well controllable. Additionally, the selectivity is very high, e.g. 50:1 for the negative tone resist HSQ, and the resulting side-walls are as smooth as with the cryo-process [88]. The main disadvantage is, that HBr can corrode even stainless steel in the presence of moisture. The corrosion can lead to particle contamination on the wafer, mechanical failure and, in worst case, leaks in the gas delivery system. Since HBr is highly toxic, this demands for specific safety treatment, like heating the gas pipeline.

Wet Etching

Wet-etching procedures sum up processes, where the ions of a liquid etchant react in an aqueous solution with the to-be-etched material and the reaction product is dissolved in liquid afterwards and removed from the substrate by diffusion or convection. In general the etch rate depends on the material, the corresponding etchant and its concentration, the temperature, partly on the convection conditions and in some cases the crystal orientation of the substrate. For example, by the use of an potassium hydroxide solution it is possible to produce anisotropic etch profiles in (110)-Si-wafers [89]. Here, the only used wet-etch process is the isotropic structuring of SiO_2 with the use of diluted hydrofluoric (HF) acid. That is why, the key elements for the wet-etching procedure are explained on the basis of this process. The chemical reaction is described by the following reaction equation:

$$SiO_2^{\text{solid}} + 6HF^{\text{liquid}} \longrightarrow H_2SiF_6^{\text{dissolved}} + 2H_2O^{\text{liquid}} \tag{1.4}$$

Hence, every single Si atom is bound to six fluorine ions, leading to a decrease in fluorine ion concentration over process time and therefore a decrease in etch rate. To generate

a homogeneous etch rate, a buffered HF solution is used. The buffer, here ammonium fluoride (NH_4F) which dissociates to NH_4^+ and F^- ions in water, compensates the fluorine loss caused by the primary reaction.

$$NH_4F \rightleftharpoons NH_4^+ + F^- \tag{1.5}$$

1.3.1.3 Drying

Besides the before mentioned disadvantage of wet etching compared to dry etching, the drying of the generated structures becomes an issue especially at the nanoscale. Due to the inhomogeneous evaporation of the aqueous solution, capillary forces between serried features emerge, which can cause them to collapse. To overcome this problem, the

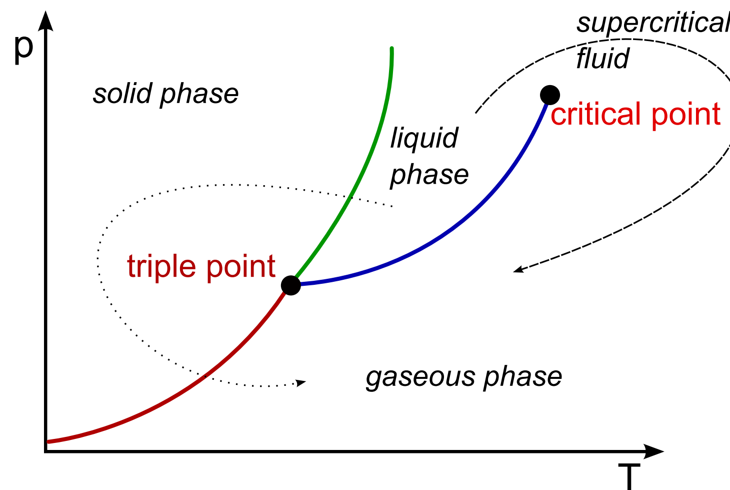

Figure 1.15: Schematic phase diagram illustrating the aggregate state of a material depending on pressure p and temperature T.

transition from the liquid to the gaseous phase has to be controlled. The normal drying procedure is governed by the material's boiling temperature, which itself depends on the pressure. The schematic phase diagram shown in Figure 1.15 exhibits not only a dependency for the boiling point (blue line), but also for the freezing point (green line) and for the sublimation point (red line). By common drying the blue line is crossed, thus the material directly transforms from its liquid phase to its gaseous phase. Alternatively it is possible to go from the liquid phase to the solid phase and by sublimation to the gaseous phase, e.g. round the triple point, which is named *freeze drying*, or to pass the critical point, which is called *critical point drying*. Both techniques are utilized during this work.

Freeze Drying

The freeze drying process can be divided into three main steps. The first step comprises the exchange of the aqueous solution, in which the sample is stored in. Suitable liquids should provide a low saturated vapor pressure. In the second step, the sample and the

medium is cooled below the freezing point, until the whole medium is in its solid phase. At this point, one can distinguish between two material groups: On the one hand, materials which expand while freezing and, on the other hand, materials which decrease in volume while freezing. An example for the first group is common water, which is used quite often even though it is more likely to destroy some features by the expansion. An example for the second group is cyclohexane, which has good freezing qualities but is highly toxic. After the material is in its solid phase, the pressure is decreased. In this last step, the material will start to sublimate continuously until the whole medium is gaseous.

Critical Point Drying

Also the critical point drying process can be divided in three main steps, with the first one being the exchange of the medium, for instance the acid used for the previous wet-etching process [90]. To limit the influences on the sample, the medium should offer a preferably low critical temperature and critical pressure. Most common is carbon dioxide (CO_2), which has a critical point at a temperature of 30 °C and pressure of $7 \cdot 10^6$ N/m^2. Because most wet-etching liquids are based on an aqueous solution and the solubility of water in liquid CO_2 is very low, the water is first exchanged with a solvent of less polarity, like isopropanol and afterwards the isopropanol is exchanged under increased pressure by liquid CO_2. In the second step, the temperature is increased at constant pressure higher than the critical pressure. After the medium passed the critical point and transformed from a liquid into a supercritical fluid, the pressure is decreased at constant temperature in the final process step. The medium will transfer from a supercritical fluid into a gas with no emerging capillary forces.

1.3.2 Thin Film Deposition

Besides the precise control of the lateral structure size by the introduced patterning techniques, the vertical dimension of the structures is of main interest in nanotechnology. Suitable deposition techniques for nanometer structures therefore have to provide high accuracy of the deposited layer thickness as well as accurate control of the purity of the generated thin film. The best procedures to deposit a thin film of material on a certain substrate cannot be chosen on general purpose and often is associated with making a compromise. An important benchmark parameter for the film quality is the homogeneity in terms of the crystal structure, e.g. grain size, the film thickness and the chemical composition in lateral direction and across the film thickness. Besides, the adhesion can play an important role and can restrict the choice of deposition techniques. In the following, the two thin film deposition techniques are introduced, which were used for the manufacturing of the developed actuators.

1.3.2.1 Electron Beam Physical Vapor Deposition

The most common technique, to deposit a thin film of pure metal, is to evaporate a certain amount and let it condensate on a substrate. In this manner, a high surface quality can be achieved. The temperature suitable to vaporize the metal can be enabled by various techniques, from simple Joule or inductive heating to laser induced or electron beam induced evaporation, which is commonly called **Electron Beam Physical Vapor Deposition (EBPVD)** [91]. With the use of a common tungsten filament, electrons are generated and accelerated by a high voltage cathode anode arrangement. Electrostatic and magnetic fields are used to steer the electron beam. Thereby, the beam is directed towards the ingot composed of the material to be deposited, which is placed in a positive charged tungsten crucible. By electron bombardment, the metal gets heated up, melts and finally, the atoms transform into the gaseous phase and coat every surface in the vacuum chamber including the sample. A schematic of the operation principle of an EBPVD system is shown in Figure 1.16.

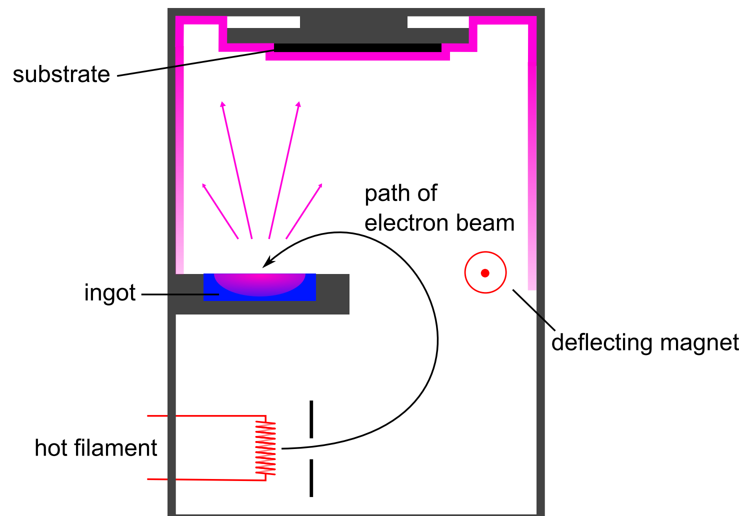

Figure 1.16: Schematic of the operation principle of an Electron Beam Physical Vapor Deposition system.

The thickness of the resulting thin film can be up to several micrometers and down to a single atomic layer grown epitaxially in Molecular Beam Epitaxy. For very thin films, the growth of the crystal layers is monitored by a method called reflection high energy electron diffraction (RHEED). For film thicknesses in the order of nanometers and more, the amount of deposited material is monitored indirectly with the use of a crystal oscillator. The resonance frequency of the crystal changes due to the increase in volume by the additional layer of deposited material. The frequency shift can be measured very accurately and therefore, the layer thickness can be determined with high precision. During this work, EBPVD was used for the fabrication of gold markers used for the alignment of different electron beam lithography layout layers. Besides the broad range of applications, EBPVD is not very common for the depositions of alloys. The different components of an alloy all have their individual vapor pressure, which makes the continuous vaporizing of a mixture of the constituents impossible. As an alternative,

one electron beam evaporator for each ingredient can be used. This technique is used for applications demanding for very thin layers of binary alloys but the range of parameters makes it quite hard to achieve a certain stoichiometric composition over thick films and for more than two constituents. Moreover the degradation of the filament, which influences the deposition rate, makes reproducible results for alloy thin films even more demanding.

1.3.2.2 Sputter Deposition

Instead of evaporating the material and letting it condensate on a substrate, it is also possible to *strike* atoms or clusters of atoms out of bulk material, called target, by inert gaseous ions. Procedures using this concept are summarized by the term **Sputter Deposition** and are up front independent on the individual vapor pressure of the material. Accordingly, the stoichiometric similarities of the thin film and the target are much higher compared to evaporating techniques. As already discussed in Section 1.1, the stoichiometric composition is of highest interest for the shape memory effect issues making sputter deposition a common technique for fabrication of SMA thin films. The fabrication of appropriate targets is commonly made by melting down the components of the alloy in a shielding gas atmosphere or under vacuum conditions. The gaseous ions are generated by a plasma, which is again generated by applying an AC current between two condenser plates, similar to the dry etching technique described in Section 1.3.1.2. For sputter deposition, one condenser plate is the substrate and the other one is the target, composed out of the desired thin film material. Normally an inert gas, like argon, is chosen, but also reactive gases like oxygen or nitrogen can be used to sputter compounds. The compound can be formed on the target, in-flight or right at the substrate surface, depending on the process parameters. The NiMnGa thin films investigated during this work are produced by a special modification of the just described process, denoted as **DC-magnetron sputtering**. The corresponding setup is shown schematically in Figure 1.17.

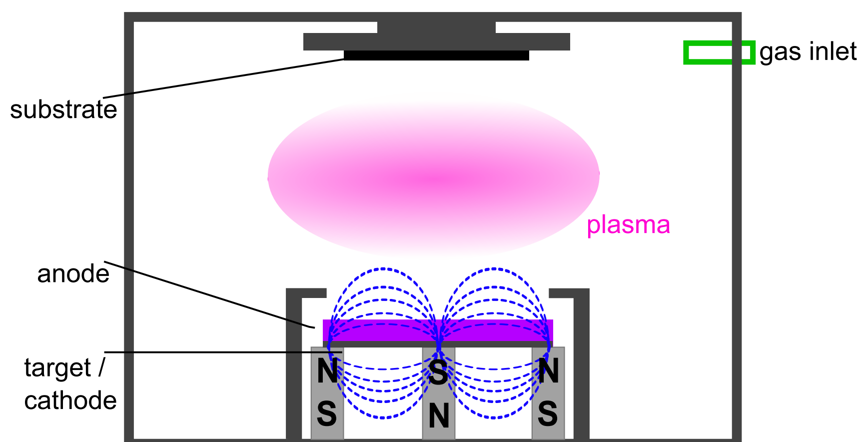

Figure 1.17: Schematic of the operation principle of a DC-magnetron sputtering system.

The condenser plates are supplied with DC voltage of several kV causing a destructive breakdown of the process gas. The generated ions not only strike the desired atoms

out of the target, but also adsorbed gases, secondary electrons and x-ray radiation does arise. Since the target material has to cross the plasma region and might tend to react with the present gasatoms, a low overall gas pressure is most suitable to achieve high quality thin films. By the positioning of several permanent magnets below the target, the adsorption of electrons at the chamber sidewalls is suppressed and additionally their collision probability with neutral gas atoms is increased. Both factors enlarge the plasma density, which allows for a lower overall chamber pressure, reducing the adsorption of gas atoms by target atoms [92].

The product of a standard sputter deposition at room temperature is an amorphous film. Nevertheless, for a functional SMA layer crystallinity is essential. To generate a crystalline film by sputter deposition, there are basically two options: First, the sputter deposition at elaborate temperatures or second a post annealing step. By a thermal treatment the amorphous film transforms into a polycrystalline film, but in parallel diffusion processes will speed up. The latter can cause an inhomogeneous distribution of the chemical composition across the film thickness, caused by the different atomic masses of the constituents, by oxidation at the surface or by compound formation at the thin film to substrate interface. The corresponding benchmark parameters, like stoichiometric homogeneity, grain size and its variation along the substrate can be influenced by adjusting the annealing temperature, time and the atmosphere. For the latter a high vacuum or a chemically inert gas is most sufficient to minimize oxidation. Besides the annealing time, the cooling procedure can influence the film properties, e.g. active cooling using liquid nitrogen suddenly stops any diffusion processes.

2 Nanofabrication

The aim on any kind of actuator design is to provide maximum deflection with the smallest possible frame size. Concerning the SMA-based actuator presented here, this task already brings up a key question: How is the actuator powered? Belonging to the thermal actuation principles, SMA-based actuators demand for some temperature variation. There are several options to provide a temperature change, which can be categorized into the group of direct and indirect principles. An example for an indirect principle is ambient heating, e.g. using a hot plate. Since the actuators presented here are nanometer sized, the deflection can only be monitored with a microscope capable of resolving nanometers. The integration of a hot plate inside e.g. a SEM may lead not only to an increase in temperature of the chip, but also of the SEM stage, which in turn can hamper its functionality. Moreover, the thermal drift of the sample exceeds the deflection of a single actuator by far and might even exceed the scan field of the SEM. An alternative indirect technique for tuning the temperature of the actuators is laser heating. However, the moving of the focus of the laser beam, or the positioning of an optical fiber above an actuator structure will increase the technical expenditure a lot. Independent on the technical challenges, the laser as well as the hot plate as an indirect principle limit the range of application of the actuator. Even though, the driving force parts are commonly not counted to the footprint of an actuator design, in a real on-chip application the overall size matters. One way to tackle this problem is to use a direct concept, e.g. heat the nanoactuator by Joule heating. As a consequence the layout is restricted to closed geometries like bridges and double beam structures. Since double beam geometries provide larger deflection values by the same driving force compared to bridges with the same dimension, a double beam geometry was chosen for all actuators investigated in this work. By choosing to heat the double beam with an electrical current, the monitoring of its electrical resistance is no extra effort. The recording of electrical resistance gives an insight to the course of the phase transformation and therefore allows to assign deflection states to the material phase. A chip layout is set-up that contains a series of actuator structures with different lengths and widths having different deflection characteristics in order to investigate size effects. Besides, several actuators with identical geometry are repeated on the chip to gain insight in the accuracy of the fabrication result. A schematic of the chosen double beam layout is shown in Figure 2.1. The geometry parameters are indicated.

As explained in detail in Chapter 1.1, the maximum recoverable shape change is only induced one time after deformation, if no training or no mechanical load is applied. To facilitate a cyclic operation, a resetting mechanism has to be implemented. One possibility is to add an additional layer, creating a bimorph structure. Since the integration of SMA-based actuators in the standard Si process technology will provide the widest range of

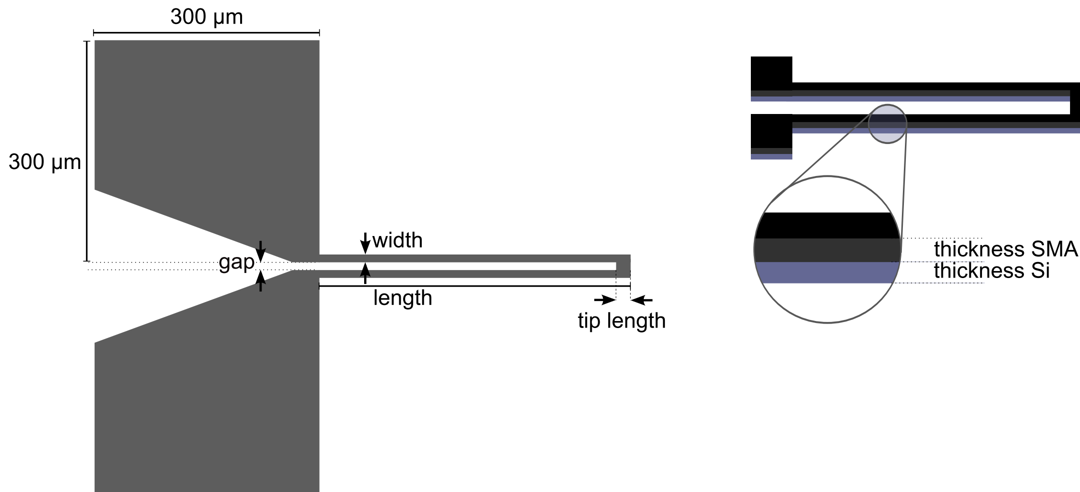

Figure 2.1: Bimorph double beam layout of the SMA-based actuators with the geometry parameters indicated.

applications, an appropriate choice for this base layer material is Si. At this point the nanoactuator layout is defined and with this the goal for the nanofabrication process.

Regarding the fabrication of SMA-based nanoactuators, a few challenges have to be coped with. The major part of SMAs, including NiMnGa, belong to the difficult-to-etch materials, which prohibits structuring of a NiMnGa film by RIE. Even the structuring of other materials on the same chip by RIE might cause some problems. If the atoms of the etch gas are very small, they might diffuse into the SMA thin film. Since SMAs are highly sensitive to their chemical composition, this will most likely reduce their shape recovery capability. Moreover, almost all SMAs demand for rather high deposition temperatures, e.g. 500 °C for NiMnGa, or an annealing step to obtain a functional crystalline material. This temperature treatment induces thermal stress, which increases the probability of delamination. Also interdiffusion between substrate and SMA material might occur. Besides, any resist will just combust during the annealing procedure, hence no pre-patterned resist mask can be used. Additionally, the unit cell mismatch between Si and NiMnGa is quite big, which further promotes delamination.

There are several options to face these challenges: First, the nanomachining of the SMA film after film deposition and second, the nanomachining of Si prior to deposition of SMA. In both cases, a Silicon-on-Insulator (SOI) wafer is used as a substrate. SOI wafers consist of a Si handle wafer, with a layer thickness in the order of several hundreds of micrometers, which is needed to enable safe handling. The first functional layer is composed of SiO_2, which is commonly called buried oxide or just box. The box provides a nonconducting layer, which is of high interest in silicon electronics or the base material with lower refractive index than Si, for Si photonic applications. The top layer of an SOI wafer is the so-called device layer, composed of Si. In our case, the Si is rarely p-doped with a nominal resistance of 1-50 Ωcm. The device layer thickness of the used SOI wafer is 340 nm Si which is combined with a 2 µm thick SiO_2 layer. This layer thickness ratio has already been successfully used for Si waveguide applications and the 2 µm thick box provides a sufficient deflection range in vertical direction.

First, the details of the structuring of the NiMnGa film prior to the Si is discussed, followed by the description of the reversed approach.

2.1 Nanomachining after film deposition

The structuring of a substrate already coated with all layers of the final layout provides several advantages. First of all, it allows for maximum geometry and design flexibility. Additionally, the layer sequence can be measured accurately on a thin film including film property variations along the substrate. However, the deposition of a film, compared with the deposition on a pre-structured surface increases film stress. As the Si substrate provides a significant lattice mismatch, the film tension might be relieved by delamination. In order to prevent delamination, a Cr layer is added to improve adhesion. The thickness of the Cr layer is varied between 50 nm and 150 nm. After Cr deposition, NiMnGa is deposited by DC-magnetron sputtering at 500 °C. A schematic of a possible process flow based on nanomachining the NiMnGa after film deposition is given in Figure 2.2.

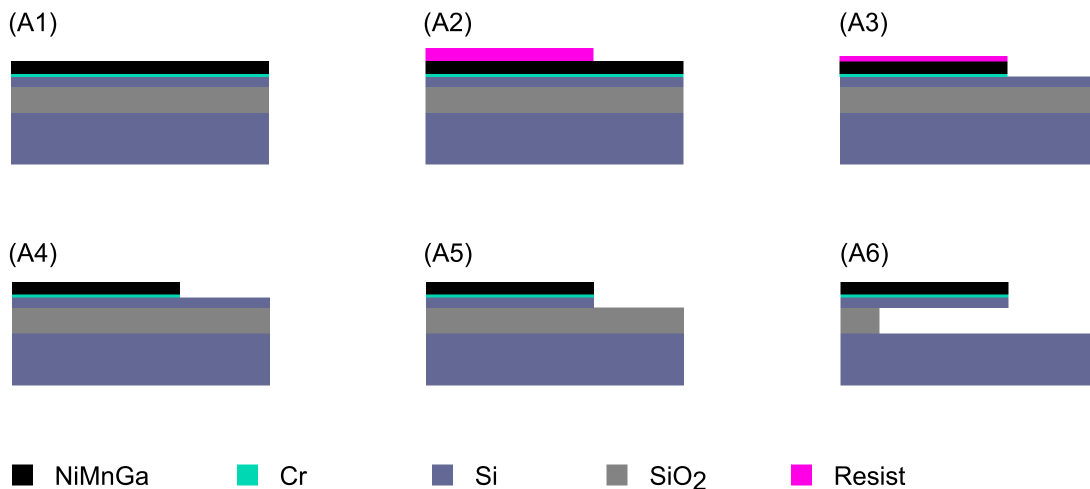

Figure 2.2: Schematic of process flow A, which is based on nanomachining the NiMnGa film.

Figure 2.2 (A1) shows the layer sequence of the initial substrate. From bottom to top, it is composed of the handle wafer of Si, a 2 µm thick box of SiO_2, a 340 nm thick Si device layer, a Cr layer and finally the NiMnGa film. In the following process, the sample is coated with a photoresist, which is patterned by EBL (A2). IBE by Ar ions is used to transfer the pattern from the resist to the NiMnGa and Cr layers (A3). Since NiMnGa belongs to the so called difficult-to-etch materials, it is not possible to chemically ablate it. Therefore, structuring by RIE is excluded. Afterwards, the resist is removed (A4) and the structured NiMnGa is used as a metal hard mask for structuring of the Si layer (A5). The last step comprises wet-etching in hydrofluoric acid to create free-standing double beam actuators (A6).

Figure 2.3 a) shows a SEM micrograph after process step A4. To create rectangular sidewalls the angle between substrate and ion beam was chosen to be zero. During the

etching the sample was set under rotation with a speed of 20 rounds per minute to provide a homogeneous etch rate along the sample. The temperature was kept constant and equal to room temperature. As can be clearly seen in Figure 2.3 a) severe redeposition occurs at the borders of the resist pattern. The generated sidewalls provide high aspect ratio and withstand an elaborated stripping procedure of 24 hours in ultrasonic aceton bath followed by two hours in an oxygen plasma.

For the next experiment series, a sequence of IBE steps is chosen to reduce the redeposition. First, the resist pattern is transferred into the NiMnGa by etching at zero degree, analogous to the first IBE procedure, but with a reduced process time. As a consequence, the NiMnGa layer is only partly etched and the created redeposition is significantly thinner. The pattern transfer step is followed by the ablation of the redeposition by IBE at an angle of 80° between substrate and ion beam. The two steps, first, pattern transfer at zero degree and second, angled etching of the NiMnGa-resist sidewalls, are repeated twice to etch through the whole NiMnGa layer. Since the redeposition is not completely removed after this procedure, another IBE step at an angle of 85° is performed subsequently. The result after 24 h ultrasonic aceton bath and 2 h oxygen plasma to remove any resist residues is shown in Figure 2.3 b).

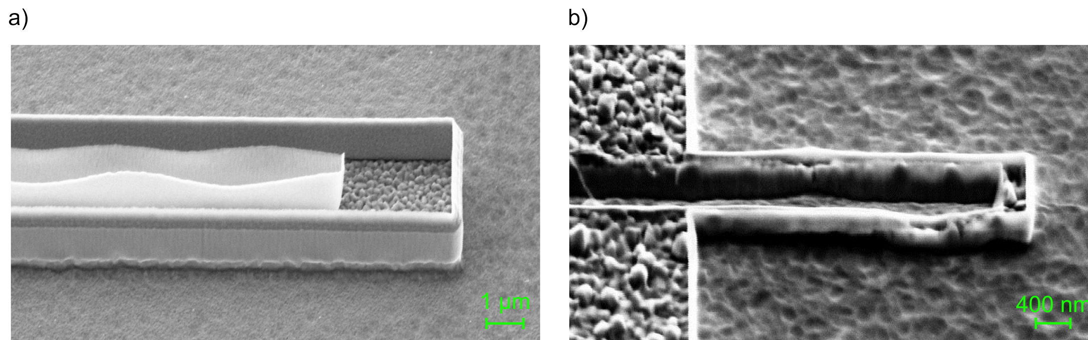

Figure 2.3: SEM micrographs of a nanomachined NiMnGa film after process step (A4). a) IBE at zero degree to transfer the resist pattern into the NiMnGa layer results in severe redeposition at the resist sidewalls. b) A sequence of IBE processes, consisting of a pattern transfer step at 0° and redeposition-removing step at 80°. The sequence is performed twice before an additional IBE step at 85° is used to further reduce the redeposition.

Even though the sidewall thickness and height has been reduced significantly, it could not be avoided completely. In detail, the redeposition is removed in some areas of the structure, but in other areas it still has a width of approximately 50 nm. An additional etching step damages the NiMnGa layer in the areas where no sidewall is left anymore. This damage is reasonable for micrometer sized double beams, but for beam widths of 200 nm and less the double beams are partly interrupted.

Summing up, the nanomachining of a thin film of NiMnGa provides a clear layer sequence, since the film growth of the NiMnGa layer is not influenced by a pre-structured substrate. Additionally this technique allows for maximum design flexibility. However, due to the limitations of the etching procedure and the corresponding occurrence of redeposition, nanostructures could not be fabricated following this approach during this work.

2.2 Nanomachining prior to film deposition

As the name implies, the NiMnGa is deposited in the last process step on a pre-structured substrate. In general, this method limits the layout options, but is associated with minimized substrate delamination and film brittleness, which might occur on a planar substrate. A schematic of the process described in detail in the following is given in Figure 2.4.

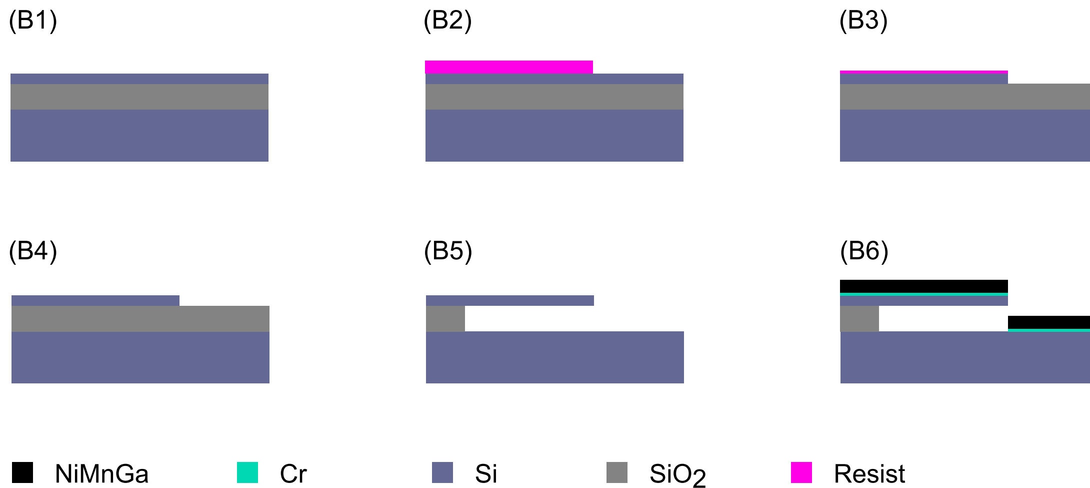

Figure 2.4: Schematic of process flow B based on nanomachining the Si prior to deposition of the NiMnGa film.

The substrate shown in Figure 2.4 B1 is a common SOI wafer chip with 340 nm Si device layer and 2 µm SiO_2. In the first process step, the substrate is coated with a photoresist. In this work, two negative tone resists MAN-2401 and MAN-2403 are used, which reduce the exposure time of the EBL compared to a positive tone resist. In normal atmosphere, the surface of the SOI wafer is oxidized and edged by hydroxyl groups due to humidity resulting in a polar surface. As a consequence, the adhesion of the nonpolar resist is very poor. A coupling agent, here TI Prime, is coated on the SOI wafer and cured for a few minutes to generate a nonpolar surface before the actual photoresist is deposited. A thickness of the TI Prime layer of a few nanometers is sufficient to achieve proper adhesion of the photoresist. Concerning the thickness of the photoresist, one has to strike the balance between minimizing the proximity effect, thus thin resist layer and thick enough to withstand the subsequent etching routine. Depending on the selectivity of the etch process, the resist layer is reduced during the pattern transfer into the substrate. If the resist layer is chosen too thin, the resist is normally partly perforated instead of being homogeneously etched. Hence, over-etching will result not only in a thinned material layer beneath the resist but also in a very rough surface. Here, a resist layer thickness of approximately 200 nm is chosen.

Process step B2 shows the exposure of the resist by EBL. Therefore, the layout is divided in a coarse layer and a fine layer. The fine layer includes the borders of the layout, here with a width of 10 µm. The coarse layer fills out the framed areas with a slightly smaller dose. The overlap of the two layers is set to 120 nm. The exposed SOI chips are placed

in a spray developer, where they are set in rotation and sprayed with the developer using air pressure. As a developer MAD-525 mixed with DI-water at the ratio of 2:1 is used. To stop the development, DI-water is sprayed onto the substrate, followed by spin drying. Compared to immersion development, spray development is faster, since the developer is not depleted during the development process. Besides, the automatization and therefore the strict compliance of mixture ratio, development time, rotation speed, drying time guarantees high reproducibility. After development, the exposed photoresist forms the mask of the subsequent etching step. The adhesion agent is not influenced by the development routine and can be removed by a dip of a few seconds in hydrofluoric acid (HF) before the Si etch process. Since its layer thickness is very thin, it is also possible to remove the TI Prime in the first seconds of the Si etch process. Since this reduces the number of process steps and therefore overall manufacturing time, the influence of the HF dip has been investigated in the beginning of the development of the fabrication process. With the used Si etch process, no influence of the HF dip is observed, besides the additional contamination that any immersion step involves. Therefore, the HF dip is skipped for the devices fabrication.

Process step B3 illustrates the pattern transfer from the resist into the Si device layer by RIE. As the process gas, HBr and oxygen at a temperature of 60 °C are used. The SOI chip is fixed by a special vacuum oil to improve the thermal contact between the Si dummy wafer and substrate. Besides the fabrication flexibility gained by the use of chips compared to whole wafers, the use of chips also minimizes the impact of varying layer thicknesses on the plasma stability. With the use of chips, the area seen by the plasma is mainly the surface of the Si dummy wafer and only partly the surface of the chip. Therefore, a changing chip surface, e.g. Si compared to SiO_2, or a different photoresist, will only marginally influence the plasma properties. This enables reproducible results. Additionally, monitoring of the plasma parameters, like color or DC bias, allows for fast detection of any abnormity during the etching routine.

In previous work, the fabrication of NiMnGa-Si bimorphs has been done by pre-structuring the Si using the cryo-process [15]. The advantages of the cryo-process are rectangular and smooth side walls. However the temperature has to be very accurate, otherwise the passivation needed for an anisotropic etch profile is affected. If the operation temperature is too high, slanted or rough sidewalls are the consequence. If the temperature is too low, black silicon is generated. Since the thermal contact can be easily affected by dirt particles at the backside of the substrate, for instance resist residuals, even the correct set of the process temperature might result in varying chip temperatures and thus varying etch profiles. Besides this challenge, the cryo-process provides very high etch rates of the order of 10 seconds per 100 nm. Moreover, the ignition of the plasma demands for a strike step with higher power. Accordingly, the strike step has a even higher etch rate. In a standard cryo-process recipe, the strike step lasts 3 seconds. However, the ignition can take place in the first second or the third second, which significantly influences the etch profile, if the whole process time is in the order of 10 seconds.

The developed HBr process is less temperature sensitive. By setting the temperature too low or too high, the angle of the generated profiles does change, for instance resulting in an undercut. However, a variation of the operation temperature within ±10 °C causes

the angle of resulting etch profiles to differ by less than 5°. Furthermore, the etch rates are of the order of 1 min per 100 nm Si. The plasma is very stable and does not demand for an additional strike step to start ignition. This leads to easier process control and reproducible results. Even though the SiO_2 layer acts as a stopping layer for the Si etch process for both processes, the extended process time will lead to an increase of lateral etch attack. As a consequence, it is favorable to set the process time very accurate. Due to unavoidable fluctuations of the etch rate, the Si layer sometimes is not etched through after the setted process time. Using the HBr process, it is no problem to repeat the etch process and possibly add another 10 seconds to the overall process time. In contrast, a re-etching using the cryo-process will most likely result in an over-etched structure, since already the strike-step will remove more than just the residual Si layer.

Summing up, the development of the HBr process allows for easier handling and higher reproducibility. Although the process time is extended compared to the cryo-process, the time needed for cooling the chamber and the sample to below -100 °C dominates the time for the fabrication step. Hence, the longer process time of the HBr process is still shorter than the time of the cryo process.

After the pattern is transferred from the resist to the Si, the resist leftover is removed by an aceton immersion bath (B4). After flushing with isopropanol and Di-water the samples are put into a HF immersion bath with 5 % buffered HF acid. Wet-etching with HF removes the buried oxide layer and generates free-standing Si double beam structures (B5). In parallel, the isotropic etching process partly removes the buried oxide beneath the contact pads leading to overhanging Si branches. An etching time of 2 hours and 15 minutes is chosen to completely dissolve the buried oxide beneath the double beam structures, while minimizing the size of the overhanging Si branches. This results in contact pads with about 3 to 4 µm extension of the undercut. Afterwards, critical point drying or freeze-drying has to be performed to avoid sticking effects. Figure 2.5 shows the results after freeze-drying with a) cyclohexane (C_6H_{12}) and b) DI-water. Using C_6H_{12}, sticking effects cannot be prevented completely. Even though some structures look as intended, over 50 % stick to the surface or the two beams of the double beam stick together. Additionally a lot of double beam structures show an internal twist. Following the freeze-drying routine with H_2O, nearly no sticking effects occur, but the device layer is party detached from the substrate. Since the wet-etching process of the buried oxide is isotropic, not only vertical etching takes place, but also the oxide in the lateral direction is removed. This leads to the release of the double beam structures but also to a undercut of the border of the contact pad. Is seems like the expansion of the H_2O while freezing leads to spalling of the undercut area.

The sample can also be dried by critical point drying to prevent the collapse of the double beam structures due to capillary forces. As explained in Chapter 1.3 the sample is first wet-to-wet transferred from the HF bath to a DI-water bath and then from the DI-water bath into an isopropyl bath. The isopropyl is evaporated by critical point drying to prevent any emerging capillary forces. Following this routine, freestanding silicon double beam structures with high aspect ratios are fabricated. Figure 2.6 a) shows a SEM micrograph of a freestanding silicon double beam structure having the dimensions of 50 nm x 9 µm x 340 nm in width, length and height, respectively. The observation of double beam

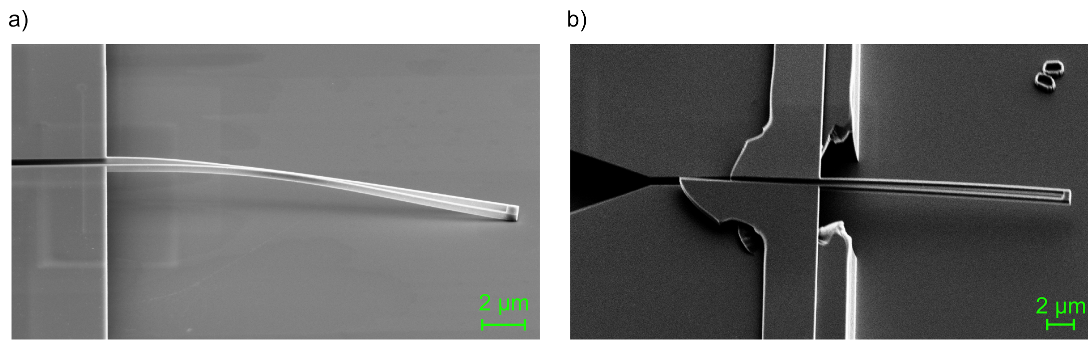

Figure 2.5: SEM micrographs of a nanomachined NiMnGa film after process step (B5). Freeze Drying using a) C_6H_{12} and b) H_2O.

structures 100 nm in width or less is accompanied with some challenges. During scanning with the SEM, Si is partly charged leading to sticking together of the two beams. A SEM micrograph, taken after the collapse occurred due to charging, is shown in Figure 2.6 b).

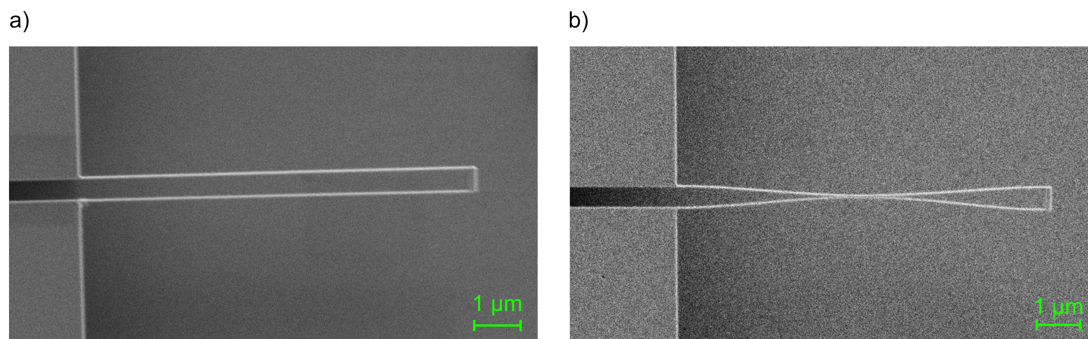

Figure 2.6: SEM micrographs of a nanomachined NiMnGa film after process step (B5) dried by critical point drying. a) Freestanding silicon double beam structure having the dimensions of 50 nm x 9 µm x 340 nm in width, length and height, respectively. b) Sticking effect due to charge induction by the scanning electron beam.

The device layer is thinned by a second dry etching process using HBr to decrease the thickness of the bottom silicon layer of the bimorph actuators to reduce their stiffness. Figure 2.7 a) shows a freestanding Si double beam with the initial device layer thickness of 340 nm illustrating the almost rectangular sidewalls and a smooth surface. Figure 2.7 b) shows a SEM micrograph of a structure after thinning the Si layer to 150 nm by a post-structuring procedure by RIE using HBr. The sidewall angle as well as the surface quality are not affected by the additional etching process. Furthermore, the width of the Si double beam structures is not influenced by the additional etch due to the high directionality of the HBr process. Both SEM micrographs are taken at an angle of 45 ° between sample and detector.

Finally, DC magnetron sputtering is used to deposit chromium as an adhesive layer at room temperature and 200 nm NiMnGa at 500 °C on the pre-structured sample (B6). Figure 2.8 shows two bimorph actuators with different Si device layer thickness a) 100 nm and b) 340 nm, respectively. Both micrographs show the deflection state at room

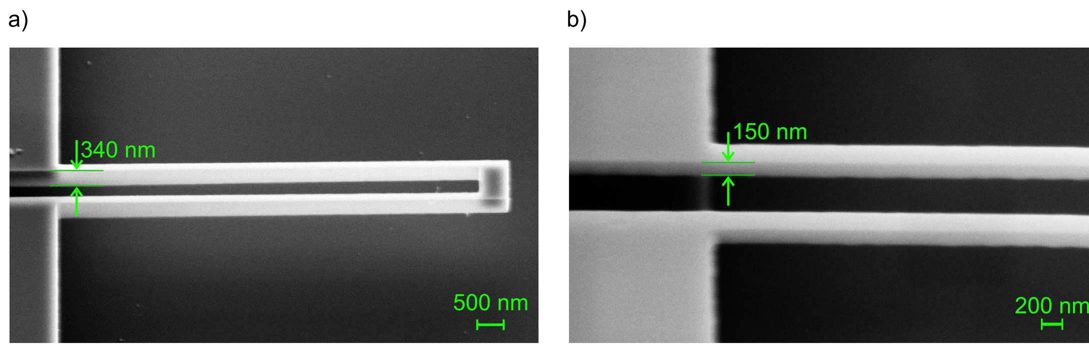

Figure 2.7: SEM micrographs of a nanomachined NiMnGa film after process step (B5) dried by critical point drying. a) Freestanding silicon double beam structure with the initial device layer thickness of 340 nm and b) with thinned of 150 nm Si layer thickness using an additional RIE process.

temperature. The reason for the different out-of plane deflections will be discussed in detail in Chapter 4. The width of the initial Si double beam structure is 50 nm for both. The DC-sputtering process is partly isotropic. Therefore, the NiMnGa is not only deposited at the top surface of the Si double beam structure but also at the sidewalls. Besides the sidewall deposition, also the grain size might increase the final actuator width. The average grain size of the polycrystalline NiMnGa film is about 50 to 75 nm in diameter, whereas the grain size of the NiMnGa/Si bimorph nanoactuators has a larger diameter of up to 150 nm. This value corresponds to the final actuator width that has initially been 50 nm for the Si double beam structure.

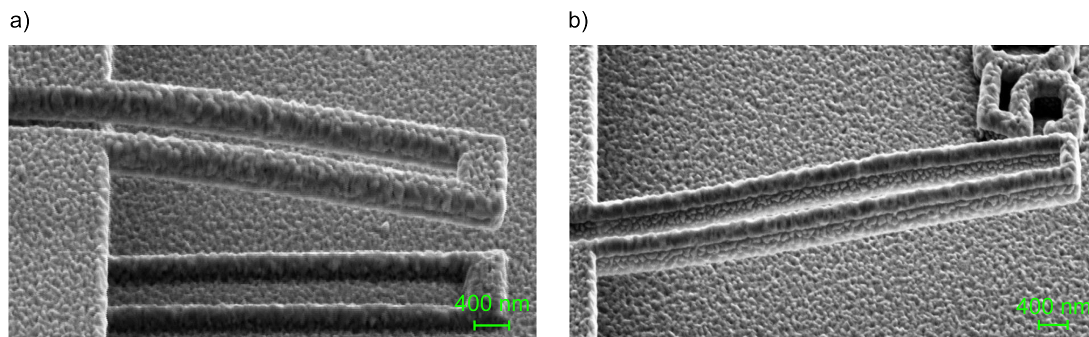

Figure 2.8: SEM micrographs of NiMnGa/Si bimorph actuator having an initial Si beam width of 50 nm. Due to sidewall deposition and the large grain size, the width increased to approx. 200 nm. The layer thicknesses are from top to bottom, 200 nm NiMnGa, 10 nm Cr and 100 nm Si (a), 340 nm Si (b).

The sidewall coating can be removed by an additional RIBE process step to generate bimorphs with the initial Si double beam widths. Therefore, the chip is mounted at an angle of 85° with respect to the Ar beam of the RIBE system. Due to the chosen angled etching process, the attack of the surface NiMnGa is minimized, whereas the ablation of the sidewalls is maximized. The Ar-etching takes place under rotation to remove the sidewall coating homogeneously at all edges . In Figure 2.9 a) a schematic of the rotation and angle of the substrate during the sidewall removal via RIBE is shown. In Figure 2.9 b) a NiMnGa/Si bimorph with the layer thicknesses of NiMnGa and Si of 340 nm and

200 nm is presented after etching for 5 min with rotation speed of 20 rpm at an angle of 85°.

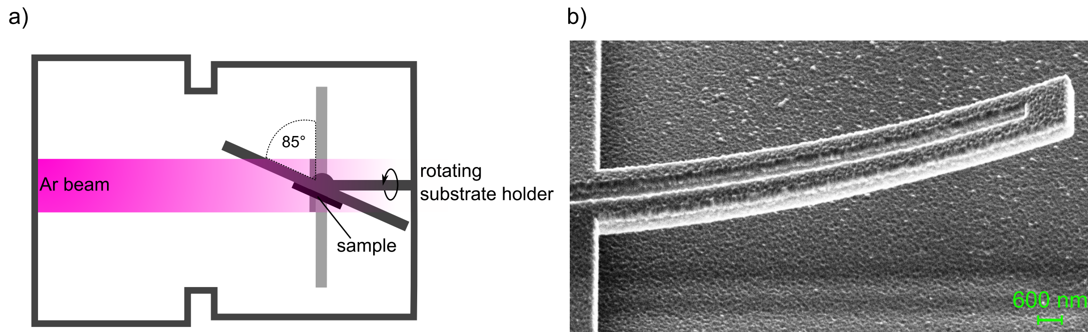

Figure 2.9: a) Schematic illustration of the chip positioning during removal of the sidewall coating generated by the NiMnGa sputter deposition process. b) SEM micrographs of NiMnGa/Si bimorph actuator after a 5 min at 20 rpm and 85° RIBE process.

In this case, the layer sequence is more pronounced, nevertheless there is still NiMnGa visible on the double beam sidewalls. Furthermore, the surface has changed significantly shown by the different grain structure of the NiMnGa film in Figure 2.8 compared to Figure 2.9. Any change of surface is surely an indicator for changed film properties making comparable results difficult. Even though the post-structuring by RIBE with Ar does not influence the NiMnGa film chemically, the impact of the mechanical ablation is quite big. Therefore, the removal of the sidewall deposition has been abandoned to make sure that the nanostructured samples provide the same film properties than the non-structured thin film reference structures.

Following the described fabrication process, freestanding NiMnGa/Si double beam structures with varying widths between 50 and 2000 nm have been prepared. A list of the investigated devices is shown in Table 1 of the Appendix. Besides, the nanomachining of Si prior to NiMnGa deposition provides the advantage of a guaranteed functional crystalline film. Any impairments of the composition and structure of the SMA by structuring processes are prevented up front.

2.3 Fabrication of SMA-based Optical Switch

The layout of the optical switch developed during this work is sketched in Figure 2.10 a). The concept is based on positioning a mechanically active SMA/Si bimorph nanoactuator in the close vicinity of a tapered Si waveguide. Due to the tapering a substantial part of the optical mode profile lies outside of the tapered waveguide. Therefore, any object being in the direct surrounding of the tapered Si structure will effect the optical mode. By approaching a Si cantilever close to the field distribution of the waveguide mode, the mode is extracted by the cantilever, while the waveguide mode is attenuated. This principle allows to actively control the transmission signal within the waveguide, without further limiting the optical bandwidth. As discussed in the previous section, the design

of the SMA/Si bimorph nanoactuators can be adjusted in a wide range to optimize the scattering amplitude. A detailed discussion of the operation principle is given in Section 4.3.

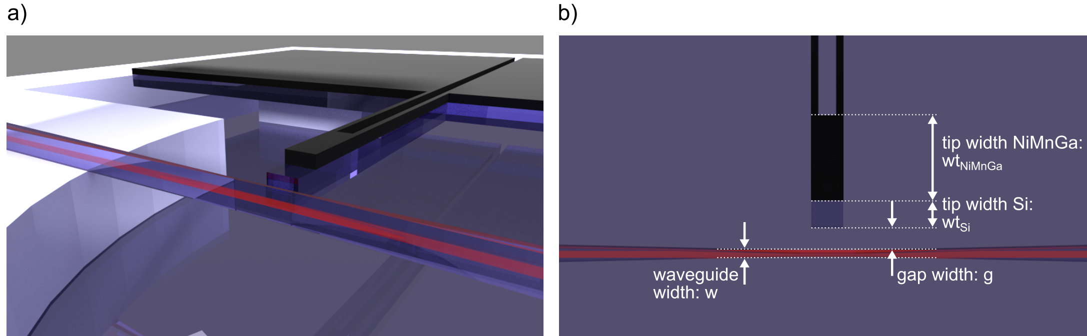

Figure 2.10: a) Schematic layout of SMA-based NEMS actuator and tapered Si waveguide. Actuation is in out-of-plane direction due to the combined shape memory and bimorph effect in the evanescent field of the tapered waveguide. b) top view of the setup shown in a) including the characteristic geometry parameters.

To achieve a change of optical transmission of about 3 dB the SMA-based nanoactuator has to provide a *sufficient* actuation stroke, thus a *sufficient* change of gap size between waveguide and Si cantilever. In Section 4.3 the relation between actuator deflection and corresponding change of the optical transmission signal is approximated by finite element analysis. Aside from the exact values, it is quite obvious that the critical dimensions of the layout have to be maintained with high precision. The most sensitive parameter in this system is the gap width g, which is defined as the distance between tapered waveguide and Si cantilever. To enable minimum transmission through the waveguide, hence maximum coupling into the Si cantilever in off state of the optical switch, the gap size has to be as narrow as possible without mechanical contact. Therefore, a process flow is developed to simultaneously fabricate SMA based nanoactuators, Si waveguide structures and grating couplers on a common SOI wafer. As grating couplers, a common single-step grating coupler design is taken from literature to couple light in and out of the Si waveguide structure.

The SMA film is deposited after nanomachining the Si, according to the process flow presented in the previous section, to guarantee for a functional crystalline NiMnGa/Si nanoactuator. Therefore, the fabrication of the waveguide structure has to be prior to the NiMnGa deposition. A schematic of the process flow is shown in Figure 2.11.

By the first EBL step, the Si-base layer of the SMA/Si nanoactuator, the Si waveguide and the associated grating couplers are fabricated simultaneously. As a substrate, a SOI wafer with 340 nm device layer and 2 µm SiO_2 oxide is chosen. The parallel fabrication, more precisely the exposure of Si waveguide structures and the base-layer of the nanoactuators in one e-beam step, enables maximum precision. Compared to a two step process which is associated with alignment inaccuracies of different layout layers, e.g. the waveguide structures and the nanoactuators, the adverse influences are minimized. Afterwards, the pattern is transferred to the Si layer by RIE using HBr at about 60 °C (b1). With the appropriate combination of gas mixture and temperature an anisotropic etch profile with

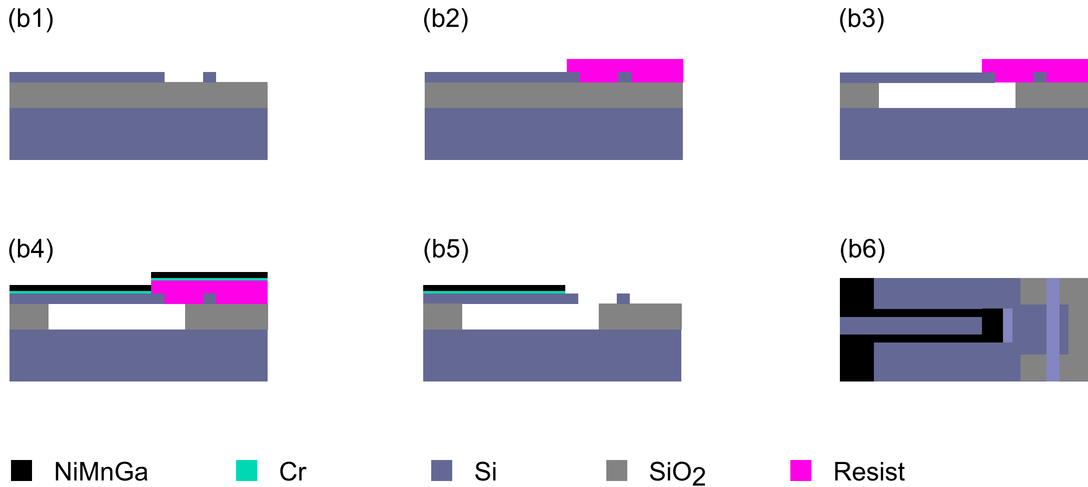

Figure 2.11: Schematic process flow for simultaneous fabrication of SMA/Si bimorph NEMS actuator and Si waveguide structures: (b1) first EBL and Si RIE, (b2) second EBL, (b3) wet-etching of SiO_2, (b4) DC-magnetron sputtering of SMA film, (b5) lift-off. (b6) Top view of the final structure.

rectangular and smooth walls is obtained. Being not of prior interest for the SMA/Si bimorph nanoactuator fabrication, the roughness of the Si structures is now crucial, since it is directly linked to the waveguide loss. Especially in the tapered region the transmission is highly sensitive to any structural changes on the waveguide surface. Figure 2.12 demonstrates minimum gap size $g = 50$ nm and minimum waveguide width at the tapered region of $w = 150$ nm. The lateral dimension of the Si double beam structure is 50 nm at a beam length of 5 µm. The layer thicknesses of the double beam structure and the waveguide are equal to the Si device layer thickness of 340 nm. As already mentioned the surface quality of waveguide structures is very important for their optical transmission properties. A test layout is fabricated with several reference waveguides consisting of two 2 mm long straight segments, one s-shaped section, having a bending radius of 50 µm and two grating couplers. For this layout, the average transmission loss is determined to be only 6 dB at 1550 nm demonstrating the excellent surface quality of the etch profile. More details on the optical measurement setup are given in Chapter 3.3.

The second EBL step defines the pattern of the SMA top layer of the bimorph nanoactuators and the control zone (b2). The accuracy of the two EBL steps, hence the two exposure layers, have to be in the order of the desired tip width and not in the order of the gap size. In the schematic shown in Figure 2.10 b) the characteristic geometry parameters are indicated including the Si tip width wt_{Si} and the NiMnGa tip width wt_{NiMnGa}. The alignment accuracy of the two EBL steps does not have to be very accurate in the direction along the waveguide and has only to be in the order of wt_{NiMnGa} perpendicular to the waveguide. Since wt_{NiMnGa} is in the order of several hundreds of nm being a magnitude larger than required gap sizes g, the presented process flow is very robust concerning fabrication irregularities.

To create freestanding SMA/Si bimorph double beams as nanoactuator devices, the buried oxide beneath the double beam structures is removed by a wet-etching step in hydrofluoric

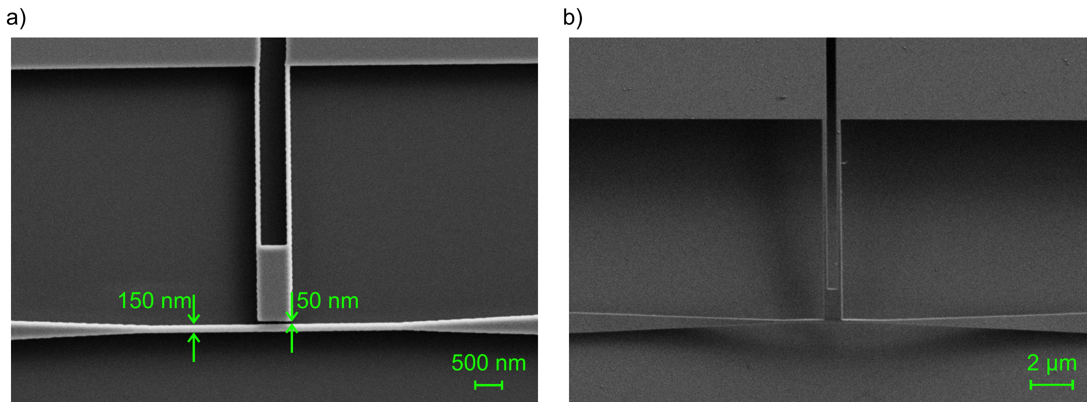

Figure 2.12: SEM micrographs of the control zone a) after process steps b1, and b) b3 as illustrated in Figure 2.11. Gap size (g = 50 nm) and waveguide width (w = 150 nm) are indicated.

acid (b3) similar to process step B5 in Figure 2.4. A corresponding SEM micrograph is shown in Figure 2.12 b).

Afterwards, the active material plus the Cr adhesion layer are deposited by DC magnetron sputtering (b4). As the final process step the SMA is removed from the control zone by a lift-off process (b5). Up to now, it has not been possible to perform the process steps b4 and b5 due to the lack of resists, which can withstand the elaborate HF wet-etching time of more than 2 hours. As an alternative, a metal lift-off mask could be used. Therefore, the resist structured in the second EBL step is used as a lift-off mask for the preparation of, e.g., a Cr mask. A Cr mask provides the advantage, that it is not attacked by HF acid. The additional process steps for the alternative fabrication using a Cr mask are illustrated in Figure 2.13.

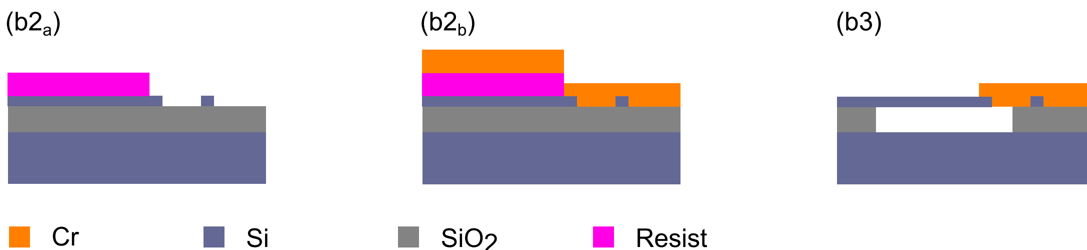

Figure 2.13: Schematic of an alternative process flow using a Cr lift-off mask: (b2$_a$) second EBL, (b2$_b$) deposition of Cr, (b3) lift-off and wet-etching of SiO_2. The process steps (b1), (b4) and (b5) are equal to Figure 2.11.

As a summary and to gain an overview of all the processing steps for the integrated optical waveguide switch, a process-chart can be seen in Figure 2.11. Therein all steps of the prior discussed workflow b are indicated, which have to be done in order to get the final chip design. An initial E-beam step for the marker exposure is indicated in step 0., which is necessary to align the proceeding E-beam steps relative to each other. Furthermore, the alternative discussed in the last section is depicted as a possible option after the second E-beam step. The option with a Cr lift-off mask provides the advantage of high temperature deposition of the active material, whereas the resist mask demands for room temperature

deposition, thus an additional annealing step after the resist lift-off. However the highly directional deposition of a Cr layer thick enough to perform the lift-off process implies fabrication difficulties. Moreover the Cr etching solution can affect the NiMnGa layer in functionality and structure. Therefore the alternative using a resist lift-off mask contains less process steps and less uncertainties which makes it the favorable option at this stage.

Besides all process steps have to be performed with high precision, extreme care and in a cleanroom environment to fabricate functional devices.

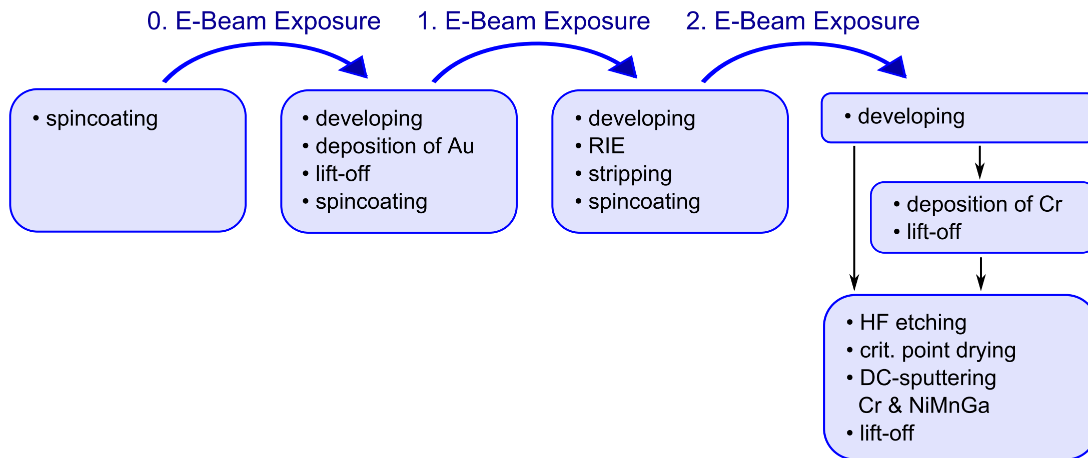

Figure 2.14: Process chart of the workflow b to fabricate the integrated optical waveguide switch, including the marker indication and possible alterations.

3 Measurement Setup

The fabrication process allows to generate a large number of devices on one chip. In-situ measurement methods inside a SEM are essential to characterize these nanostructures. In previous work, the actuation by Joule heating with real-time resistance sensing as well as recording the deflection with the SEM have been presented. In the first section of this chapter, the details of this measurement setup are discussed. Even though Joule heating provides the most promising actuation for several applications and therefore also forms the major part of the experiments done during this project, it generates in parallel a temperature gradient along the double beam structure. This gradient makes it quite challenging to identify size effects, since the phase transformation starts and finishes at different sections of the double beam. Therefore, a key task in this thesis has been to develop a setup capable of homogeneous heating. The underlying concept as well as its implementation are discussed in detail in the second section. Finally, the optical measurement setup is illustrated, which is used to characterize the SMA-based optical switch device.

3.1 Temperature Gradient

To actuate individual double beam structures by Joule heating, they have to be electrically connected. A nanomanipulator setup is used for successive probing without transferring the sample out of the SEM between each measurement. The nanomanipulators enable to move a tungsten wire onto a double beam contact pad with nm resolution. The associated contact resistances are quite high and may even be subjected to fluctuations. Therefore, a current controlled four-terminal measurement scheme is used. Two connections are enabled by nanomanipulators addressing a single double beam structure and two connections being fixed and positioned on a relatively big contact pad, which is electrically connected to all double beams on the sample. With this setup the flexibility provided by the nanomanipulators is used, without suffering under the associated contact resistances. A schematic of the chip layout including illustrations of the measurement setup is shown in Figure 3.1.

A current is passed through the outer probes and induces a voltage in the inner voltage probes by a Agilent B2900 Source Measure Unit (SMU). For increasing electrical power, the double beam gets heated up, while its intrinsic resistance is measured to monitor the progress of phase transformation in real time. The SMU is controlled by a Labview script, which stores the current, voltage and the time in milliseconds and also plots the resistance versus electrical power curve in real time. With the latter, it is possible to notice the progression of phase transformation. Thereby, the maximal heating power can

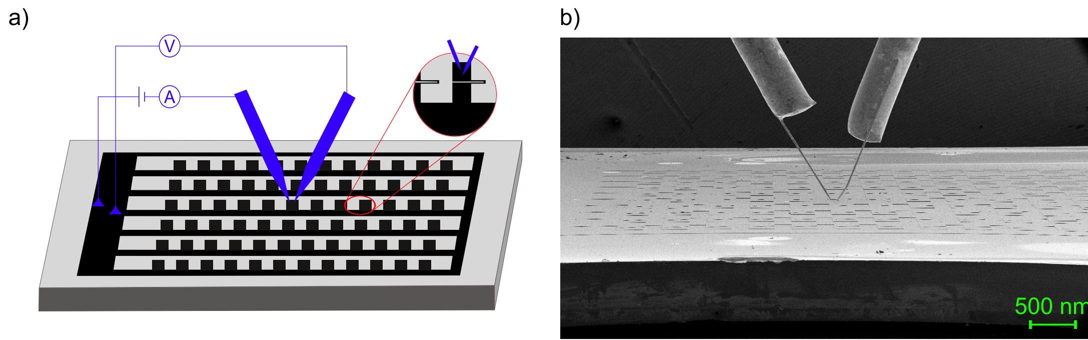

Figure 3.1: a) Schematic of the in-situ Joule heating measurement setup illustrating the chip layout consisting of one large contact pad and various double beam structures as well as the electrical connections. b) SEM micrograph of the chip mounted at 85° to gather the out-of plane deflection of the SMA/Si bimorph double beams. The tungsten tips connected to the nanomanipulators enable to electrically connect single double beam structures in succession.

be set high enough to make sure that the structure undergoes a full phase transformation and, on the other hand, small enough in order not to risk destroying the beam due to overheating. Simultaneously to the electrical resistance measurements, the deflection of the nanoactuator is determined by recording video sequences with the SEM. As discussed in Chapter 4, the SMA/Si bimorph structures deflect out-of plane. Therefore, the sample is mounted at an angle of 85° with respect to the detector. A cross correlation approach is used to determine the net deflection of the actuator from the recorded SEM video data. The raw deflection data is influenced by the drift caused either by a movement of the sample or the electron beam itself. Individual parts of the structure have to be tracked to determine global movements within the measurement versus deflection or deformations to compensate the drift. The correlation approach is based on a convolution of the whole image with a pre-defined area of the first frame of the video sequence. Therefore, a free software package [67] is executed with the commercial software Matlab. At first, a non-deflecting geometry feature is defined as a marker region in the first frame of the SEM video data. The position of the marker region is then tracked throughout the measurement stack, hence throughout all frames of the video sequence, which determines the movement of the markers in horizontal and vertical direction. In the following, the video sequence is corrected by the marker movement, so that non-deflecting geometry features stay fixed. A suitable structure for drift compensation can be, for example, the base of the double beam or the etched inscription. In Figure 3.2, an representative marker array (blue) is positioned on two suitable geometry features. In the first frame of the drift-corrected video sequence a marker region is defined at the font of the actuator (red). Using the same algorithm, the marker region is tracked, and thus, the position of the beam tip is found for each frame of the corrected video sequence. The displacement of the individual markers is then used to determine the net deflection of the actuator font. By way of illustration, only the displacement of one marker of the red highlighted marker region is indicated in Figure 3.2.

For all measurements, the frame rate is the limiting factor concerning the temporal resolution. At a resolution of 1280 x 1024 pixels the maximum frame rate is 8.3 Hz, at

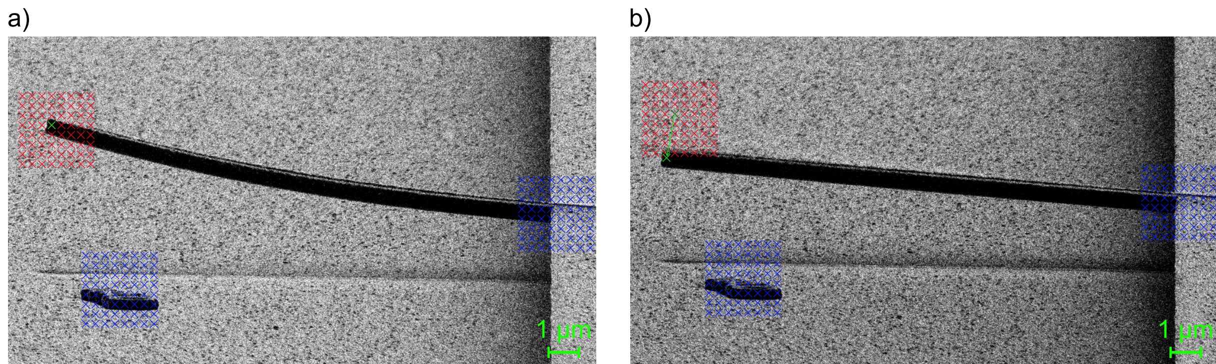

Figure 3.2: Scheme to determine the net deflection of the actuator by digital image correlation. A set of markers on fixed geometry features, highlighted in blue, are used to compensate the drift of the SEM video sequence. In the corrected video, markers are positioned at the tip of the double beam, here highlighted in red, to evaluate its deflection. To illustrate the deflection by Joule heating, two frames of a representative video sequence are shown a) frame 1 and b) frame 177 of 351 frames.

higher resolution even smaller. A higher spatial resolution leads to a better resolution of the deflection value but, on the other hand, the frame rate has to be maximized to allow for a high-resolution monitoring of the deflection over heating power. For most measurement data, an accurate tracking is enabled by a fixed marker size of 25 x 25 pixels and a precision of their position tracking of 1/1000 th of a pixel. Due to the varying initial bending of the double beams and the contrast change due to the temperature distribution sometimes this setting has to be modified.

3.2 Homogeneous Temperature

One key element in this project is to develop a measurement setup, which enables to homogeneously heat the sample. In comparison to Joule heating, the application is not the major prospective, but to get a deeper understanding of the physical processes, which govern the deflection for decreasing lateral dimensions. To monitor the deflection of the nanometer sized specimens, the prospective is to assemble the whole setup inside the SEM, which demands for vacuum feed-throughs for all connections. Besides this technical challenge, the heat transfer is dominated by heat conduction, and not by heat convection as for atmosphere, since the SEM working pressure is of the order of 1×10^{-6} mbar. As a consequence, the setup has to guarantee that the (hot) sample is not in direct thermal contact to the SEM stage. The SEM stage, including the stage motors used for moving the sample during the scanning, might be impaired by the temperature variation. Hence, not only the chip temperature but also the substrate holder temperature should be monitored. In the first setup, a Peltier element is included in the sample holder to heat the sample and to accelerate the reverse transformation by cooling with the same Peltier element. A schematic of the first design of the substrate holder is shown in Figure 3.3. As can be seen up front, the design is chosen to be very robust concerning mechanical influences. The former setup has always suffered under insufficient rigidity due to the rotation ability of

the glass plate inside the holder. For more information about the previous sample holder design, see [15].

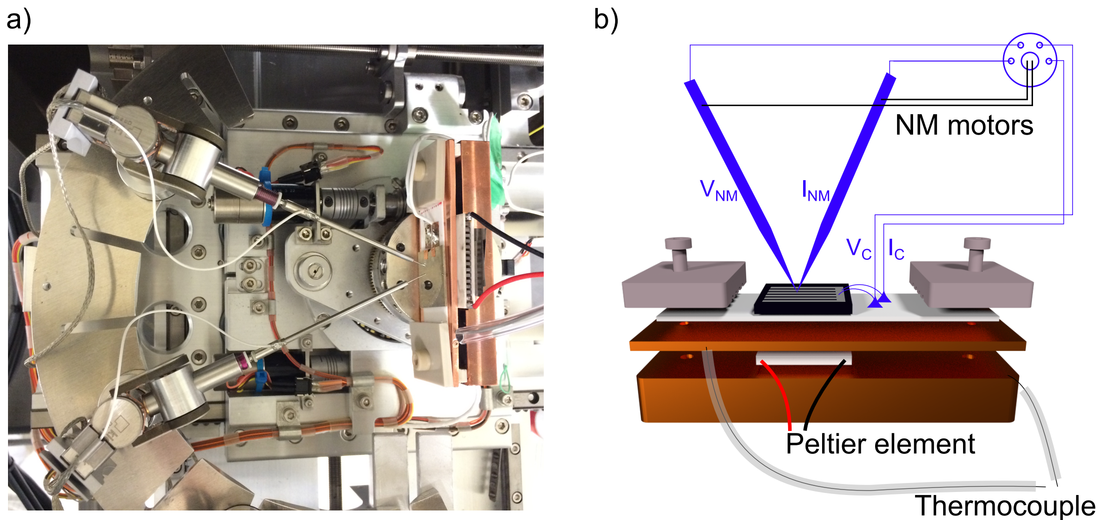

Figure 3.3: Schematic of the first in-situ homogeneous heating measurement setup. a) Picture of the sample holder and the nanomanipulators setup. b) Schematic of the sample holder layout including the necessary connections. First experiments showed significant heating of the copper base, which is in direct contact to the SEM stage. Therefore a second design has been developed, which is shown in Figure 3.4.

The chip temperature and the substrate holder temperature is monitored to avoid any damage of the SEM stage during the measurement. Therefore, a thermocouple is plugged into the copper base of the sample holder. A second thermocouple is plugged into a copper plate which is in direct thermal contact to the chip and the Peltier element. Due to the copper plate, a homogeneous temperature distribution is enabled along the chip. The sample itself is glued to a ceramic transfer plate to simplify the handling and mounting. Again, a four-point-probe setup is used to measure the electrical resistance. Therefore, two contacts are set by two movable nanomanipulators, and two contacts are bonded to the contact pad, as already illustrated in Figure 3.1 a). The wires are not directly connected to the cables connected to the vacuum feed-through, they are bonded to small copper contacts, which itself are glued to the ceramic transfer plate to minimize the mechanical force on the bond contacts. A standard cable is soldered to the copper contact plates, which can withstand tension created during the mounting process much better. First tests with this sample holder show promising results but lack in available temperature range. Even though the Peltier element is capable of generating up to 200 °C in atmosphere it is not possible to go beyond 150 °C due to heating up of the copper base. The copper base is in direct thermal contact to the SEM stage, which might be temperature sensitive and therefore should not heat up significantly. Besides, if the copper base heats up, the effect of using the Peltier element to cool the sample, hence to fasten the reverse-transformation, vanishes. Thus, a second design is developed and manufactured, which is illustrated in Figure 3.4.

In the new design, the copper base is used only to guarantee for mechanical stability and to monitor the approximate temperature of the SEM stage by measuring the copper

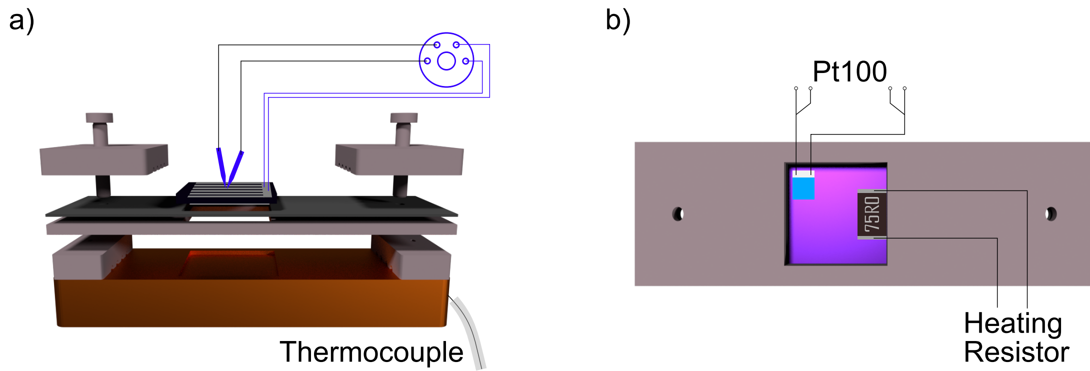

Figure 3.4: Schematic of the final in-situ homogeneous heating measurement setup. The thermal contact between sample and SEM stage is reduced by screws of PEEK and an additional plate of PEEK. The heating resistor, as well as the PT100 temperature sensor, are directly glued to the chip to enable accurate temperature control and sensing. The electrical connections correspond to the previous setup.

base temperature with a thermocouple. The heating of the chip is enabled by a 75 Ω resistor directly glued to the backside of the chip. Also glued to the backside of the chip is a PT100 temperature sensor. To enable optimal temperature sensing the PT100 is also measured by a four-point-probe setup. The sample transfer plate is changed from ceramic to a certain type of heat-resistant polymer sheet, to gain flexibility concerning its geometry, e.g. size of the cut-out area. The screws as well as the plate to enable good mechanical stability of the polymer transfer sheet, are made out of PEEK to guarantee for a low thermal contact between chip and copper base. The electrical connections are kept the same as for the previous assembly.

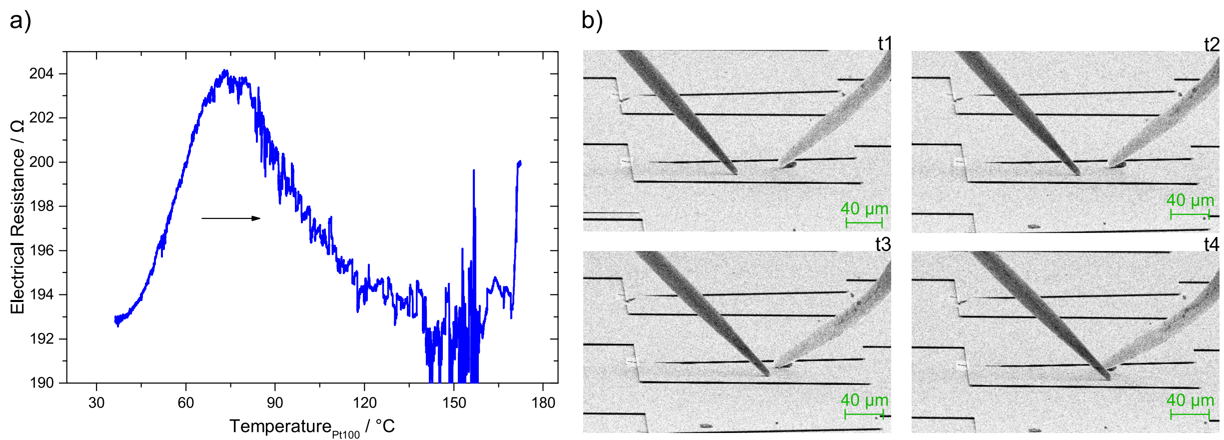

Figure 3.5: a) Temperature dependent electrical resistance characteristics of a NiMnGa/Si double beam. Due to the temperature change the whole sample moves in reference to the nanomanipulators which leads to a bending and scratching of these on the contact pad, as indicated in b) for the frames t1 to t4.

Figure 3.5 a) shows the electrical resistance as a function of the chip temperature measured by the Pt100 of a double beam structure. The characteristic slope of the phase transformation is clearly visible. However, compared to electrical resistance data measured

by Joule heating the signal shows a high noise level, especially at higher temperatures. This is due to the change of the contact resistance during the measurement. Due to the homogeneous heating not only the single double beam but the whole chip undergoes a considerable thermal drift. This leads to a displacement of the nanomanipulators in reference to the chip. This movement is captured in a video of which 4 frames t1 to t4 are shown in Figure 3.5 b). The movement of the tips of the nanomanipulators is of the order of several tens of µm and not only in the in-plane direction but a combination of in-plane and out-of plane, because of the 85° tilting of the sample. Therefore, the quality of the electrical contact changes with changing temperature. Additionally, the manipulators create macroscopic scratches on the NiMnGa surface. Therefore, the electrical resistance measurement by homogeneous heating has been made by fixed bond contacts outside the SEM and the new sample holder is used solely to study the deflection characteristics.

For the electrical resistance measurement outside of the SEM, two 25 µm thick aluminium wires are bonded onto one contact pad of an SMA/Si nanoactuators and two wires are bonded onto the second contact pad to generate a four-point-probe measurement setup. Heating and temperature sensing of the chip is enabled by the same combination of a 75 Ω heating resistor and a PT100 sensor as for the experiment inside the SEM. By an increase in heating current, the chip temperature is increased and the resistance change is measured. The resistance characteristics does not suffer under changes in the electrical contact properties and therefore allows for a reliable data set to compare with. The electrical resistance characteristics by homogeneous temperature variation of different double beam structures are presented in Section 5.1.

3.3 Optical Setup

At first, the quality of the fabricated waveguide structures is optically characterized inside the SEM. If there is significant roughness visible, the transmission is expected to be very poor. If there is no significant roughness visible, hence roughness values below approximately 10 nm, the next step is to perform optical transmission experiments. Compared to the qualitative statement obtained by SEM inspection, the transmission value through a well-defined waveguide geometry allows a quantitative analysis of the fabricated structure. Therefore, any chip design provides a certain amount of reference structures. The reference structures as well as the actual devices consist of two grating couplers, two 2 mm long straight segments and one s-shaped section, having a bending radius of 50 µm. In the actual device design, an additional tapered waveguide section is added, whereas the waveguide width is kept constant in the corresponding reference structures. For the realization of the optical transmission experiments, laser light with a wavelength of 1550 nm is coupled into an optical fiber. The optical fiber and the grating coupler structures have to be aligned by adjusting the distance and the angle of the optical fiber in reference to the grating couplers to couple light into the waveguide structures, t. Therefore, each fiber is fixed to a multi-axis stage providing µm resolution in the three translational directions x-y-z and the two rotational directions r-ϕ. The transmitted light is coupled in the same manner out into an optical fiber and its intensity is measured by a common photodiode. The signal is monitored by an oscilloscope, which simplifies the adjustment of the optical fibers. A schematic of the measurement setup is depicted in Figure 3.6 a). The measurement routine is shown in Figure 3.6 b) illustrating the array of devices to easily measure them in sequence.

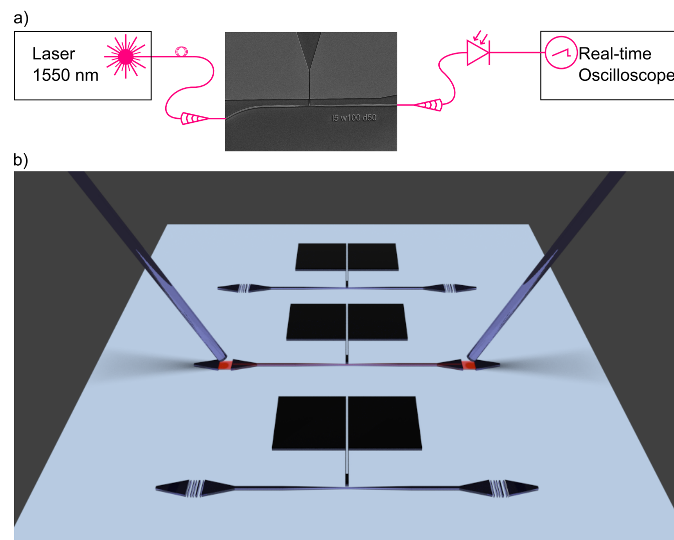

Figure 3.6: a) Measurement setup: Light of an optical fiber at wavelength $\lambda = 1550$ nm is coupled into a Si waveguide by a grating coupler, transmitted to the SMA-based actuator and coupled into an outgoing optical fiber. The transmitted light is measured by a photodiode and recorded on a real-time oscilloscope. b) Measurement routine: After aligning the optical fibers once, the individual devices are measured in sequence by moving the chip orthogonal to the waveguide axis.

4 Simulation

In parallel to the fabrication and experimental characterization of SMA-based nanoactuators, the simulation of their thermo-mechanical behavior has been a major part of this work. It is favorable to estimate the actuation stroke before starting the fabrication to develop an efficient actuator design. Especially for nanoscale devices, the fabrication is quite time-consuming and expensive. Therefore, it is not advisable to examine the functionality of a design idea by building up a test device. Additionally, the set up of an appropriate model is beneficial from the fundamental research point of view. Even a simple continuum model describing the macroscopic behavior of a specimen can help to identify emerging size-effects. In the following, the computational methods to simulate the SMA/Si bimorph behavior are introduced. Starting with the general description of the used software package and the basic assumptions, the mechanical behavior is discussed at first. Afterwards, the simulation of the electrical properties is introduced before the performance of the SMA-based optical switch is presented in the last section. Parts of the results presented in the following chapter have already been published in Reference [93].

4.1 Structural Mechanics

At first, the initial state has to be defined to estimate the mechanical performance of a NiMnGa/Si actuator. The initial state for an actuator device is normally denoted as the condition at room temperature (RT). As discussed before, the chosen fabrication method is to pre-structure the Si sample and deposit the NiMnGa at $T_{sputter} = 500\,°C$ subsequently. After the high temperature deposition the NiMnGa/Si bimorph cantilevers are cooled down to RT. The emerging RT state provides a certain stress distribution associated with a non-zero deflection value, hence the RT state is non-trivial. In the following the initial state is defined as the high temperature state, where the bimorph is supposed to be strain-free and planar. In Figure 4.1 a scheme illustrating the general deflection behavior of a NiMnGa/Si cantilever upon cooling is shown. Starting from planar state at $T_{sputter}$, the Si as well as the NiMnGa layer contract according to their individual thermal expansion coefficients. Caused by the differential thermal contraction, stress is induced, leading to a deflection of the bimorph. The larger thermal expansion coefficient of NiMnGa compared to Si leads to an upward out-of plane deflection. This behavior is independent on the phase transformation and sets the deflection in regime I in Figure 4.1, which denotes the temperature range $T_{sputter} > T > M_S$. In the following, the deflection resulting from the different thermal expansion coefficients of NiMnGa and Si will be denoted as the **bimorph-effect**.

In regime II, $M_S > T > M_F$, NiMnGa transforms from its high temperature phase, austenite, to martensite. Since austenitic and martensitic phase do not exhibit the same material properties, the change of phase will result in a change in the stress distribution, hence a change in the deflection behavior. Any effect arising from different material properties of the two phases of the NiMnGa, e.g. different Young's moduli or varying thermal expansion coefficients, will be denoted as a **ΔP-effect**. Besides the change of material parameters, the transformation itself can enable stress relaxation due to the pseudo-plastic stress-strain behavior of martensite. If the conditions are fulfilled and stress relaxation occurs, the deflection will be reduced. This effect is solely due to the martensitic transformation and will be denoted as the **SME-effect**.

In regime III, $M_F > T > RT$, the bimorph deflects corresponding to the differential thermal expansion coefficients, similarly to the temperature range $T_{sputter} > T > M_S$, hence due to the bimorph-effect. In Figure 4.1 the deflection in z-direction as well as the corresponding contraction of the NiMnGa and Si layer is illustrated for the different temperature spans.

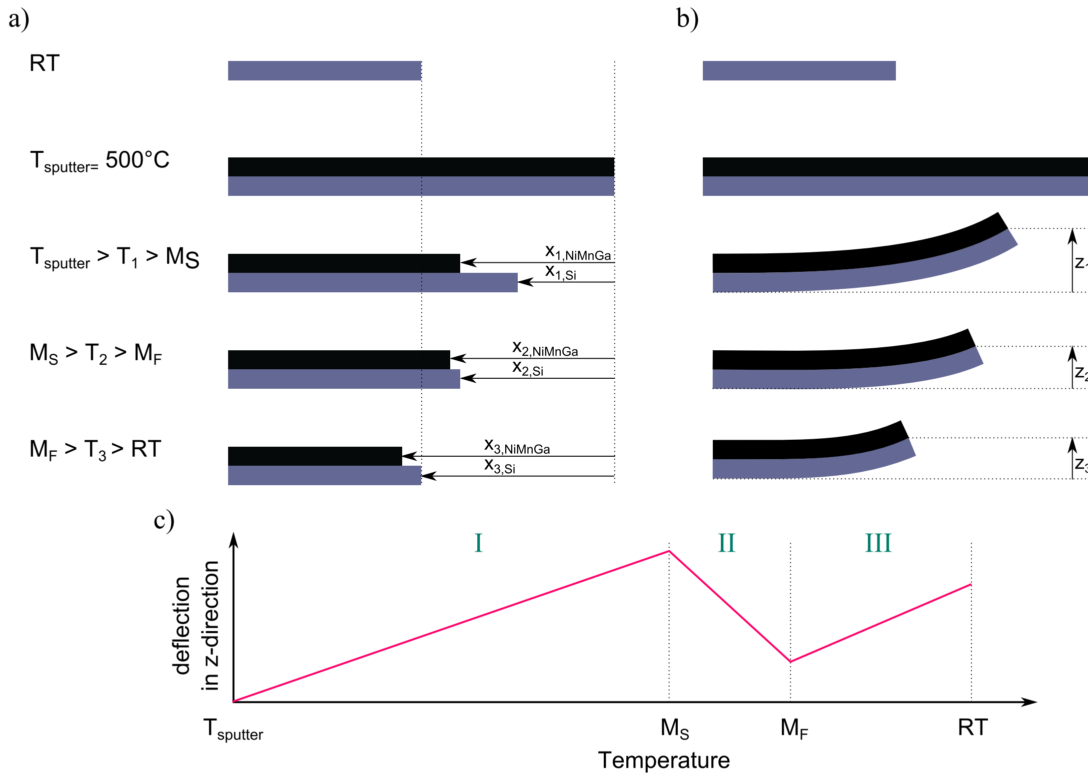

Figure 4.1: Scheme of the deflection behavior of a NiMnGa/Si cantilever upon cooling. a) Illustration of the differential thermal contraction of the NiMnGa and Si layer and b) corresponding out-of plane deflection of the bimorph. c) Schematic of the resulting temperature-dependent deflection characteristic including the definition of regimes I, II and III as described in the text.

The finite element software package COMSOL Multiphysics is used, particularly the Heat-Transfer module and the Nonlinear-Solid-Mechanics module, to model the deflection upon cooling from the sputter deposition temperature to RT. In order to distinguish between the Bimorph-, ΔP- and SME-effect a finite element (FE) model with several extensions

is set up. The basic version, in the following referred to as model #1, solely considers the bimorph effect, describing regimes I and III in Figure 4.1. The bimorph effect is caused by the differential thermal expansion of the NiMnGa and Si layer with temperature T. Approximated by the Heat-Transfer module of COMSOL, the induced thermal strain ϵ_{th} is given by

$$\epsilon_{th} = \alpha(T - T_{ref}) \tag{4.1}$$

with the material parameter for thermal expansion α and the reference temperature T_{ref}. The reference temperature corresponds to the temperature at which the material is stress-free, thus we assume $T_{ref} = T_{sputter} = 773.15\,\text{K}$ for the NiMnGa/Si bimorphs. The thermal strain value directly refers to the displacements in x-direction sketched in 4.1 a). In the following the thermal strain is taken as a load for the Structural-Mechanics module to determine the resulting deflections in z-direction. By applying Hooke's law, the stress σ is referred to the strain ϵ:

$$\sigma = \sigma_0 + D : (\epsilon - \epsilon_0 - \epsilon_{th}) \tag{4.2}$$

with the initial stress σ_0, the initial strain ϵ_0, and the symmetric 6×6 elasticity matrix D. For an isotropic material the elasticity matrix is reduced to:

$$D = \frac{E}{(1+\nu)(1-2\nu)} \begin{pmatrix} (1-\nu) & \nu & \nu & 0 & 0 & 0 \\ \nu & (1-\nu) & \nu & 0 & 0 & 0 \\ \nu & \nu & (1-\nu) & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{(1-2\nu)}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{(1-2\nu)}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{(1-2\nu)}{2} \end{pmatrix} \tag{4.3}$$

including only the Young's modulus E and the Poisson ratio ν. Also different pairs of elastic moduli can be used, as long as two moduli are defined, the other ones can be calculated. Since all devices are made out of polycrystalline NiMnGa films, the assumption of isotropic material properties is made for all presented numerical calculations.

Summing up, the input parameters for the simulation of the bimorph effect of an isotropic material are the Young's modulus, the Poisson ratio and the thermal expansion coefficient including its reference temperature.

In general, there are two options to get the necessary material parameters: Either scan the literature or measure them directly. Direct measurement has not been possible, since the removal of the substrate to measure single material properties is not easy to manage. Also, the realisation of, e.g. a tensile test for a 200 nm thin film is associated with severe challenges regarding film brittleness and clamping issues. Moreover, the temperature dependence of the individual material properties is of major interest making the direct measurement even more elaborate. Additionally, the material properties are influenced by substrate and film thickness, therefore, the measurement of a freestanding reference thin film might not give the correct parameters. On the other hand, a very broad parameter range is found in literature, as the properties of NiMnGa highly depend on the stoichiometric composition [94]. Besides, a large part of research focuses on single

crystalline NiMnGa, in particular the study of its behavior of different crystallographic orientations, i.e. material anisotropy. Therefore, the Young's modulus has been chosen as a temperature-independent value in the basic model within the reported parameter range [95, 96], while being aware of the resulting uncertainty. The thermal expansion coefficient is known to differ very much between austenitic and martensitic phase, hence, a model for each value is set up. This corresponds to assuming a different slope in temperature regime I, e.g. $T_{sputter} > T > M_S$, compared to temperature regime III, e.g. $M_F > T > RT$, see Figure 4.1 c). Since Si is one of the most investigated materials, its properties are well known being part of the COMSOl library. The initial material parameters for the basic model (model #1) are listed in Table 4.1.

Table 4.1: Material parameters for modelling of the bimorph effect using model #1 as described in the text. The Poisson ratio of Si is included in the COMSOL material library; for the corresponding value of NiMnGa the value of Ni is assumed. The thermal expansion coefficient of Si is used according to the analytical formula presented in [97]: $\alpha_{Si}(T) = (-3.0451 + 0.035705\,T - 7.981\,10^{-5}\,T^2 + 9.5783\,10^{-8}\,T^3 - 5.8919\,10^{-11}\,T^4 + 1.4614\,10^{-14}\,T^5) \cdot 10^{-6}\,1/K$.

	NiMnGa		**Si**
thermal expansion coefficient α, $1/K$	$\alpha_A = 23 \cdot 10^{-6}$ $\alpha_M = 33 \cdot 10^{-6}$	[98, 99]	$\alpha_{Si}(T)$
reference temperature T_{ref}, K	773.15		773.15
Young's modulus E, GPa	100		169 [100]
Poisson ratio ν	0.33		0.22

By using model #1 the behavior of a NiMnGa/Si bimorph during cooling from sputter temperature to room temperature is estimated. The resulting deflection characteristic assuming either α_A or α_M for the modeling cases of 2d and 3d are presented in Figure 4.2. Due to symmetry, only half the geometry of the NiMnGa/Si double beam actuator is considered. In the 3d case, this corresponds to one cantilever with a certain beam width, length and different layer thicknesses of NiMnGa and Si. In the 2d approximation, the width of the double beam is neglected, reducing the geometry to its cross-section. The contact pads of the NiMnGa/Si actuator geometry restrict the bimorph movement. This is approximated by a fixed constraint set to the onset of the bimorph, which is indicated in Figure 4.2 b) by a green surface (3d), and by a green line (2d).

At the sputter deposition temperature, the bimorph is in planar state. Upon cooling, it deflects upward, dependent on the ratio of thermal expansion coefficients of both. Since $\alpha_M > \alpha_A > \alpha_{Si}$ the maximum deflection is obtained by assuming martensite material parameters for the complete temperature range. Besides this observation, it is clearly visible that the influence of the dimension (2d versus 3d) is negligible for the deflection characteristics.

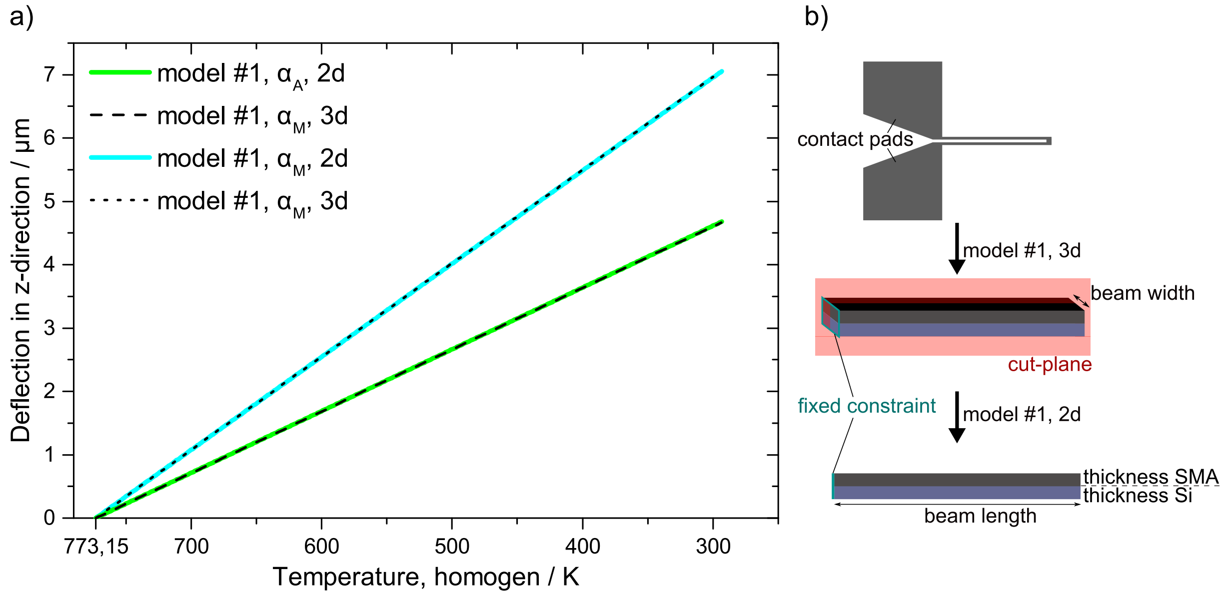

Figure 4.2: a) Simulated deflection characteristics using the basic model in 2d and 3d for a NiMnGa/Si bimorph actuator with the beam width and length of 250 nm, 15.45 µm, respectively. The layer thicknesses of NiMnGa and Si are 200 nm and 250 nm, respectively. b) Illustration of the geometry as it is implemented for the 3d and 2d cases in model #1.

In the temperature regime II in Figure 4.1, $M_S > T > M_F$, NiMnGa transforms from austenite to martensite. Therefore a variable for the martensitic phase fraction $\xi_M(T)$ is defined. Referring to the one-dimensional macro model of Tanaka et al. proposed in 1986 [55], the martensite volume fraction is given by an exponential step-function depending on temperature that varies between 1 (martensitic state) and zero (austenitic phase).

$$\xi_M(T) = \frac{1}{1 + exp(k_M(T - T_M)} \tag{4.4}$$

The parameter k_M is a simulation parameter related to the steepness of the phase transition, whereas T_M is determined by the transition temperatures $T_M = M_S + M_F/2$. With the steepness of the phase transformation, the width of the temperature span for $0 < \xi_M(T) < 1$ is expressed. For a single crystallographic unit cell, there is no meaning of a martensitic phase fraction of, for instance $\xi_M(T) = 0.4$, since the cell can either be 100 % martensitic ($\xi_M(T) = 1$) or 100 % austenitic ($\xi_M(T) = 0$). Besides computational reasons for assuming a step function with no singularity, a non-integer martensitic phase fraction makes sense, if considering not a single unit cell but a representative volume. Here, the martensitic phase fraction presents the averaged phase of a certain volume of SMA, hence the mean value of all unit cells in the considered volume. Moreover, the transition is not abrupt, but starts and ends at two different temperatures. In between, the material has a certain probability to be in martensitic or austenitic state. This probability can also be approximated by a phase fraction between 0 and 1.

By this simple notation it is possible to define each material parameter by a mixture of its value in martensitic and in austenitic state.

For example the thermal expansion coefficient of NiMnGa can be written as a mixture of α_A and α_M:

$$\alpha(T) = (1 - \xi_M(T)) \cdot \alpha_A + \xi_M(T) \cdot \alpha_M \tag{4.5}$$

Although this continuous model does not describe the shape memory effect itself accurately, it enables to implement two independent parameter sets, one for the martensite and one for the austenite. The corresponding model extension, model #2, is able to approximate the ΔP-effect. The influence of a ΔP-effect resulting from different thermal expansion coefficients of martensite, α_M, and austenite, α_A, is shown in the deflection characteristics presented in Figure 4.3.

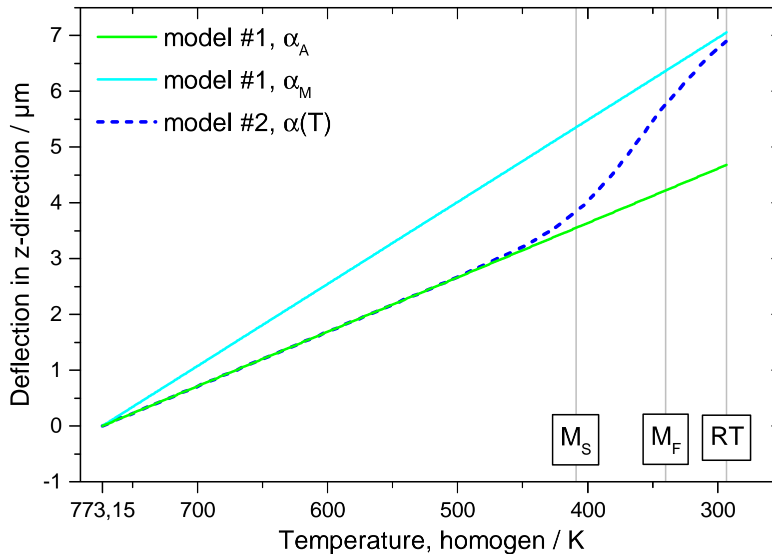

Figure 4.3: 2d-Simulation of deflection characteristics using the basic model (model #1) and its first extension (model #2) for a NiMnGa/Si bimorph actuator with 15.45 µm beam length and layer thicknesses of NiMnGa and Si of 200 nm and 250 nm, respectively.

The simulated actuator deflection exhibits a sharp increase after reaching M_S and then levels off upon cooling below M_F (dashed blue line). The basic model (model #1) using a constant thermal expansion coefficient of either austenite α_A or martensite α_M sets the upper and lower limit for the out-of-plane deflection. The characteristic slope in the deflection behaviour determined by model #2 is due to different thermal expansion coefficients in austenitic and martensitic phase of NiMnGa. The first extension demonstrates the ΔP-effect, hence varying deflection behavior upon cooling. However, model #2 does not include any information about the properties, kinetics or yield conditions of the martensitic transition itself.

In the following model extension, model #3, the Young's Modulus is defined in sections by three strain regimes to evaluate the impact of the characteristic microstructural changes due to the martensitic transformation.

$$\sigma_M(\epsilon) = \begin{cases} \epsilon E1 & \text{if } \epsilon < \epsilon_{yield} = 0.001 \\ \epsilon E2 & \text{if } \epsilon_{yield} \leq \epsilon \leq \epsilon_{crit} \\ \epsilon E3 & \text{if } \epsilon > \epsilon_{crit} = 0.06 \end{cases} \tag{4.6}$$

In order to keep the model as simple as possible, the same Young's modulus is used for the initial elastic regimes in austenite and in martensite, e.g. $E_A = E1 = E3$. In martensite state, in the strain range between ϵ_{yield} and ϵ_{crit}, the quasi-plastic stress-strain behavior is described by a low stiffness parameter $E2$ being one order smaller than the stiffness in the elastic regime. This enables thermal stress reductions once the temperature reaches $T < M_S$. The stress relaxation process in turn reduces the deflection change caused by the bimorph-effect.

According to the Voigt model, parallel coupling of both phases is assumed in the transformation regime. Thus, the stress of the martensite-austenite compound is described by the rule of mixture:

$$\sigma(\epsilon, T) = (1 - \xi_M(T)) \cdot \sigma_A(\epsilon) + \xi_M(T) \cdot \sigma_M(\epsilon), \text{ with } \sigma_A(\epsilon) = \epsilon E1. \tag{4.7}$$

The influence of the two extensions on the basic model is shown in the deflection characteristics presented in Figure 4.4.

The simulated deflection based on model #3 follows the behavior of model #1 using α_A in the temperature range between 500 °C and martensite start temperature M_S. Below M_S the deflection approaches the behavior of model #2 upon a certain point, when the martensitic transformation enables stress relaxation, which leads to reduced deflection upon cooling. The resulting deflection versus temperature curve provides a peak deflection value, which is reached between M_S and RT. This characteristic temperature is denoted as the peak temperature T_P in the following and indicated by an arrow in Figure 4.4.

Since the stress relaxation is dependent on the strain in the NiMnGa layer, which is in turn dependent on the thickness ratio of NiMnGa to Si, the influence of stress relaxation due to martensitic transformation on deflection is also layer ratio dependent. In Figure 4.5 the temperature-dependent deflections calculated by model #3 are shown for five different SMA film thicknesses, while the Si film thickness is kept constant. For $t_{SMA} > 300$ nm, the deflection is dominated by the bimorph- and ΔP-effect. For SMA/Si layer thickness ratios larger than one, the average strain is below the lower limit ϵ_{yield} for onset of quasi-plastic deformation defined in Eq. 4.6, hence, no stress relaxation occurs. For $t_{SMA} < 300$ nm, the average strain increases, thus, stress release gains significance. Therefore, a change of deflection due to martensitic transition is observed, resulting in a peak in the deflection characteristic. Again, the corresponding peak temperatures T_P are indicated by arrows in Figure 4.5, which in parallel visualizes the thickness dependence of T_P.

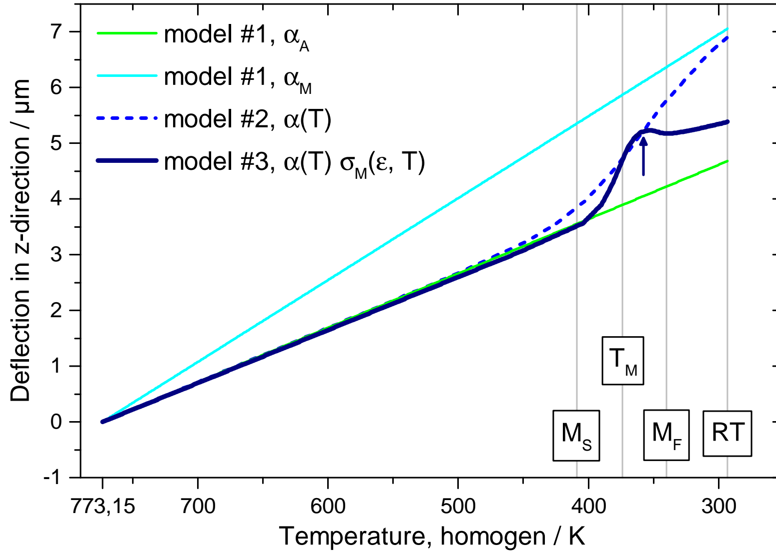

Figure 4.4: 2d-Simulation of deflection characteristics using models #1, #2 and #3 for a NiMnGa/Si bimorph actuator with 15.45 µm beam length and layer thicknesses of NiMnGa and Si of 200 nm and 250 nm, respectively. Model #3 reveals a change of deflection due to martensitic transition. The characteristic peak temperature T_P is indicated by an arrow. Figure reproduced in accordance to [93].

From these observations it is clear, that the relative contributions of the different effects depend on the temperature spans and the chosen bilayer thickness ratio. Hence, the maximum actuation stroke depends on the chosen temperature span and thickness ratio. Generally it can be either negative or positive. Therefore, the simulation shown in Figure 4.5 is repeated for NiMnGa film thicknesses between 25 and 400 nm in 5 nm steps to evaluate the ideal thickness ratio for maximum contribution of SME versus maximum contribution of ΔP effect. Three temperature spans are taken into account to directly compare the trend of certain actuation strokes. The temperature span of complete transformation, $RT < T < M_S$, and a smaller temperature span between RT and T_P are chosen. Additionally, the actuation stroke is determined for the temperature span between RT and T_M. The mean transition temperature, $T_M = 1/2 \cdot (M_S + M_F)$, is also considered here, since T_P varies with film thickness. A fixed temperature span provides better comparability and simplifies the control of the actuator in a future application. For the minimal NiMnGa film thicknesses T_M approaches T_P.

Figure 4.6 shows the simulated actuation stroke of NiMnGa/Si bimorph actuators as a function of SMA film thicknesses for different temperature spans upon heating. Two thickness regimes can be distinguished reflecting the counteraction of ΔP effect and SME. For SMA film thicknesses above 200 nm, the stroke is positive not depending on the temperature span. In this case, the ΔP effect dominates. Consequently, by heating the bimorph above room temperature, the deflection will decrease. For SMA film thicknesses below 200 nm, however, the actuation stroke becomes negative, which refers to a lower beam deflection at room temperature compared to the given deflections at temperatures $M_{S/F}$. In this case, the stress relaxation associated with the martensitic transformation

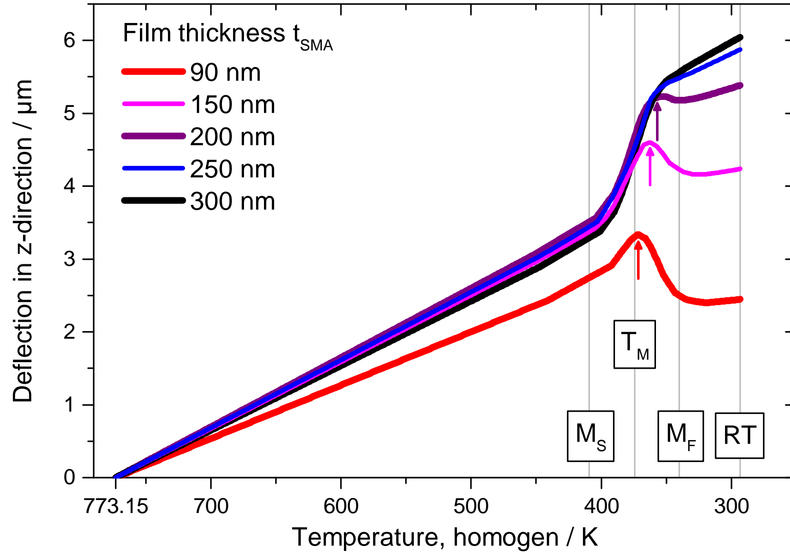

Figure 4.5: 2d-Simulation of deflection characteristics using model #3 for NiMnGa/Si bimorph actuators with 15.45 µm beam length and layer thicknesses of NiMnGa, t_{SMA}, as indicated. The Si layer thickness is 250 nm. The characteristic peak temperatures T_P are indicated by arrows. Figure reproduced in accordance to [93].

is clearly observed. The corresponding optimum NiMnGa thickness revealing maximum negative stroke is about 90 nm. The corresponding maximum negative actuation stroke is nearly 900 nm (approximately 6 % of beam length) for the temperature span between RT and peak temperature T_P.

Besides the critical strain values ϵ_{yield} and ϵ_{crit}, the influence of the substrate has to be considered. At the interface of NiMnGa film and Si substrate, the stress value is increasing with increasing out-of-plane deflection. Upon a certain stress level, plastic deformation can occur, which will lead to a reduction of NiMnGa material being able to contribute to the SME. However regions below a certain minimal stress level may not exceed the limit for quasi-plastic deformation at all. This is the case at the NiMnGa surface, which thus also does not contribute to the SME. The basic model #1 is used to model the stress profiles at different temperatures to estimate reasonable upper and lower stress limits. In Figure 4.7, the stress profile along the thickness direction of a NiMnGa/Si bimorph is shown. As expected, maximum stress values are observed at the NiMnGa/Si interface. Due to the higher thermal expansion coefficient of martensitic NiMnGa, the corresponding stress values exceed the stress values for the austenitic case, which is indicated by the solid (model #1, α_A) and dashed (model #1, α_M) red line in Figure 4.7.

For onset of quasi-plastic deformation, a stress limit of $\sigma_{c1} = 100$ MPa is assumed. The lower stress limit corresponds to a strain value of 0.1 % due to the Young's modulus of 100 GPa, which has been already introduced in Eg. 4.6. A strain value of 0.1 % for onset of quasi-plastic deformation lies within the data reported in literature. The upper limit, $\sigma_{c2} = 500$ MPa, most likely refers to an underestimation of the fraction of noncontributing

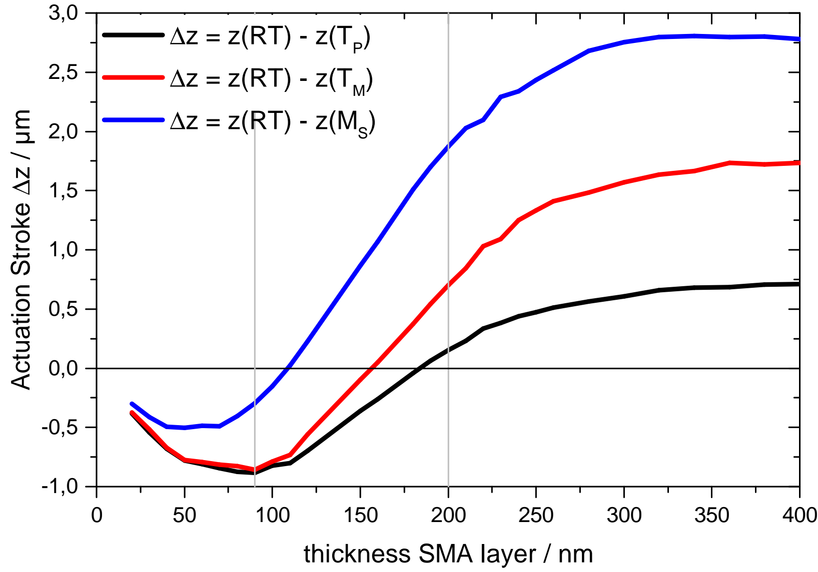

Figure 4.6: Simulated actuation stroke as function of SMA film thickness for different temperature spans upon heating at room temperature. The deflection values are taken from Figure 4.5. The Si layer thickness is 250 nm. Figure reproduced in accordance to [93].

material, since additional effects, such as creep occurring at high deposition temperatures may also induce stress relaxation [101, 102]. In Figure 4.7, the fraction of NiMnGa material in the range $\sigma_{c1} < \sigma < \sigma_{c2}$ is highlighted in light blue. The fraction of NiMnGa exceeding σ_{c2} is about 4 % at room temperature (highlighted in red).

Taking these limits into account, we estimate the fraction of SMA material contributing to the SME. For this purpose the NiMnGa film thickness is varied from 100 to 800 nm for a fixed layer thickness of Si of 250 nm. Figure 4.8 shows the volume fraction of SMA material exceeding σ_{c1}, and the fraction exceeding σ_{c2}. An exponential fit is used to approximate the simulation data. For NiMnGA film thicknesses larger than 300 nm, material fractions above the lower and upper stress limit remain almost unchanged. In this case, the fraction of SMA material exceeding the lower stress limit is about 42 %, while the fraction exceeding the upper limit is about 3.2 %. Below 100 nm NiMnGa film thickness the fraction of material exceeding the lower limit reaches 100 % whereas the fraction exceeding σ_{c2} approaches 100 % while approaching zero film thickness. As a consequence the fraction of NiMnGa contributing to the SME reaches a maximum of nearly 80 % for a SMA film thickness of about 90 nm.

Summing up the FEM simulation reveals the counteraction of ΔP effect and stress relaxation. Referring to the study on the layer ratio dependence of ΔP and SME effect, we observe maximal stress relaxation, thus maximum negative stroke, associated with martensitic transformation for the same SMA/Si thickness of 90 nm / 250 nm, which also maximizes the estimated fraction contributing to the SME. Furthermore, an ideal thickness ratio SMA/Si of $1 / 2.\overline{7}$ is predicted, which should provide a maximal stress relaxation effect. To validate this predictions, the deflection of single bimorph double

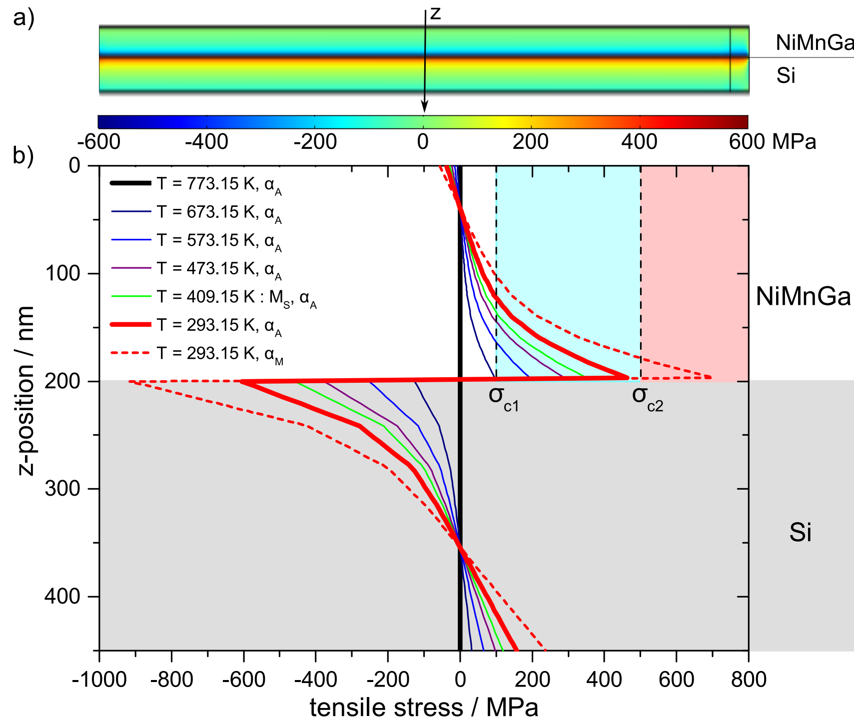

Figure 4.7: Simulated tensile stress of a NiMnGa/Si bimorph actuator with 250 nm beam width and 15.45 µm beam length. The layer thicknesses of NiMnGa and Si are 200 nm and 250 nm. a) Stress distribution of the cross-section at room temperature for model #1 assuming α_A (solid) and α_M (dashed). b) Stress profile along z-direction of the cut-line indicated in a). Figure reproduced in accordance to [93].

beam structures has to be measured, while applying homogeneous heating or cooling. Implementing this condition is a challenge for nanoactuation. Homogeneous heating implies that the whole substrate is heated, causing thermal expansion effects of the whole chip. The resulting movements are in the order of hundreds of µm, hence bigger than the field of vision in the SEM that is needed to resolve the deflection of single bimorph nanoactuators. An alternative to homogeneous heating of the whole chip but only the single investigated structure, is to apply Joule heating. This alternative allows for selective heating, and thus precise deflection measurements. On the other hand, Joule heating is associated with a temperature gradient along the beam. Since each beam is connected to the contact pad which performs like a cold temperature reservoir, the onset of the beam stays quite cold compared to the beam tip. As a consequence, the phase fraction of martensite varies along the beam, e.g. some sections of the beam may have already transformed to austenite, while other sections are still in martensitic state. Therefore, any measurement of any variable which depends on temperature or phase fraction will measure a superposition of different phase states occurring along the beam. Moreover, the temperature is not a parameter which is set, but a parameter which is influenced by the thermal conductivity and the electrical power. Hence, the assignment of electrical power to temperature is highly dependent on material parameters which makes it difficult to refer individual characteristics of an observable to a certain temperature.

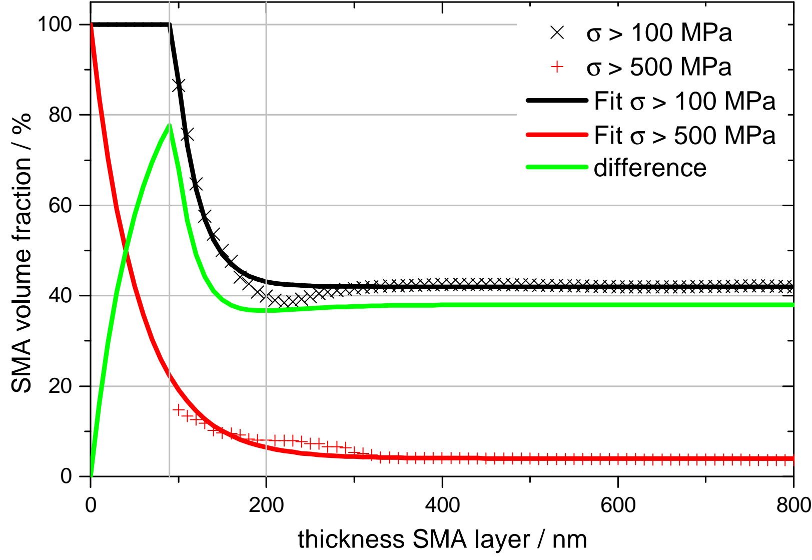

Figure 4.8: Simulated volume fraction of NiMnGa material exceeding the assumed limit for quasi-plastic deformation of 100 MPa (black) and above the assumed limit for onset of plastic deformation of 500 MPa (red) as a function of SMA film thickness. Solid lines are fit curves to the FEM simulation data plotted as symbols. The volume fraction in the intermediate stress range (green line) represents the SMA material contributing to the SME. The Si layer thickness is 250 nm. Figure reproduced in accordance to [93].

Summing up, the SME effect is smeared out and the electrical power is controlled instead of the temperature. In parallel, the Joule heating alternative directly gives rise to monitor the electrical resistance versus electrical power. In the following section, the underlying concept of the simulation of electrical resistance during Joule heating of single bimorph structures is introduced.

4.2 Electromagnetics

The Heat-Transfer module of COMSOL Multiphysics and the Structural-Mechanics module are combined with the AC/DC module to approximate electrical characteristics of NiMnGa/Si bimorph double beams upon Joule heating. The AC/DC module is used to solve Maxwell's equations (Eq. 4.8) by the finite element method. Numerically stable edge element discretization is applied to approximate stationary and dynamic electric $\vec{E}$ and magnetic $\vec{B}$ fields.

$$\begin{aligned} \vec{\nabla} \cdot \vec{D} &= \rho \\ \vec{\nabla} \cdot \vec{B} &= 0 \\ \vec{\nabla} \times \vec{E} &= -\frac{\partial \vec{B}}{\partial t} \\ \vec{\nabla} \times \vec{H} &= \vec{j} + \frac{\partial \vec{D}}{\partial t} \end{aligned} \tag{4.8}$$

In a linear, homogeneous and isotropic material the displacement field $\vec{D}$ and the magnetizing field $\vec{H}$ can be approximated by $\vec{D} = \epsilon \vec{E}$ and $\vec{H} = \frac{1}{\mu}\vec{B}$, using constant material parameters for permittivity ϵ and permeability μ. The electric charge density ρ and the electric current density $\vec{j}$ are introduced as sources of the electromagnetic field. Moreover, the electric current density can be expressed by the electrical conductivity σ by $\vec{j} = \sigma \vec{E}$. Throughout this work, magnetic properties of the ferromagnetic SMA NiMnGa are not investigated neither numerically nor experimentally.

The variable for the martensitic phase fraction $\xi_M(T)$ introduced in the description of model #2 in the previous section, is used to define physical properties in the phase transformation regime by a mixture of its value in martensitic and in austenitic state, according to Eq. 4.5. In addition, a second variable is defined similar to $\xi_M(T)$, that describes the transformation from martensite to austenite, denoted as $\xi_A(T)$. As a consequence, separate transformation temperatures can be included for forward and reverse transformation. The two variables are connected by a simple if-then-else query depending on a certain time t_c. In a full phase transformation cycle, t_c is the time after which the specimen has completely transformed in forward direction into a single phase state and before reverse transformation starts.

$$\xi(T) = \begin{cases} \xi_A(T) & \text{if } t \leq t_c \quad \text{martensite} \rightarrow \text{austenite} \\ \xi_M(T) & \text{if } t > t_c \quad \text{austenite} \rightarrow \text{martensite} \end{cases} \tag{4.9}$$

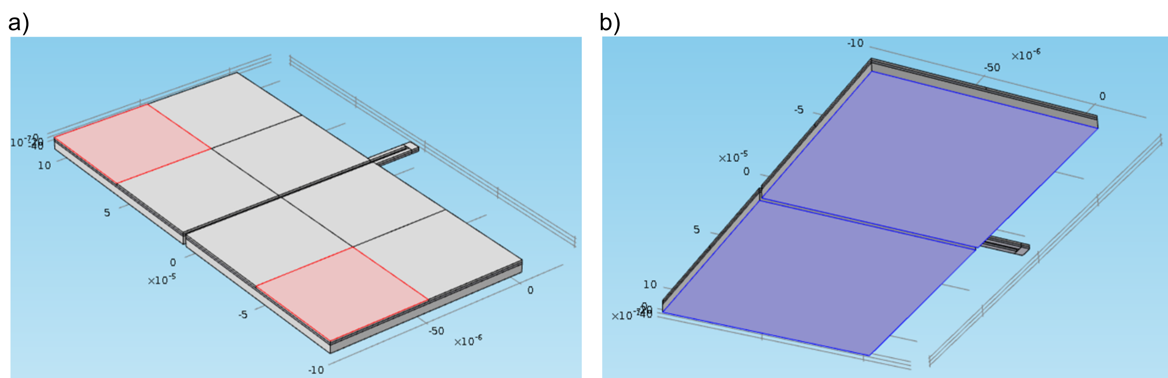

Figure 4.9: Boundary conditions for the Joule heating FEM model. a) Electrical contacts (red) serving as the input and output port for the electrical current. b) Heat sink and fixed constraint at the bottom of the SiO_2 layer (blue).

Certain boundary conditions have to be chosen to model Joule heating instead of homogeneous heating. Since the pressure in the SEM is of the order of 10^{-6}mbar convective cooling is negligible. Cooling of the bimorph is dominated by conductive cooling due to the connection of the *hot* double beam to the substrate holder acting as an infinite cold temperature reservoir. The temperature of the heat sink is set to room temperature. There are several options to model the heat sink, the most obvious choice would be to position the heat sink at the bottom of the Si handle wafer. Indeed the Si handle wafer has a thickness of 725 µm, which goes beyond the computational limits. Therefore, the Si handle layer thickness is set to 4 µm. The resulting temperature profile is compared to

approximations further reducing the model geometry. First, the heat sink is positioned directly at the bottom of the Si device layer and second, it is positioned at the bottom of the 2 µm thick SiO_2 layer. Whereas the first approximation results in unrealistic temperature profiles, the second is sufficient. Hence, the Joule heating model geometry considers the NiMnGa layer, the Si layer and the SiO_2 layer. In Figure 4.9 b) the heat sink is highlighted in blue. As a mechanical boundary condition, a fixed constraint is applied to the bottom layer of the oxide layer, thus overlapping with the heat sink. Besides, the thermal expansion reference temperatures is set to 773.15 K for NiMnGa and Si to generate an upwards deflected structure as the initial state at room temperature. The mechanical behaviour is described without taking stress relaxation by the martensitic transformation into account to further reduce the computational effort. Furthermore an electrical current has to be applied to the double beam structure to enable Joule heating. Therefore, the electrical potential is set to zero at all outer boundaries besides two finite areas located at the contact pad which serve as the input and output port for the electrical current. The two ports are highlighted in red in the schematic shown in Figure 4.9 a). Not the whole contact pad is chosen as the electrical port to exclude thermal effects at the onset of the beam.

The necessary material parameters for the AC/DC module and the Heat Transfer module for quasi static conditions are the permittivity ϵ, the electrical conductivity σ and the thermal conductivity κ. As introduced before, the thermal expansion coefficient α, the Poisson ratio ν and the Young's modulus E set the mechanical response. Table 4.2 lists the material parameters chosen for the Joule heating simulation.

Table 4.2: Material parameters for modelling the electrical, thermal and mechanical behaviour upon Joule heating.

	NiMnGa	**Si**	**SiO_2**
electrical conductivity σ, S/m	$\sigma_{SMA}(T)$	10	0
thermal conductivity κ, W/mK	$\kappa_{SMA}(T)$	$\kappa_{Si}(T)$	1.4
permittivity ϵ	4.2	4.2	4.2
Poisson ratio ν	0.33	0.22	0.17
Young's modulus E, GPa	100	169	70
thermal expansion coefficient α, $1/K$	$\alpha_{SMA}(T)$	$\alpha_{Si}(T)$	$0.5 \cdot 10^{-6}$

The temperature dependence of the thermal expansion coefficient of NiMnGa and Si is considered according to the previous section. The electrical conductivities of Si and SiO_2 are given by the SOI wafer supplier, whereas the other materials constants of Si and SiO_2 are taken from literature and included in the COMSOL material bibliography.

First, the temperature dependence of the electrical conductivity of NiMnGa is considered. Therefore, the temperature-dependent resistance of a thin NiMnGa film is taken into account. A representative curve is shown in Figure 4.10. The characteristic course is used to determine the transformation temperatures, but also the temperature-dependent

resistance in austenite and martensite can be identified. Therefore, straight lines are fitted to the electrical resistance characeristics in the temperature range $RT < T < A_S$ and $A_F < T < T_{max}$. The first refers to the martensitic resistance characteristic $R_M(T)$ whereas the latter refers to the austenitic resistance characteristic $R_A(T)$.

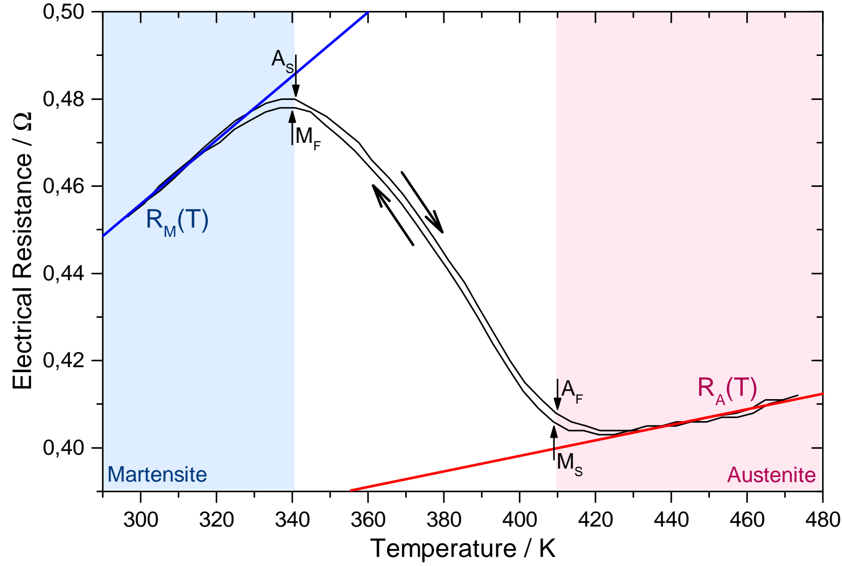

Figure 4.10: Temperature-dependent electrical resistance characteristic of a NiMnGa film with the nominal layer thickness of 1 µm. The austenitic phase and the martensitic phase are highlighted and the corresponding transformation temperatures are indicated. The resistance in martensite, $R_M(T)$, and austenite, $R_A(T)$, are determined by a standard line fit procedure.

The relation between electrical conductivity and the introduced functions $R_{M,A}(T)$ is given by Eq. 4.10 with the thickness of the SMA layer d.

$$\sigma_{M,A}(T) = \frac{1}{R_{M,A}(T)} \cdot \frac{1}{d} \tag{4.10}$$

Finally the electrical conductivity is given by a mixture of the conductivity in austenitic and martensitic state by the martensitic phase fraction.

$$\sigma_{SMA}(T) = (1 - \xi(T)) \cdot \sigma_A(T) + \xi(T) \cdot \sigma_M(T) \tag{4.11}$$

The thermal conductivity of the NiMnGa layer is approximated by the Wiedemann-Franz Law, describing the thermal conductivity of metals by the electrical conductivity. This assumption is based on the approach that the amount and velocity of free electrons dominate the heat transport and the electrical current in metals.

$$\kappa_{SMA}(T) = L_o \cdot \sigma_{SMA}(T) \cdot T \tag{4.12}$$

The constant proportionality factor L_o was found empirically by Ludvig Lorenz, but later proved quantum mechanically: $L_o = \frac{\pi^2 \kappa_B^2}{3e^2} = 2.45 \cdot 10^{-8} \frac{W\Omega}{K}$.

The thermal conductivity of Si has been measured by C.J. Glasbrenner and G. A. Slack in the range from 3 K to the melting point of 1681 K [103]. Figure 4.11 shows its characteristic in the temperature span between room temperature and about 1200 °C revealing a significant decrease at higher temperatures until it saturates. In previous work of C. Lay, a size effect in the Si layer has been identified by comparing Ni/Si double beam structures of different lateral dimension with each other [15]. He postulated a thermal conductivity of bulk Si of $\kappa_{Si}(T)^* = 5.016 \cdot 10^4/\mathrm{T} - 19.29$, which has to be adapted by a geometry-dependent form factor F by $\kappa_{Si}^{nano}(T)^* = F \cdot \kappa_{Si}(T)^*$. Taking this knowledge into account, a bulk thermal conductivity is defined, which is adaptable by a form factor to approximate the thermal conductivity of nanostructures. To avoid the thermal conductivity to become negative, an exponential approximation to the empirical data is chosen in contrast to $\kappa_{Si}(T)^*$. The explicit equations are shown in Figure 4.11 as well as an example for the impact of a form factor equal to $F = 0.4$. A form factor of $F = 0.4$ reduces the thermal conductivity to 40 % of its bulk value for the whole temperature range. Because of the exponential temperature dependency of κ_{Si}, the effect is more pronounced for lower temperatures and nearly vanishes for high temperatures. For instance, the difference in thermal conductivity of bulk Si and $\kappa_{Si}^{nano}(T)$, assuming a form factor of $F = 0.4$, is roughly 100 W/mK at room temperature and approximately 10 W/mK at 1000 °C.

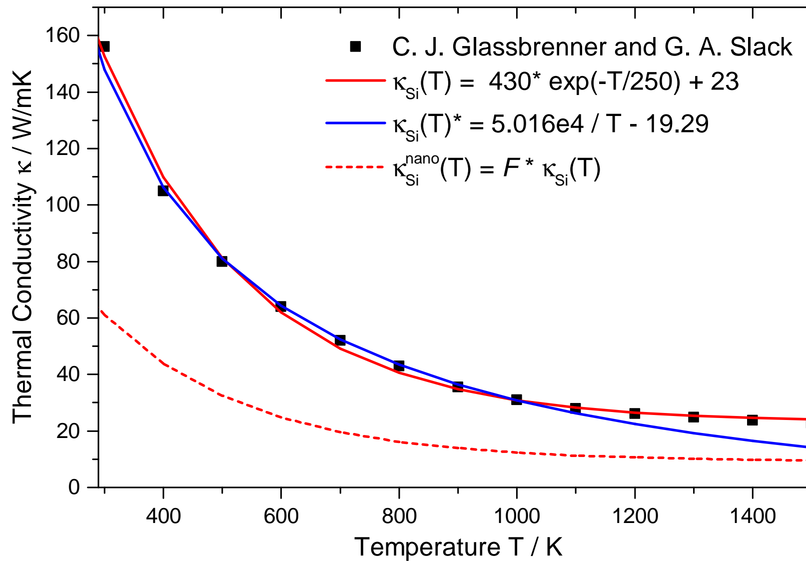

Figure 4.11: Thermal conductivity as a function of temperature according to [103]. The dashed lines illustrate the influence of an additional form factor $F = 0.4$ as described in the text.

At this point all necessary boundary conditions are set and the properties of participating materials are defined. A time dependent study is used to ramp the current up to its pre-defined maximum value and back to zero after a certain dwell time. In Figure 4.12, the initial stress profile at room temperature (a) and the slope of the applied current versus time (b) are shown. In b) also t_c is indicated, where the simulations switches from

$\xi_A(T)$ to $\xi_M(T)$. In Figure 4.12 c) the simulated time-dependent electrical resistance depending is illustrated. In Figure 4.12 d) surface plots of the double beam structure for the times A, B, C, and D are shown, as indicated in Figure 4.12 c). The color mapping in the NiMnGa layer refers to the martensitic phase fraction, with blue indicating martensitic phase ($\xi(T) = 1$) and red indicating austenitic phase ($\xi(T) = 0$). As it is clearly visible, the double beam deflects downward upon heating and transforms into austenite beginning from the beam's tip, while the beam's onset stays in martensitic state. Hence, qualitatively the characteristics of the deflection and resistance behaviour upon Joule heating are represented by the FEM model.

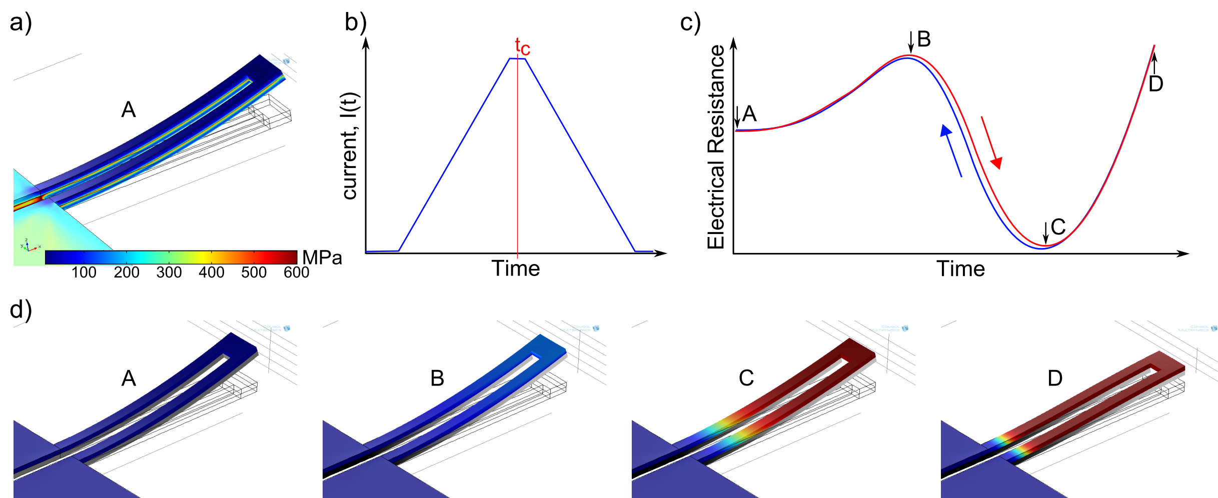

Figure 4.12: a) Initial stress profile and corresponding deflection at room temperature; b) current ramp to perform Joule heating simulation; c) simulated electrical resistance characteristic; d) surface plots that show the martensitic phase fraction at the time steps indicated by the points A - D. Color code: Red (austenite), blue (martensite).

In the following the influence of the input parameters is studied to investigate the potential of quantitative analysis of simulation results with experimental data.

The presented model provides quite some flexibility in the parameter choice. On the one hand, there are parameters with no direct physical meaning included in the definition of $\xi(T)$, and on the other hand, the range is very broad for the material parameters. A common procedure to set these parameters is to use verified experimental data to specify the simulation parameters. A set of verified experimental data in this case are measurements on double beam structures with micrometer dimensions, which are known to show no explicit size-effect. Due to the measurement inside the SEM the number of observables is quite limited. More precisely, experimentally accessible are only the electrical resistance and the deflection versus applied electrical power. Especially the relation between electrical power and temperature distribution of the double beam structure is unknown in this case. A double beam structure is modeled with a certain set of parameters, which are in the following varied successively to illustrate the dependency of some key parameters on the simulation result. A list of the parameters in the basic model and in the individual variations is shown in Table 4.3.

Table 4.3: Material parameters for modelling the electrical, thermal and mechanical behaviour under Joule heating.

	basic	**var1**	**var2**	**var3**	**var4**	**var5**	**var6**	**var7**	**var8**	**var9**
l, µm	24	34								
w, µm	2		1							
t_{Si}, nm	1000			500						
t_{SMA}, nm	1000				500					
d, nm	645					545				
F	1						0.4			
E_M, GPa	100							50		
α_M, 10^{-6} 1/K	33								23	
A_S / A_F, K	338.15/390.15									328.15/380.15

In this work, the overall objective of the simulations is to model the electrical resistance and deflection of micrometer sized NiMnGa/Si bimorph double beam structures with high precision and to compare corresponding experimental data of nanostructured bimorphs with this model to identify size effects. A major challenge is to unambiguously assign an effect to the right cause. Therefore, the first parameter variation focuses on the effect of scaling, e.g. the influence of changed geometry settings. In Figure 4.13, the influences of smaller beam width w, larger beam length l, smaller Si and SMA layer thicknesses t_{Si} and t_{SMA} are shown. In addition, the variation of thickness parameter d is shown, corresponding to the thickness of the NiMnGa reference film to determine the electrical conductivity.

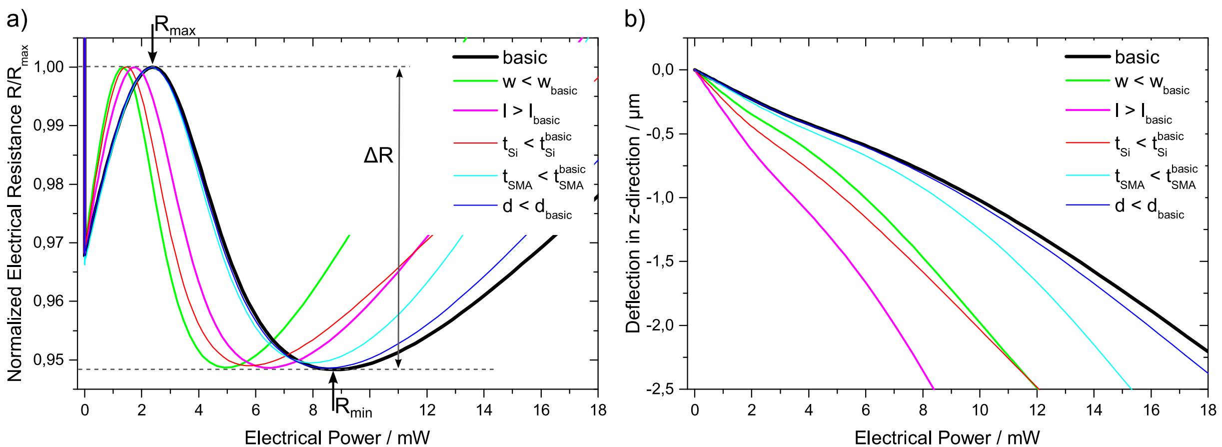

Figure 4.13: a) Normalized electrical resistance characteristics and b) deflection in out-of-plane direction for the basic model and geometry variations as indicated.

By increasing the length of a double beam structure, by decreasing its width or thickness the absolute resistance value increases. Therefore, in Figure 4.13 a) the normalized

electrical resistance is shown, to be able to present all geometry variations in one graph. Besides, any measured data will have some fluctuations in the absolute electrical resistance value, which requires some kind of normalization anyway. It can be recognized that the relative change of resistance due to the phase transformation, given by $\Delta R = R_{max} - R_{min}$, stays constant for all geometry variations. In contrast, the resistance maximum is shifted to lower power values for increasing double beam length, decreasing double beam width and decreasing Si layer thickness. This effect is caused by a change in the temperature distribution in the double beam. By reducing the width of the cantilevers, the heat transfer is reduced resulting in an higher temperature for the same electrical power value. Thus, the transformation start temperatures are reached earlier, in terms of electrical heating power. The approximated thermal conductivity of NiMnGa is much higher than the thermal conductivity of Si and does not decrease with temperature. Thus, a reduced SMA layer thickness has only a marginal effect on the resistance characteristics.

In Figure 4.13 b) the deflection characteristics for the indicated geometry variations is shown. The changed temperature profiles lead to larger deflections for all geometry variations. This effect is enhanced due to higher mechanical flexibility of thinner and longer structures.

With regard to the identification of size effects it is of highest interest to know the dimensions of the investigated double beams very accurately. As described in detail in Chapter 2, the sputter deposition of NiMnGa causes side-wall deposition. Since the temperature profile is dominated by the properties and dimensions of the Si structure, it can be assumed that the side-wall deposition does not influence the resistance characteristics significantly.

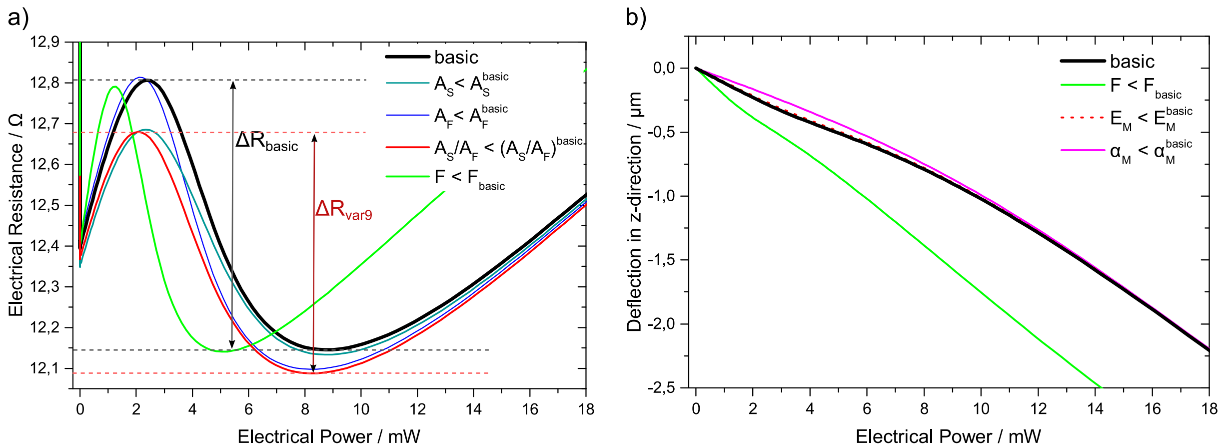

Figure 4.14: a) Normalized electrical resistance characteristics and b) deflection in out-of-plane direction for the basic model and variations of the transition temperatures and material properties as indicated.

In the next parameter study, the influence of varying material properties is investigated. As discussed before, the parameter range for NiMnGa is quite broad in literature and direct measurement was not possible. Therefore, the influence of a smaller Young's modulus and a smaller thermal expansion coefficient in the martensitic state is investigated. Both parameters solely affect the mechanical response of the system,

therefore, no effect is present in the electrical resistance characteristics. As shown in Figure 4.14 b), a decrease of α_M from $33 \cdot 10^{-6}$ 1/K to $23 \cdot 10^{-6}$ 1/K (var #8), which is equal to α_A, changes the slope of the deflection characteristics, but has negligible influence on the absolute deflection value. It can be hardly seen, but even a reduction of the martensitic Young's modulus from 100 GPa to 50 GPa (var #7) does not significantly influence the deflection behavior.

Furthermore, the influence of shifted transition temperatures is examined. Even though there is no crystallographic reason for an individual change of A_S and A_F, the simulation was not only performed assuming a common shift, but also a separate shift with A_S / A_F staying fixed. In Figure 4.14 a), the electrical resistance characteristics for the transition temperature variation is shown. A shift of the austenite start temperature to lower values results obviously in a shift of R_{max} to lower power values. Associated with this shift is a smaller absolute maximum electrical resistance value R_{max}. Accordingly, a shift towards lower austenite finish temperatures results in a shift of R_{min} to lower power values and lower resistance values. The combination of both variations, hence a joint shift of A_S and A_F by 10 K each (var #9), results in shift of the electrical resistance characteristics towards lower power values and a decrease of ΔR, as indicated by ΔR_{var9} in red in Figure 4.14 a).

At last, the influence of the Si material parameters on the resistance and deflection characteristics is examined. Since Si is one of the most investigated materials nowadays, we assume the Young's modulus and the thermal expansion coefficient to be known. Since the electrical conductivity highly depends on the dopand concentration, the manufacturer provides a range of values. However, since even the maximum value given is several orders of magnitude smaller that the electrical conductivity of NiMnGa, the influence of the Si electrical conductivity is negligible in this material system. In comparison, the Si layer has a large impact on the heat transfer, resulting in a shift towards lower power values for decreasing Si layer thickness (var #3), as illustrated in Figure 4.13. By introducing a form factor F (var #6), the thermal conductivity of Si is reduced. The reduced Si layer thickness results in a changed temperature gradient along the beam, hence a shift of R_{max} and R_{min} towards lower power values and larger deflection values. Both effects are shown in Figure 4.14 for a form factor of $F = 0.4$.

Summing up, several parameters can cause similar changes, which makes it difficult to directly link a measured characteristic to a certain material parameter. To further visualize this statement, the effect of the presented parameter variation on the electrical resistance characteristics is summarized in Table 4.4. The sign "-" marks a negligible effect size compared to the impact of the other variations.

In particular, a shift of the electrical resistance maximum can have several causes. Therefore, a major challenge to improve the simulation model is to improve the available experimental data. For the introduced system, the temperature dominates the key characteristics but the temperature profile cannot be measured in-situ. The link between electrical heating power and generated heat distribution can be modeled, but the result highly depends on material parameters, which cause quite big systematic errors, as not only NiMnGa, but also Si in nanometer dimension contributes some uncertainty.

Table 4.4: List of the effects of the parameter variations on the electrical resistance characteristics.

var #	description	effect on resistance characteristics
1	$l > l_{basic}$	R_{max}, R_{max} shifted to lower power values
2	$w < w_{basic}$	"
3	$t_{Si} < t_{Si}^{basic}$	"
4	$t_{SMA} < t_{SMA}^{basic}$	-
5	$d < d_{basic}$	-
6	$F < F_{basic}$	R_{max}, R_{max} shifted to lower power values
7	$E_M < E_M^{basic}$	-
8	$\alpha_M < \alpha_M^{basic}$	-
9	$A_S \,/\, A_F < (A_S \,/\, A_F)^{basic}$	small shift R_{max}, R_{max} to lower power values & smaller ΔR

Therefore, the development of a measurement setup which is capable to monitor the temperature of the double beams is essential. In Chapter 3 the measurement setup for Joule heating is introduced and the improvements of the new setup for homogeneous heating are discussed in detail.

4.3 Wave Optics

The performance of a SMA-based optical switch is approximated by a FEM model using the Wave Optics module provided by COMSOL Multiphysics. The Wave Optics module calculates the wave propagation by direct discretization of Maxwell's equations (Eq. 4.8).The electric field is expressed as the product of a slowly varying envelope function and a rapidly varying exponential phase function to minimize the computational effort. Additionally, the modelled geometry is reduced to the interaction zone of the optical switch, as illustrated in Figure 4.15.

First, the local eigenmodes of the facets of the waveguide geometry are determined. The eigenmodes represent the natural transverse resonances and offer a means for low loss transmission of the electromagnetic wave in propagation direction. The eigenmodes, and in general the transmission through a waveguide structure, depend on the wavelength of the incident light (here: 1550 nm), the refractive indices of the waveguide and the surrounding, which are 3.48 for Si and 1.444 for SiO_2, respectively, and the geometry settings. As a starting point, the waveguide width is chosen to be 200 nm and the length of the waveguide is set to be 5 µm. Figure 4.16 shows the results of the initial boundary mode analysis. Due to the tapering of the rectangular waveguide to a waveguide width of $w = 200$ nm, the waveguide becomes single mode and solely the fundamental TEM_{00} mode

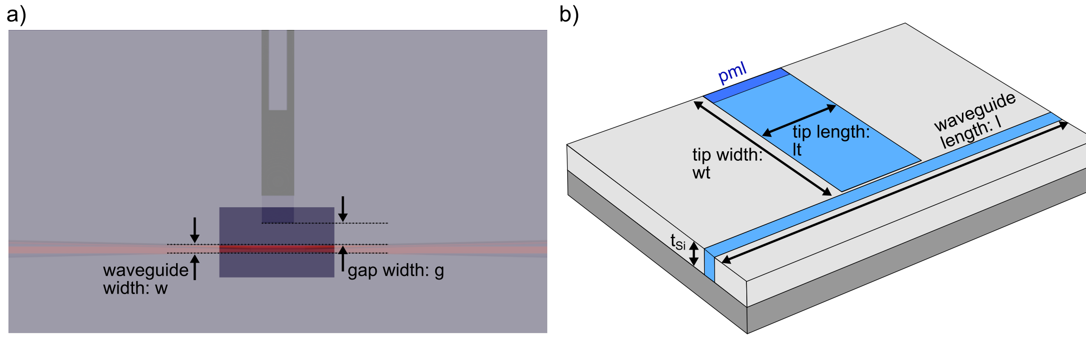

Figure 4.15: a) Schematic layout of SMA-based NEMS actuator and tapered Si waveguide. b) Simplified geometry used for the simulation of propagating modes including key geometry parameters.

is guided. The fundamental mode has a non-zero out-of-plane electric field component E_z and a non zero in-plane magnetic field component H_y, which is commonly denoted as a TE mode for dielectric slab waveguides. The corresponding TM mode having a 90 ° rotation of the two transverse modes, E_y and H_z, is not favorable. This is due to tapering the waveguide only in one direction, while keeping the thickness, e.g. height of the waveguide fixed to $t_{Si} = 340\,\mathrm{nm}$ leading to a larger mode diameter in the z-direction. As visualized in 4.16 a) a discontinuity arises along the z-direction in the transverse electric field. Since the dielectric constants or the permittivity of the lower cladding, e.g. the SiO_2 layer, the upper cladding, air, and the Si waveguide are different. Due to equal permeability constants no step arises in the transverse magnetic field, see Figure 4.16 b).

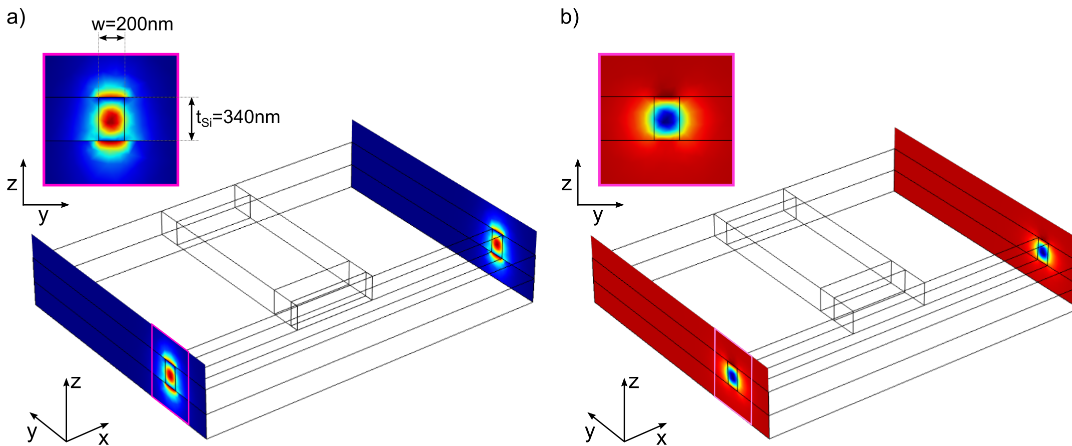

Figure 4.16: a) Tangential electric mode field E_z and b) tangential magnetic mode field H_y for an effective mode index of 2.0663. Due to the asymmetric tapering only the TE component of the fundamental TEM_{00} mode is guided.

In the following, the facets of the waveguide structure are defined as an input and output port to evaluate the transmission. In these ports, the S-parameters can be directly extracted from the simulation result. In the S-parameter approach, in general, an electrical network is regarded as a 'black box', containing various components which interact with other components through ports. The network is characterized by a square matrix called the Scattering- or S-matrix, which can be used to calculate the networks response to input

signals applied to the ports [104]. In a 2-port network, the S-matrix relates the incident electromagnetic waves to the reflected and transmitted waves according to Eq. 4.13 with the variables ($a_{1,2}$) and ($b_{1,2}$) as illustrated in Figure 4.17.

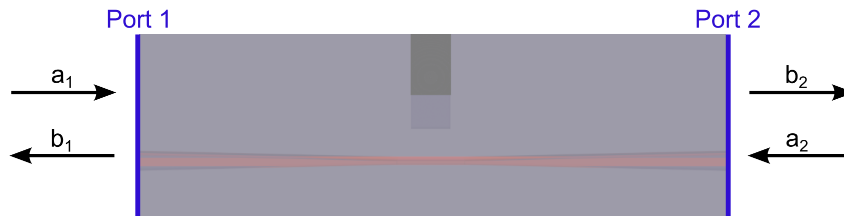

Figure 4.17: Illustration of the S-parameter approach, relating the input signal ($a_{1,2}$) applied to the ports to the output signal ($b_{1,2}$).

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \tag{4.13}$$

Expanding the matrices to equations gives:

$$b_1 = S_{11}a_1 + S_{12}a_2, \qquad b_2 = S_{21}a_1 + S_{22}a_2 \tag{4.14}$$

Here, *Port 1* is the input port and *Port 2* is the output port, where no incident wave is applied, hence a_2 is zero. This reduces Eq. 4.13 to $b_1 = S_{11}a_1$ and $b_2 = S_{21}a_1$. Therefore, S_{11} describes the amount of reflection, whereas S_{21} relates the incident wave to the transmitted wave.

Figure 4.18 shows the result of the frequency analysis for a waveguide tapered to 200 nm and different gap sizes. For a gap size of $g = 300$ nm, Figure 4.18 d), the electric field is maximal inside the waveguide structure and only marginally influenced by the Si structure next to it. By repeating the frequency analysis for decreasing gap size, the electric field decreases in the waveguide and increases in the Si double beam structure. For a gap size of 50 nm, Figure 4.18 a), the major part of the electric field leaks into the Si structure and transmission through the waveguide is suppressed significantly.

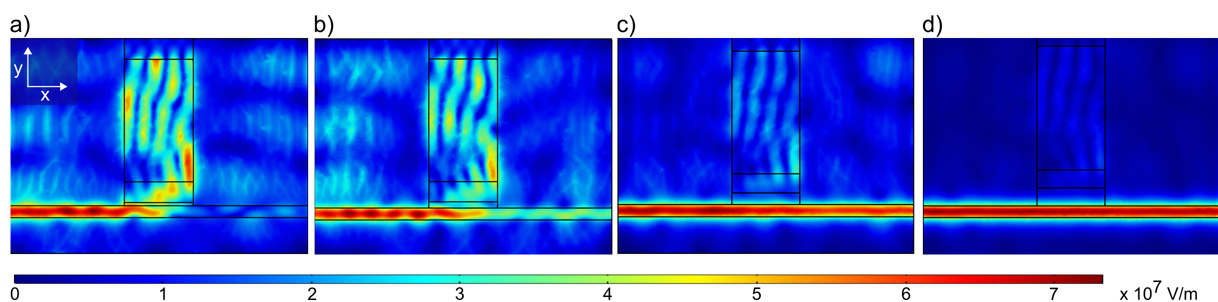

Figure 4.18: Results of the frequency analysis. Surface plot of the electric field norm $|E_z|$ at the cut plane at half waveguide thickness for waveguide width of $w = 200$ nm and gap size g of a) 50 nm and b) 100 nm c) 200 nm and d) 300 nm.

The qualitative analysis shown in Figure 4.18 can be completed by extracting the S-parameter for the transmission signal (S_{21}) and plot it versus gap size. The resulting dependency is shown in Figure 4.20 for a waveguide width of 200 nm. A clear exponential dependency is found, which is in line with the expectations, since the evanescent field decreases exponentially with increasing distance to the waveguide. The simulation is repeated for a waveguide width of 100 nm to estimate the influence of the waveguide width. For this waveguide width, no port mode can be found. Therefore, the geometry is modified. A tapered waveguide structure with an initial profile of 340 nm x 200 nm and a final profile of 340 nm x 100 nm is added next to the waveguide with the minimal waveguide width of 100 nm. The top view of the adapted geometry is shown in 4.19 a). With this modification the fundamental TEM_{00} mode is excited in the input port facet and guided along the waveguide structure. The surface plots of the electric field norm at the cut plane at half waveguide thickness are shown in Figure 4.19 for gap size g of 400 nm (b) and 100 nm (c).

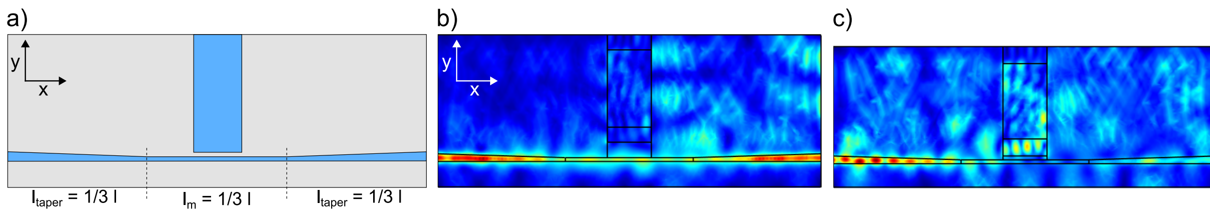

Figure 4.19: a) Top view of the modified geometry used to model a minimal waveguide width of $w = 100$ nm. Results of the frequency analysis show a lot of scattering of electric field norm $|E_z|$ for gap size g of b) 400 nm and c) 100 nm.

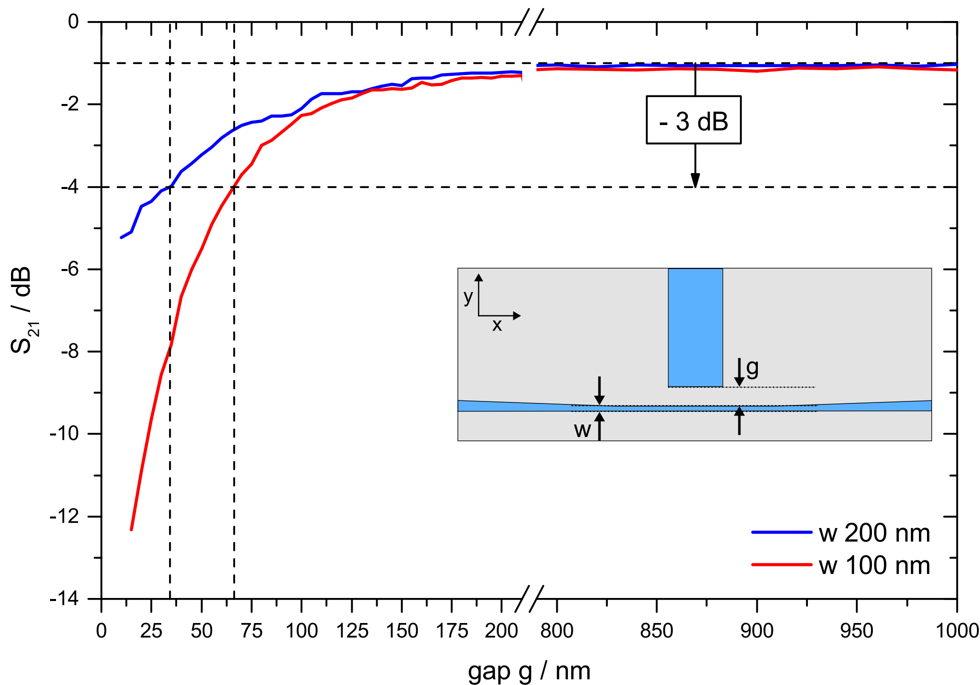

Figure 4.20: Transmission parameter S_{21} as a function of gap size for waveguide widths of 100 nm and 200 nm. A similar result is obtained in two dimensional calculations [105].

The simulations clearly demonstrate, how the transmission signal is modified by the gap size. Additionally, a decrease of the minimal width of the tapered waveguide enhances this dependency. For a minimal waveguide width of 100 nm the model suggests a gap size about 70 nm to achieve a sufficient on-off ratio of at least 3 dB. For a less tapered structure, e.g. $w = 200$ nm, smaller gap sizes below 50 nm are necessary.

During this work, also an alternative concept based on waveguide interruption has been investigated, which is illustrated in Figure 4.21. Similarly to the just presented simulation, the transmission dependency on actuator position has been modeled for the alternative design. Since the travelling wave can couple into and out-of the Si actuator as well as excite cavity modes for certain parameters, the resulting dependency between transmission and actuator position is nonlinear. Furthermore, the facets of the interruption of the waveguide structure have to be perfectly smooth, otherwise the transmission signal will be reduced to almost zero, while the influence of the actuator will be largely reduced. Due to these issues, this concept has not been further developed.

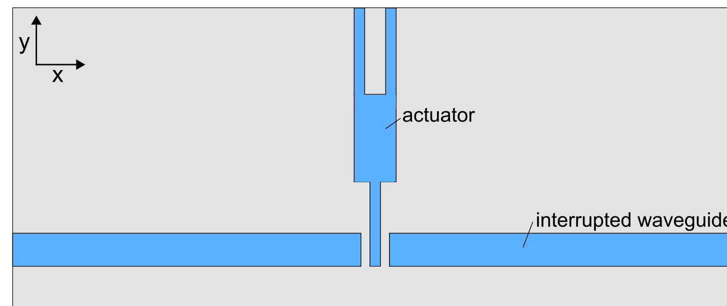

Figure 4.21: Schematic layout of an alternative concept, in which the actuator moves inside the gap of an interrupted Si waveguide.

5 Experimental Results

5.1 Electrical Performance

Electrical resistance measurements are performed using the in-situ measurement setup inside the SEM and inside the cryostat to gain insight in the progress of martensitic phase transformation and its dependency on the size of the specimen. As discussed in detail in 3.1, the first comprises high flexibility and therefore the characterization of a large number of nanostructures. Besides, actuation by Joule heating is more suitable for application. Due to this, the major part of experiments are performed by Joule heating and this chapter starts with the presentation of the corresponding results. A list of the investigated devices is shown in Table 1 of the Appendix.

5.1.1 Temperature Gradient

Using the in-situ measurement setup, allows to sweep the temperature of the double beam structures by Joule heating, thus the observable is electrical resistance versus electrical power. Each double beam structure is measured at least three times to guarantee a reproducible resistance characteristic. In each cycle, the current is linearly increased with time up to a preset value of maximum electrical heating power P_{max}. This value is kept constant for a unspecified time, after which the current is steadily decreased to zero. In general, P_{max} is not sufficient for a complete transformation in the first cycle. Therefore, in the subsequent cycles, P_{max} is increased, while keeping in mind that the choice of a too high value for P_{max} can lead to overheating. Figure 5.1 a) shows the electrical resistance characteristics of the first three cycles of a bimorph actuator of the first chip generation with a beam width of 500 nm and a beam length of 20 µm, respectively. In the second last cycle, the nanoactuator undergoes a full transformation, but at P_{max} the resistance changes with no variation of electrical power. This resistance change with time, in general resistance decrease, most likely indicates a material modification due to the high temperatures at the beam tip. To guarantee for a stable material system in the final measurement, P_{max} is kept constant until the resistance value remains unchanged. Hence, after the second last cycle a permanent shift in resistance compared to the first cycle is observed. The last cycle is similar to the second last cycle, besides, no further resistance drop at P_{max} occurs. The slope of the last cycle is reproducible, but in general the measurement routine is finished at this step.

To get an insight in the SMA characteristics, especially in terms of occurring size effects, the continuum model introduced in Chapter 4 is used. The model is capable of reproducing

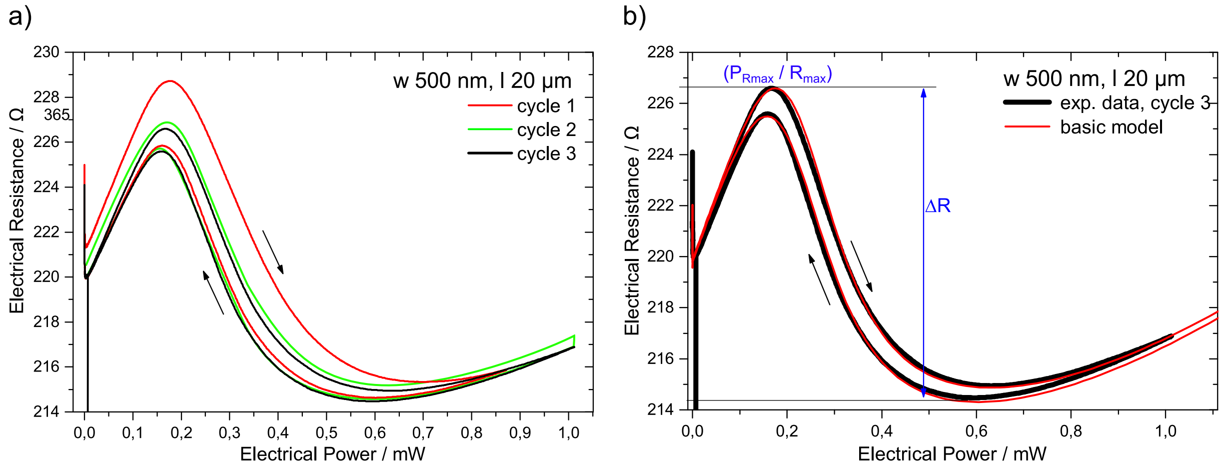

Figure 5.1: a) Electrical resistance - power characteristics recorded during the first three cycles. After the 3rd cycle, reproducible performance is obtained. b) Electrical resistance versus electrical power of the reference bimorph actuator with the layer thicknesses of NiMnGa and Si of 200 nm and 340 nm, respectively. The beam width and length are 500 nm and 20 µm, respectively. The result of the FEM simulation is shown as well, illustrating good agreement.

scaling effects, hence any deviation between simulation and experiment can be a hint towards emerging size effects. Key input parameters, for the simulation of the electrical resistance behavior upon Joule heating, are the electrical and thermal conductivity of the participating materials. The electrical conductivity of NiMnGa is given by the 200 nm thin film resistance data by $\sigma_{M,A}(T) = \frac{1}{R_{M,A}(T)} \cdot \frac{1}{d}$. The thermal conductivity of NiMnGa is given by the Wiedemann-Franz-law. For Si, the electrical conductivity is provided by the supplier and the thermal conductivity is approximated by $\kappa_{nano} = F \cdot \kappa_{bulk}$ with κ_{bulk} from [103] and F being a geometry dependent form factor. For more details see Chapter 4. A reference system is needed to determine the simulation parameters. For the first chip generation, the chosen reference bimorph has a beam width of 500 nm, a beam length of 20 µm and the layer thicknesses of NiMnGa and Si are 200 nm and 340 nm. The reference bimorph is fabricated on the same chip as the (nano)structures, to exclude effects caused by fabrication irregularities, like variations in the NiMnGa composition or different layer thicknesses of NiMnGa and/or Si. The measured electrical resistance value in Ω versus electrical power in mW is shown in Figure 5.1 b) together with the corresponding FEM characteristic.

Both characteristics show good agreement demonstrating the applicability of the model described in Chapter 4 for the martensitic phase fraction. To fit the absolute value of the electrical resistance, the d parameter has to be modified. The interpretation of a modified d value is similar to a change of the SMA layer thickness. To generate the high coincidence shown in Figure 5.1 the d value, thus the SMA layer thickness as it is implemented in the FEM model, is increased by an amount of approximately 15 % (d^*). The deviation between nominal layer thickness of 200 nm, and SMA layer thickness as implemented in the simulation $t_{SMA} = 230$ nm can be simply due to a fabrication inaccuracy during film deposition. Another reason for the deviation might be the different grain size of NiMnGa

on the Si nanostructures compared to the NiMnGa grain structure on the flat Si surface. As discussed in Chapter 4, the structuring does have some significant influence on the thermal conductivity of Si. This effect can be estimated by introducing a Form factor F. For Figure 5.1 this form factor is set to $F = 0.6625$, which goes in line with recent results on Ni/Si nano double beam structures. The chosen simulation parameters are shown in Table 5.1.

Table 5.1: Material parameters for modelling the electrical resistance characteristics upon Joule heating of the reference bimorph with 500 nm beam width, 20 µm beam length and the layer thicknesses of NiMnGa and Si of 200 nm and 340 nm. These values are kept constant for the device geometries of the first chip generation, which are called *basic model* in the following.

	basic model
$R_A(T)$, Ω	$5.58 + 0.001 \cdot T$
$R_M(T)$, Ω	$3.72 + 0.0088 \cdot T$
d^*	1.162
A_S / A_F, K	338.15 / 366.15
M_S / M_F, K	361.15 / 333.15
F	0.6625

The austenite start and finish temperatures, as well as the martensite start and finish temperatures, are adapted to fit the slope of the experimental data and reproduce the hysteresis most accurately. The resulting transition temperatures are listed in Table 5.1. The austenite start temperature merges with the 200 nm thin film value, whereas the austenite finish temperature is decreased by about 30 K compared to the 200 nm thin film reference. This observation is an enhancement of the trend visible already by comparing the 1 µm film with the 200 nm thin film.

In Figure 5.1 b) two key features can be highlighted, which can serve as comparative parameters: First, the peak position P_{Rmax} of the maximum resistance value R_{max} and second, the ΔR value defined as $\Delta R = \frac{R_{max} - R_{min}}{R_{max}}$.

Keeping the set of parameters given in Table 5.1 fixed, the different double beam geometries of the first chip generation are modeled. The resulting electrical resistance data are normalized and plotted in Figure 5.2, together with the corresponding experimental data.

It can be clearly seen that the transformation points of maximum and minimum electrical resistance are approximated quite well in the case of beam width w = 500 nm. In particular, the R_{max} value is barely shifted to lower power values with decreasing beam width. The ΔR value increases for decreasing beam width, which is not reproduced by the FEM model.

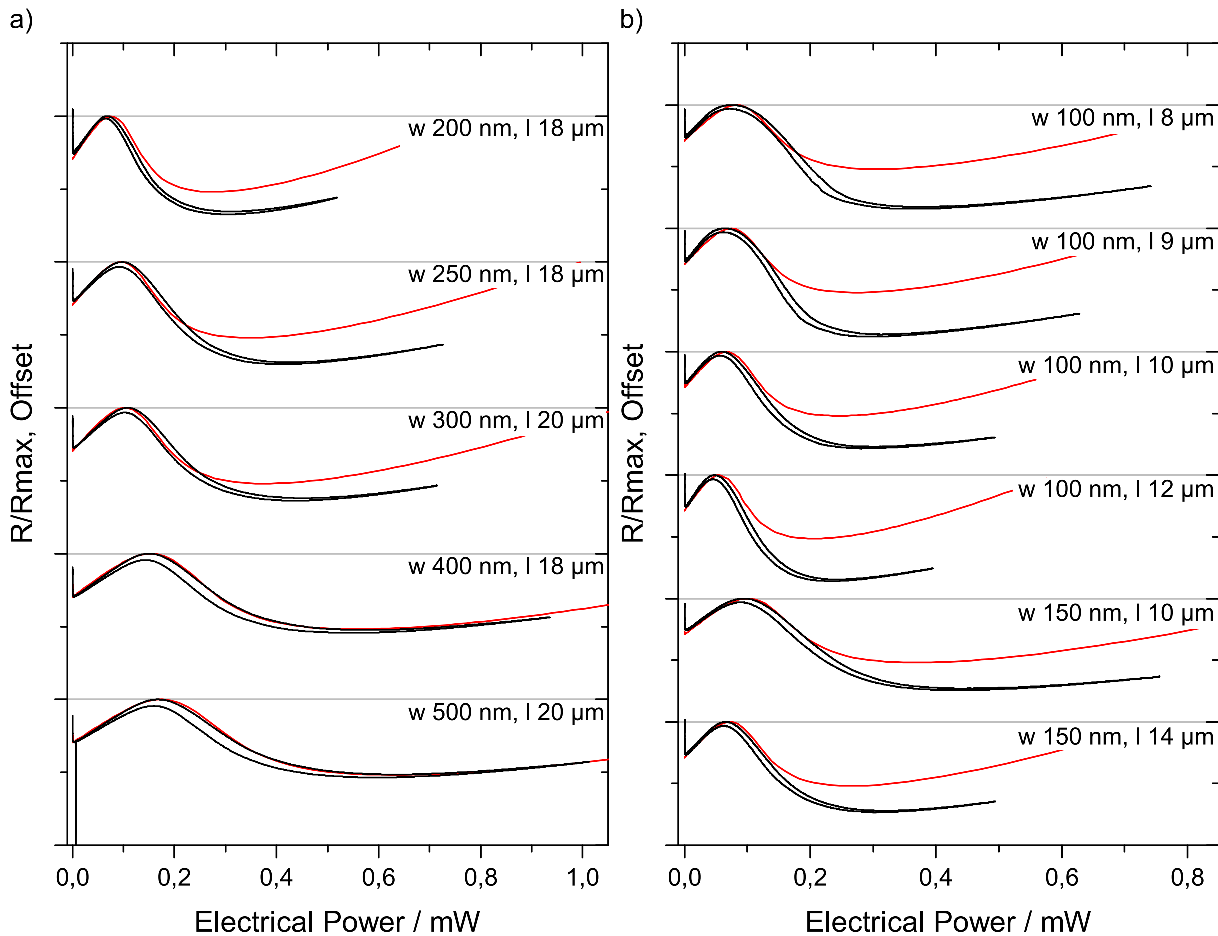

Figure 5.2: Normalized electrical resistance versus electrical power of NiMnGa/Si bimorph nanoactuators (black) and of corresponding simulated characteristics (red) with the layer thicknesses of NiMnGa and Si of 200 nm and 340 nm for different beam widths w and lengths l as indicated.

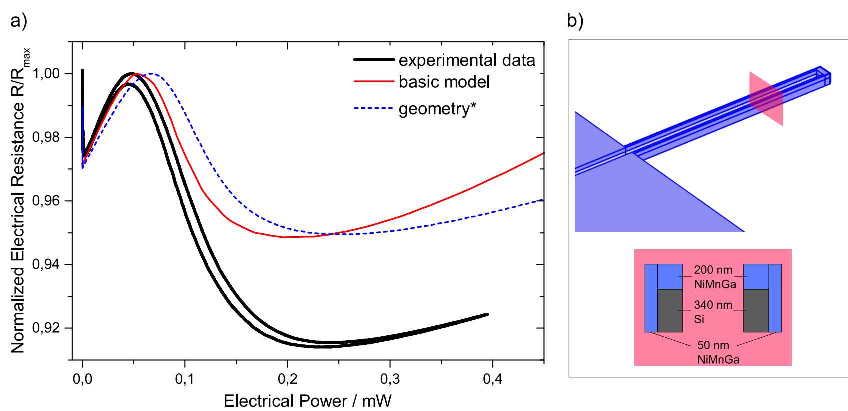

Figure 5.3: a) Normalized electrical resistance versus electrical power of a bimorph nanoactuator with the layer thicknesses of NiMnGa and Si of 200 nm and 340 nm and beam width and length of 100 nm and 12 µm, respectively. Besides the raw data the results of the FEM model for the basic geometry (red line) and a geometry taking NiMnGa sidewall deposition into account (blue line). The adapted geometry* as it is implemented in the FEM model is illustrated in b).

In the following, the effect of sidewall deposition is examined. Therefore, the bimorph geometry in the simulation has been modified by adding an extra layer of NiMnGa on the SMA/Si bimorph sidewalls with a thickness of 50 nm. The modified geometry is illustrated in Figure 5.3 b). Taking sidewall deposition into account does only marginally influence the normalized resistance versus electrical power signal, hence the approximation of modeling only the simplified geometry is considered to be sufficient.

The result of the electrical performance tests of the second chip generation, with layer thicknesses of 200 nm NiMnGa and 100 nm Si, is shown in Figure 5.4. The list including all investigated device geometries is shown in Table 1 of the Appendix. Again, the electrical resistance is normalized to its maximum value R_{max}.

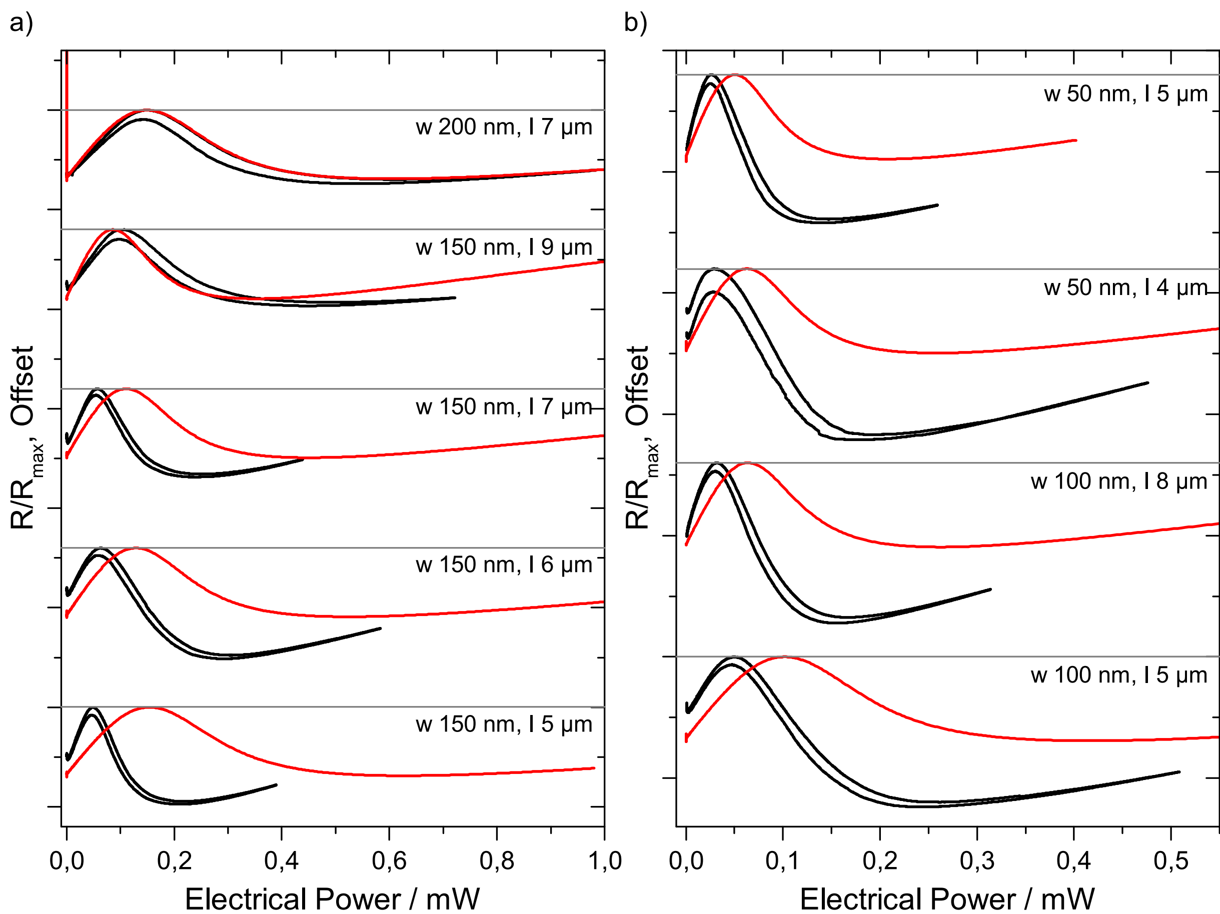

Figure 5.4: Normalized electrical resistance versus electrical power of NiMnGa/Si bimorph nanoactuators (black) and of corresponding simulated characteristics (red) with the layer thicknesses of NiMnGa and Si of 200 nm and 100 nm for different beam widths w and lengths l as indicated.

Again the electrical resistance characteristics are approximated by the FEM model. Therefore, the input parameters of the basic model, hence the input parameters for the first chip generation, had to be modified. Therefore the largest structure of the second chip generation was taken as a reference system. The chosen simulation parameters are shown in Table 5.2, which are denoted as *basic model**.

Table 5.2: Material parameters for modelling the electrical resistance characteristics upon Joule heating of the second generation of bimorph devices with the layer thicknesses of NiMnGa and Si of 200 nm and 100 nm.

	basic model*
$R_A(T)$, Ω	$5.956 + 0.0044 \cdot T$
$R_M(T)$, Ω	$3.958 + 0.00795 \cdot T$
d^*	1.9
A_S / A_F, K	340.65 / 398.15
F	0.9

Comparing the normalized electrical resistance characteristics of the different bimorph nanoactuators reveals a significant dependence of the position of the resistance maxima on the beam widths. While the resistance maximum of the bimorph with a beam with of 200 nm compares to the simulation, the resistance maxima of the nanoactuators with smaller beam width show a significant shift to lower power values. As for the first chip generation, the ΔR value increases for decreasing beam width. Table 5.3 shows ΔR normalized to the maximum resistance value for both chip generations.

Table 5.3: Normalized ΔR associated with the resistance drop during phase transformation for the 1st chip generation and the 2nd chip generation with the layer thicknesses of NiMnGa and Si of 200 nm and 340 nm and 200 nm and 100 nm, respectively. The beam width w and beam length l is indicated.

1st chip generation

w, l	$\Delta R \cdot 100$
500 nm, 20 µm	5.35
400 nm, 18 µm	5.43
300 nm, 20 µm	6.36
250 nm, 18 µm	7.01
200 nm, 18 µm	6.07
150 nm, 14 µm	7.37
150 nm, 10 µm	7.14
100 nm, 12 µm	8.59
100 nm, 10 µm	7.83
100 nm, 9 µm	8.78
100 nm, 8 µm	8.42

2nd chip generation

w, l	$\Delta R \cdot 100$
200 nm, 7 µm	3.69
150 nm, 9 µm	3.83
150 nm, 7 µm	4.43
150 nm, 6 µm	5.57
150 nm, 5 µm	4.87
100 nm, 8 µm	6.61
100 nm, 5 µm	6.19
50 nm, 5 µm	6.11
50 nm, 4 µm	7.05

An analysis of the shift can be made by having a detailed look on the peak values of the electrical resistance behavior. The shift in electrical power at R_{max} can be translated into a temperature by the use of the FEM model. The temperature profile links the peak position to the transition temperatures. Figure 5.5 a) shows a series of temperature profiles along the beam length for five different values of electrical heating power for a nanoactuator with 200 nm beam width an 7 µm beam length. This illustrates, that a increase in temperature gradient occurs with increasing electrical power. The inset in 5.5 b) illustrates the x-direction along the beam length, as well as the profile of the martensitic phase fraction at 0.40087 mW electrical power. The onset of the beam stays quite cold even at tip temperatures of 500 K. Thus, the double beam onset stays martensitic and the beam does not completely transform into austenite. For higher electrical heating power, temperature gradients of more than 50 K/µm occur along the beam direction.

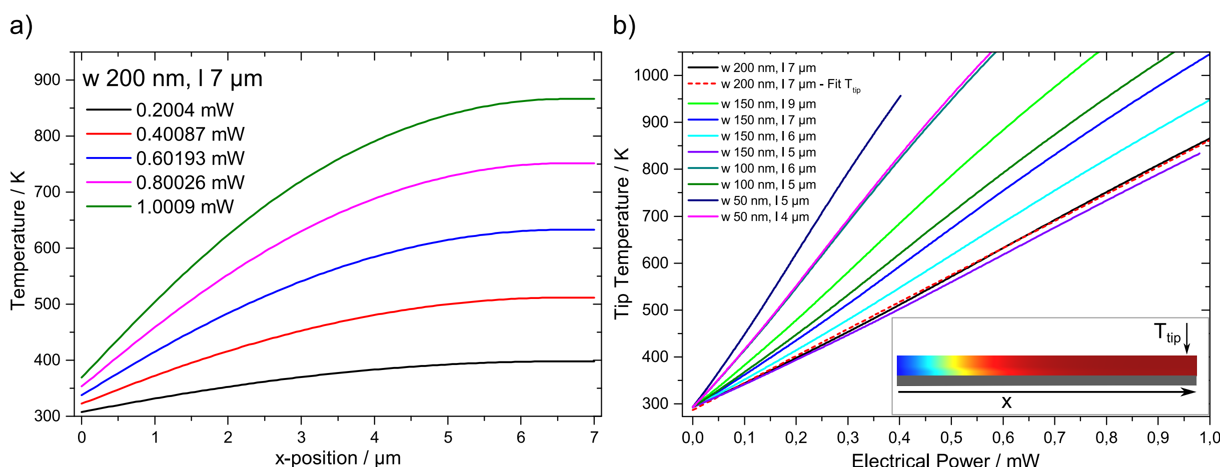

Figure 5.5: a) Illustration of the temperature profile of a nanoactuator with the beam width of 200 nm and beam length of 7 µm, respectively. b) Tip temperature as a function of electrical heating power for different beam widths w and lengths l as indicated. The layer thicknesses of NiMnGa and Si are 200 nm and 100 nm. The inset shows a schematic of the martensitic phase fraction at 0.40087 mW electrical power. Red indicating austenite and blue indicating martensitic phase.

Accordingly, the electrical resistance can be plotted versus a reference temperature. Most appropriate is the tip temperature, as it merges with the maximum temperature of the structure. The relation of the tip temperature with electrical power varies with the temperature profile, hence it is different for every cantilever geometry. Figure 5.5 b) gives an overview of the simulated $T_{tip}(P)$ characteristics for the devices of the second chip generation. To perform the variable transformation, from electrical power to tip temperature, a linear approximation of the dependency shown in Figure 5.5 b) is used. The linear fit, denoted as $Fit\ T_{tip}$ for a nanoactuator with 200 nm beam width and 7 µm beam length is included in 5.5 b) illustrating the deviation of the linear approximation. Figure 5.6 shows the corresponding electrical resistance versus tip temperature as exported directly from the simulation and the electrical resistance versus tip temperature calculated by the use of the linear approximation of $T_{tip}(P)$. It can be clearly seen, that the deviation is very small. Additionally, the measured electrical resistance data is shown as a function of the tip temperature by using the linear approximation of the simulation result. Both

characteristics are very similar, demonstrating that the procedure of converting electrical power to the reference temperature T_{tip} is appropriate.

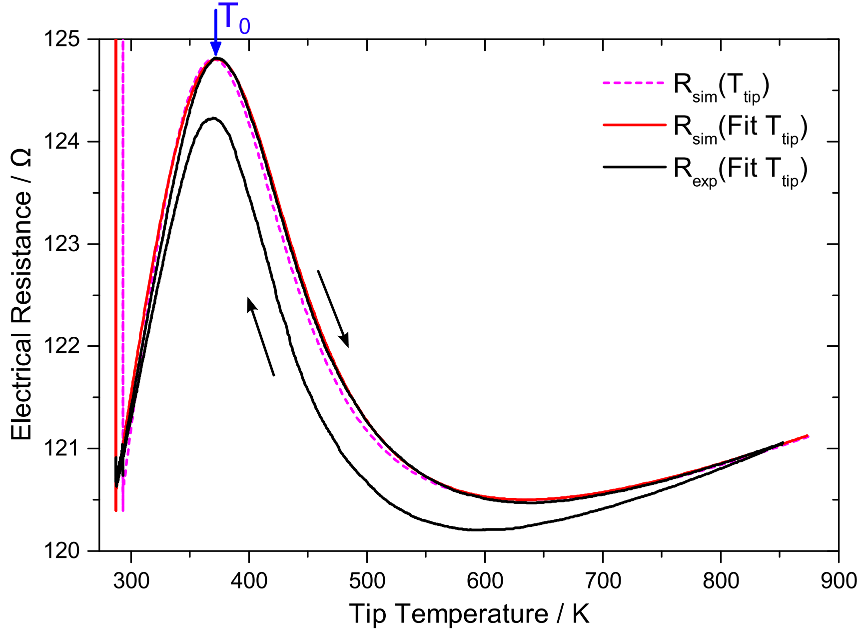

Figure 5.6: Normalized electrical resistance versus tip temperature of a NiMnGa/Si bimorph nanoactuator with 200 nm beam width, 7 µm beam length and the layer thicknesses of NiMnGa and Si of 200 nm and 100 nm. The data set directly exported from the simulation ($R_{sim}(T_{tip})$), the converted simulated data set (R_{sim}(Fit T_{tip})) and the converted experimental data set (R_{exp}(Fit T_{tip})) show good accordance, demonstrating that the linear approximation of $T_{tip}(P)$ is appropriate.

Figure 5.7 shows the normalized electrical resistance as a function of tip temperature for all nanoactuators of the second chip generation. The resistance maxima shift to lower temperature values by about 40 K for beam widths in the range of 150 nm to 50 nm. The corresponding temperature shifts, denoted as ΔT_0, are listed in Table 5.5.

To determine possible reasons for the deviation between simulation and experiment a parameter variation has been done with one of the smallest SMA/Si bimorph structures of the second chip generation, with a beam width of 50 nm and a beam length of 5 µm, respectively. For all the numerical calculations shown in Figure 5.4 the same Si Form factor $F = 0.9$ is used. The thermal conductivity of Si depends on the overall size of the geometry. In the presented structures the layer thickness is constant and equal to 100 nm. Therefore, a constant Si Form factor F can be chosen. However, the lateral dimension might further reduce the heat transfer in the Si layer, which can be implemented by a decrease of the F value. The effect of a reduced F parameter is shown in Figure 5.8. The reduced thermal conductivity of the Si layer leads to a shift in x-direction, but does not influence the ΔR value. Hence, the shift of A_S towards lower values can be interpreted by a lower thermal conductivity value. In previous work, Si Form factors in the range of $0.41 < F < 0.63$ were observed in Ni/Si bimorph double beam structures with beam width w in the range of $250\,\text{nm} < w < 450\,\text{nm}$ and layer thicknesses of NiMnGa and Si of 200 nm and 250 nm [93]. In addition, a reduction of the thermal conductivity up to

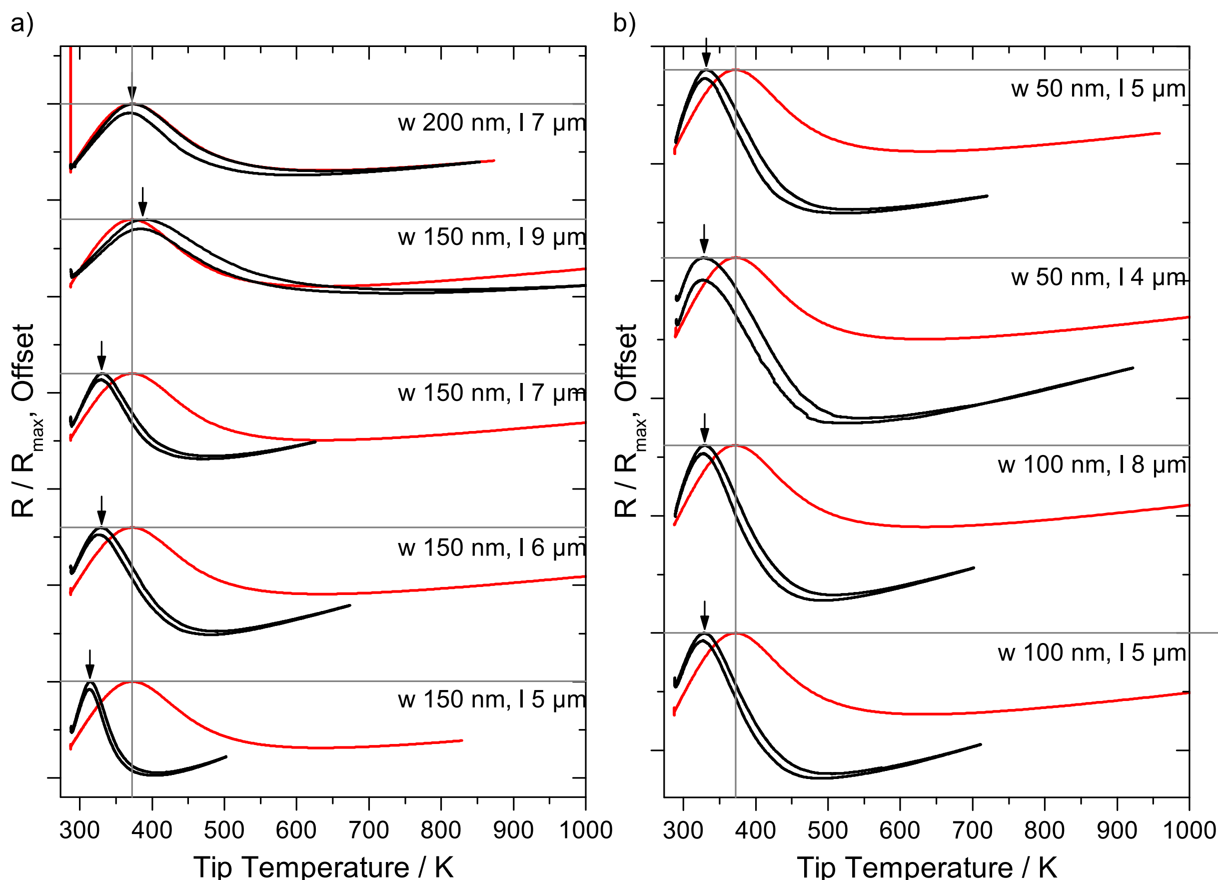

Figure 5.7: Normalized electrical resistance dependent on tip temperature of NiMnGa/Si bimorph nanoactuators (black) and of corresponding simulated characteristics (red) with the layer thicknesses of NiMnGa and Si of 200 nm and 100 nm for different beam widths w and lengths l as indicated.

60% compared the bulk value was observed in measurements on the Si device layer of SOI substrates with film thicknesses of 300 nm [106]. This corresponds to a Form factor of $F = 0.6$ of a non-structured Si thin film. In Si nanowires with diameters in the range of 115 to 22 nm, values in the range of $0.08 < F < 0.24$ were observed [107]. Moreover, an increase in surface roughness, e.g. due to an etching procedure, further reduces the thermal conductivity of Si [108]. Having the range of F parameters in mind, the used Form factor $F = 0.9$ serves as an upper limit. The critical dimensions of the fabricated NiMnGa/Si bimorph double beam structures are up to 50 nm small. Referring to the study on Si nanowires, a Form factor of $F < 0.24$ is possible. Furthermore, the Si layer was reduced by reactive ion etching, which might roughened the surface, resulting in even smaller F values. By decreasing the F parameter in the simulation from 0.9 to $F = 0.3$ the experimentally observed shift towards lower power values of R_{max} can be reproduced. Having the range of F parameters in mind, it is not possible to exclude a shift in A_S, but the effect can also be due to a further reduction in the thermal conductivity of Si. Hence, the ΔT_0 shown in Table 5.5 represent an upper limit, since a decreased κ_{Si} value would reduce the amount of ΔT_0.

Table 5.5: List of the approximated tip temperatures at the austenite start temperature T_0 for different beam widths w and lengths l as indicated. The layer thicknesses of NiMnGa and Si are 200 nm and 100 nm. The difference between the simulated and constant $T_0 = 372K$ value and the measured value is denoted as ΔT_0.

beam width w, length l	T_0	ΔT_0
200 nm, 7 µm	372 K	0 K
150 nm, 9 µm	389 K	-17 K
150 nm, 7 µm	331 K	41 K
150 nm, 6 µm	329.5 K	42.5 K
150 nm, 5 µm	314.5 K	57.5 K
100 nm, 8 µm	329.5 K	42.5 K
100 nm, 5 µm	329 K	43 K
50 nm, 5 µm	331.5 K	40.5 K
50 nm, 4 µm	328 K	44 K

In the following, the effect of a shifting of transition temperatures is examined. Therefore, the austenite start and finish temperatures as they are implemented in the FEM model are shifted by 40.5 K, which corresponds to the ΔT_0 value determined with the previously described method. The resulting electrical resistance characteristic is also shown in Figure 5.8.

A reduction of the transition temperature results in shift towards lower power values, but is also associated with a broadening of the resistance characteristic. The results of manually shifting the transition temperatures show, that a change in the ΔR value can refer to a change in A_S and A_F. However, a shift in ΔR can also be linked to non-transforming material. Recent results have shown, that for nanostructured devices and ultra-thin films the amount of e.g. residual austenite remaining untransformed can reach values of up to 40%. This effect is observed as a reduction of ΔR. Here, an increase in ΔR is measured for decreasing specimen size. This would imply, that the amount of residual austenite decreases with decreasing beam width, which is unexpected and needs further investigation.

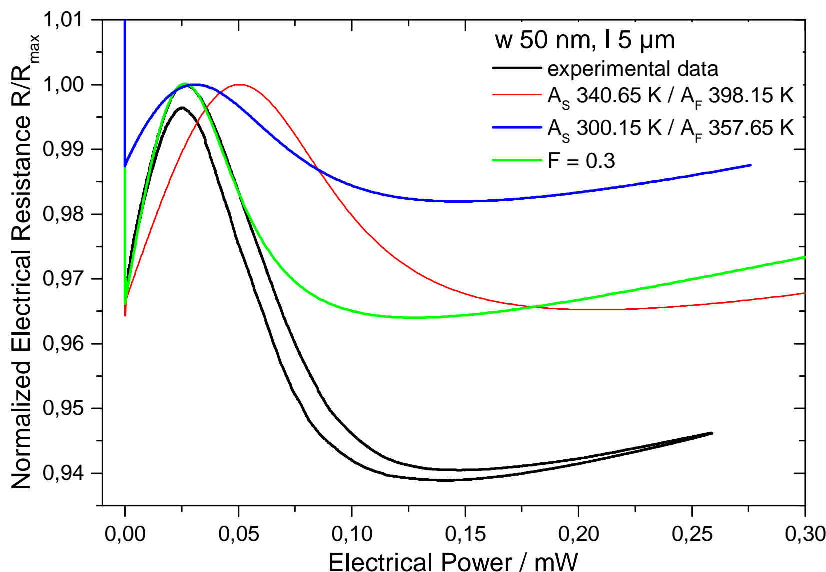

Figure 5.8: Normalized electrical resistance versus electrical power of a bimorph nanoactuator with the layer thicknesses of NiMnGa and Si of 200 nm and 100 nm and beam width and length of 50 nm and 5 µm, respectively. Besides the raw data the results of the FEM model for varying input parameters are shown: The effect of reduced thermal conductivity of the Si layer by $\kappa^*_{Si} = F \cdot \kappa_{Si}$ (green line) and the impact of shifted transition temperatures as indicated (blue line).

5.1.2 Homogeneous Temperature

In the following, results of the electrical resistance section the results of the electrical resistance measurement inside a cryostat are presented. With this setup, the flexibility given by the nanomanipulators is lost and the corresponding deflections cannot be recorded. However, the temperatures can be set homogeneously on the whole chip, allowing for a clear assignment of resistance values to transition temperatures. In Figure 5.9, the electrical resistance versus temperature is shown for a nonstructured area of the first chip generation, with layer thicknesses of 200 nm NiMnGa and 340 nm Si and the second chip generation, with layer thicknesses of 200 nm NiMnGa and 100 nm Si.

Both measurements show low noise level and low absolute resistance values. The characteristic slope in the resistance behavior of NiMnGa upon heating is observed, however, the resistance increase for temperature above A_F is missing. By approaching the Curie temperature of NiMnGa the increase of resistance with temperature is known to be reduced significantly. As a consequence the identification of the corresponding transformation is quite imprecise. For A_S and M_F, the tangential method is used, whereas a horizontal line is assumed to determine A_F and M_S. For the second generation, with layer thicknesses of 200 nm NiMnGa and 100 nm Si, no hysteresis is observed, hence A_S and M_F merge, as well as M_S and A_F. The transition temperature values are listed in Table 5.6.

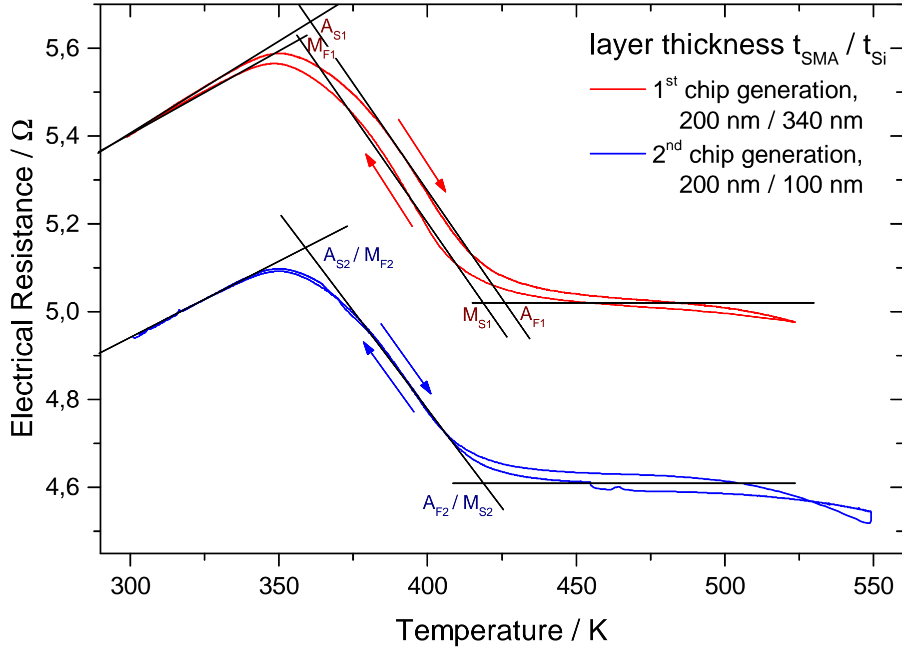

Figure 5.9: Thin film electrical resistance versus temperature measured by four point-probe measurement inside a cryostat on a nonstructured are of the first chip generation, with layer thicknesses of 200 nm NiMnGa and 340 nm Si and the second chip generation, with layer thicknesses of 200 nm NiMnGa and 100 nm Si. The transformation temperatures are indicated and determined by the tangential method.

Compared to the reference thin film sample all transition temperatures are higher on the non-structured areas of the two chip generations. This can be due to a systematic error concerning the temperature sensor or due to a variation concerning the material itself. In addition, the hysteresis is reduced to almost zero for the layer thicknesses of NiMnGa and Si of 200 nm and 100 nm.

Table 5.6: Transition temperatures determined by four-point-probe measurement inside a cryostat. The layer thickness of NiMnGa is 200 nm, the thickness of Si t_{Si} is indicated.

	A_S / K	A_F / K	M_S / K	M_F / K
1st chip generation, t_{Si} = 340 nm	360.65	426.15	418.15	357.65
2nd chip generation, t_{Si} = 100 nm	358.65	418.15	418.15	358.65
thin film reference, t_{Si} = 1 µm	338.15	390.15	391.15	339.15

In the following, each bimorph is measured with the same temperature sensor, the same SMU and, since they are on the same chip, also same film properties. The result is shown in Figure 5.10. Again, the increase of the electrical resistance after the completion of the phase transformation cannot be observed. Thus, the corresponding transition

temperatures, A_F and M_S, are not determined. Since the hysteresis is negligible for the majority of measurements, solely A_S is determined by the tangential method. The resulting A_S temperatures are listed in Table 5.7.

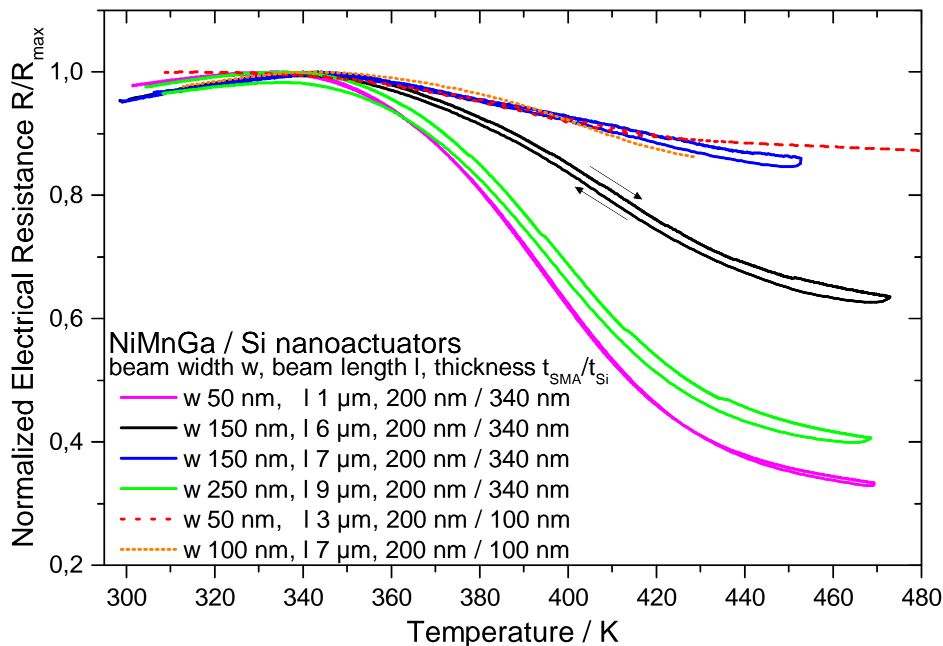

Figure 5.10: Temperature-dependent electrical resistance measured by four-point-probe setup inside a cryostat of NiMnGa/Si nanoactuators with the beam width w, the beam length l and the layer thicknesses of NiMnGa t_{SMA} and Si t_{Si}.

Table 5.7: Austenite start temperature A_S determined by four-point-probe measurement inside a cryostat. The beam width w, beam length l and the layer thicknesses of NiMnGa t_{Si} and Si t_{Si} are indicated. The difference between the measured thin film austenite start temperatures shown in Table 5.6 and the measured values for the individual bimorph nanoactuators is denoted as ΔT.

w / nm	l / µm	t_{Si}/ nm	t_{SMA} / nm	A_S / K	ΔT / K
50	1	340	200	357.15	3.5
150	6	340	200	360.15	0.5
150	7	340	200	344.15	16.5
250	1	340	200	356.15	4.5
50	3	100	200	344.15	14.5
100	7	100	200	353.15	5.5

The electrical resistance characteristic upon homogeneous heating shown in Figure 5.10 clearly reveals the ΔR change already observed in the Joule heating experiments. The

determination of the austenite start temperature is subject to major uncertainty due to the small resistance change between room temperature and A_S. However, the comparison with the thin film data measured on a non-structured part of the same sample, Table 5.6, reveals that the approximated A_S temperature of nearly all actuators is smaller than the thin film value. The largest shift of $\Delta T = 16.5$ K is observed for a nanoactuator of the first chip generation, hence layer thicknesses of NiMnGa and Si of 200 nm and 340 nm, with a beam width of 150 nm and beam length of 7 µm. Even though the absolute values of the transition temperatures are higher than expected, which might be due to a systematic error in the cryostat measurement, the trend corresponds to the outcome of the Joule heating experiment, that the austenite start temperature is shifted to lower values for decreasing device geometry. However, the different resistance characteristics of the two chip generations, which solely differ in the thickness of the bimorph base layer, indicate the high impact of the substrate thickness. However, a higher substrate-to-SMA layer thickness ratio does not implicitly result in a reduction of the SME. A hint towards this direction, is the reduction of the hysteresis to almost zero for layer thicknesses of NiMnGa and Si of 200 nm and 100 nm compared to 200 nm and 340 nm, respectively. Furthermore, the observation of decreasing ΔR value for decreasing beam width indicates a smaller amount of non-transforming material for a higher surface-to-volume ratio. Both effects demand for further investigation but also illustrate the high potential of nanoscale SMA devices.

5.2 Mechanical Performance

In the following the mechanical performance upon Joule heating of various NiMnGa/Si nanoactuators is presented. As described in Chapter 3.1 the SEM is used to capture videos parallel to the electrical resistance measurement. After drift correction the deflection behavior of the bimorph can be quantitatively analyzed. To compare the performance of actuators with different beam lengths the deflection at room temperature is set to zero and the absolute deflection value in µm is divided by the total beam length resulting in a relative deflection in %. The data for the first chip generation with the layer thicknesses of NiMnGa and Si of 200 nm and 340 nm is shown in Figure 5.11. The list including all investigated device geometries is shown in Table 1 of the Appendix.

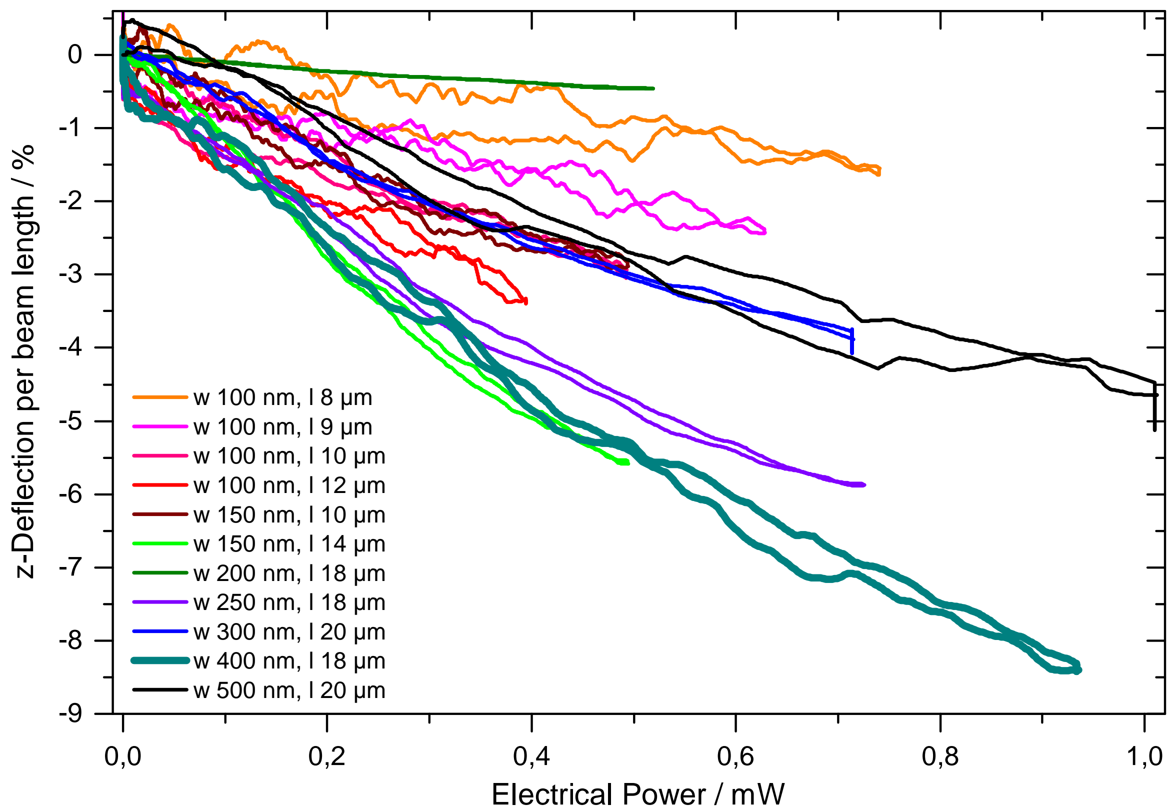

Figure 5.11: Deflection in z-direction in % of the total beam length normalized to the deflection value at room temperature versus electrical power for NiMnGa/Si bimorph nanoactuators of different beam widths w and different beam lengths l as indicated. The layer thicknesses of NiMnGa and Si are 200 nm and 340 nm.

Up front, several key features of the experimental data can be named. First the deflection behavior shows no significant slope change and only shows a slight bending at large electrical power values. Besides this hardly visible left turn, a very small hysteresis compared to the resistance measurements is observed. Additionally the noise of up to 50 nm is clearly visible, which is in contrast to the low noise of the resistance data. The temperature variation does not cause a shift of the SEM beam, but it can cause charging

of the surface, resulting in contrast and brightness changes of the SEM micrographs. This variation is enhanced by the detection angle of 85° and results in a non-constant brightness of the video data. Changing brightness values impede the tracking routine of the markers and lead to visible fluctuation of the deflection data.

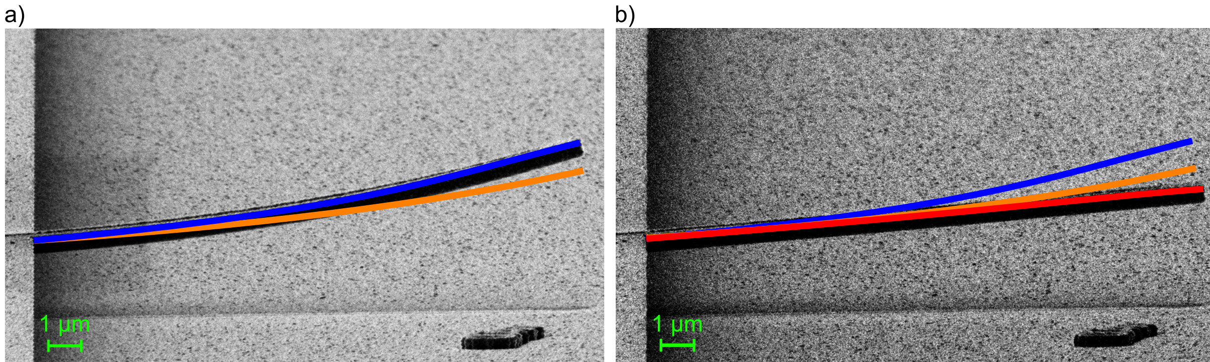

Figure 5.12: SEM micrographs of a NiMnGa/Si bimorph nanoactuator with 400 nm beam width and 18 µm beam length at room temperature (a) and at maximum temperature (b). The lines indicate the deflection at room temperature (blue), $T_1 > RT$ (orange) and T_{max} at maximum heating power (red). The layer thicknesses of NiMnGa and Si are 200 nm and 340 nm.

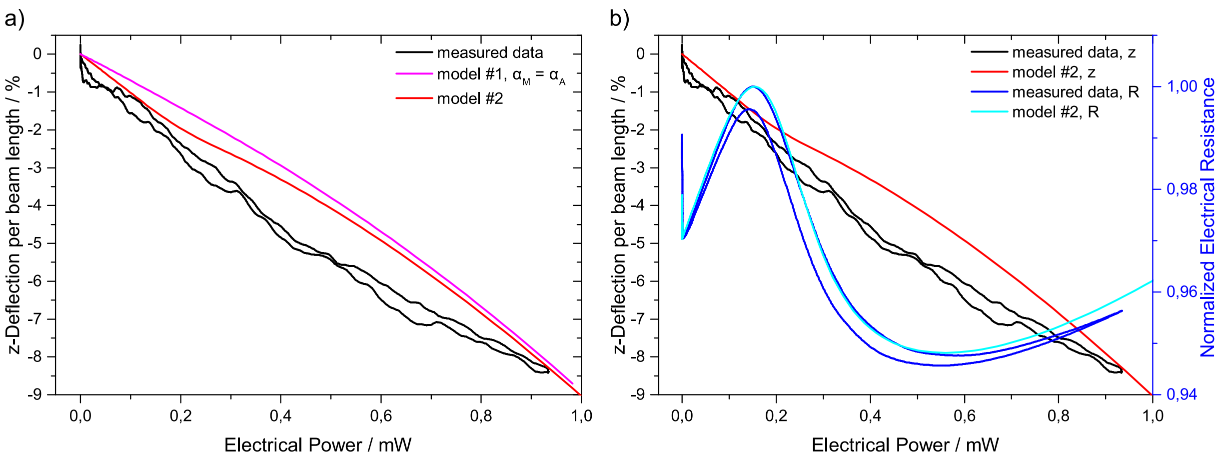

Figure 5.13: Deflection in z-direction in % of the total beam length normalized to the deflection value at room temperature (black) versus electrical power of the NiMnGa/Si bimorph nanoactuator shown in 5.12. The layer thicknesses of NiMnGa and Si are 200 nm and 340 nm, respectively, the corresponding beam width and length are 400 nm and 18 µm. In a) additionally the simulated deflection behavior using equal (model #1) / different (model #2) thermal expansion coefficients for the martensite and austenite is shown. In b) the normalized electrical resistance data and the corresponding simulation result is presented illustrating the good approximation of the electrical behavior and the deviation in the deflection behavior.

Figure 5.12 illustrates the temperature dependent course of deflection by representative SEM micrographs of a NiMnGa nanoactuator with 400 nm beam width and 18 µm beam length. In initial state at room temperature, the bimorphs are deflected upwards. The

RT deflection is set to zero in Figure 5.11 and highlighted in blue in Figure 5.12. Upon Joule heating, hence with increasing electrical power, the bimorph deflects downwards (orange). Figure 5.12 b) shows the maximum deflection value at the maximum electrical heating power (red).

For a more detailed analysis, the deflection versus electrical power is plotted together with the electrical resistance data and the simulation result in Figure 5.13 b) for the nanoactuator shown in Figure 5.13 a) having a beam width of 400 nm and a beam length of 18 µm. Whereas the basic model approximates the electrical resistance characteristic accurately, the deflection behavior is approximated correctly only in the first part. The deviation of simulated deflection and measured deflection increases significantly after the onset of the phase transformation, indicated by the slope change in the resistance curve. For comparison, the estimated deflection for a material with no transformation, hence constant thermal expansion coefficient (model #1), is included in Figure 5.13 a). It is clearly visible, that the hypothetical material shows a right turn, whereas the measured curve shows a slight left turn or no slope change. However, both FEM models do approximate the maximum deflection value quite well.

The result of the mechanical performance test of the second chip generation with a nominal layer thickness ratio of 200 nm NiMnGa and 100 nm Si is shown in Figure 5.14. The list including all investigated device geometries is shown in Table 1 of the Appendix. Again, the deflection at room temperature is set to zero. Furthermore, the deflection value is normalized to the total beam length. The corresponding electrical resistance characteristic is shown in the same plot to illustrate the progress of phase transformation.

It is clearly visible, that the deflection behavior of the second chip generation is very different from the first generation. First, the absolute deflection value is smaller. Whereas the normalized deflection is up to 8 % for the first generation, the deflection stays below 2 % for the second generation. This is associated with a worse signal to noise ratio, reflected by the fluctuations in Figure 5.14 a) - f). Besides, all measured bimorph nanoactuators of the second generation, having a thinner Si layer, show a significant slope change. In the first part, the deflection is negative, hence is directed downwards, but upon a certain power value the deflection increases again with smaller slope. The degree of deflection per heating power is given by $\frac{\Delta z}{\Delta P}$. For downward deflection direction this value is negative. The nanoactuators of the second chip generation do not only show a significant decrease of $|\frac{\Delta z}{\Delta P}|$, but even an upward deflection for the smallest beam width, hence a positive $\frac{\Delta z}{\Delta P}$ value is observed.

Figure 5.15 shows SEM micrographs recorded at a) room temperature and b) at maximum heating power. Compared to the first generation, the beams are deflected slightly downward at room temperature, which is highlighted in blue. Upon heating, the downward deflection value increases, indicated in orange. Upon a certain electrical power value, the deflection is either significantly reduced or takes place in the reverse, upward, direction, highlighted in red in Figure 5.15 b).

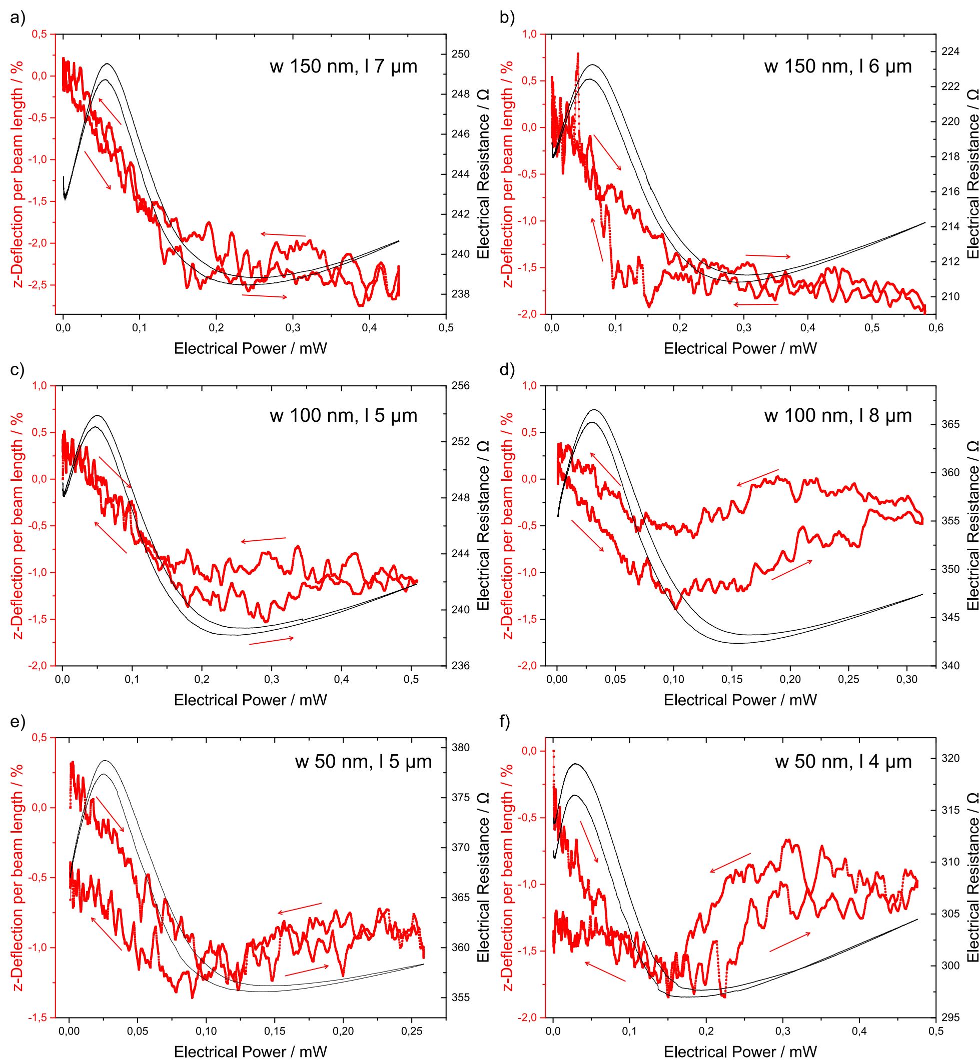

Figure 5.14: Deflection in z-direction per beam length normalized to the deflection value at room temperature (black) versus electrical power and corresponding electrical resistance data. The beam widths w and beam lengths l of the individual nanoactuators are indicated. The layer thicknesses of NiMnGa and Si are 200 nm and 100 nm.

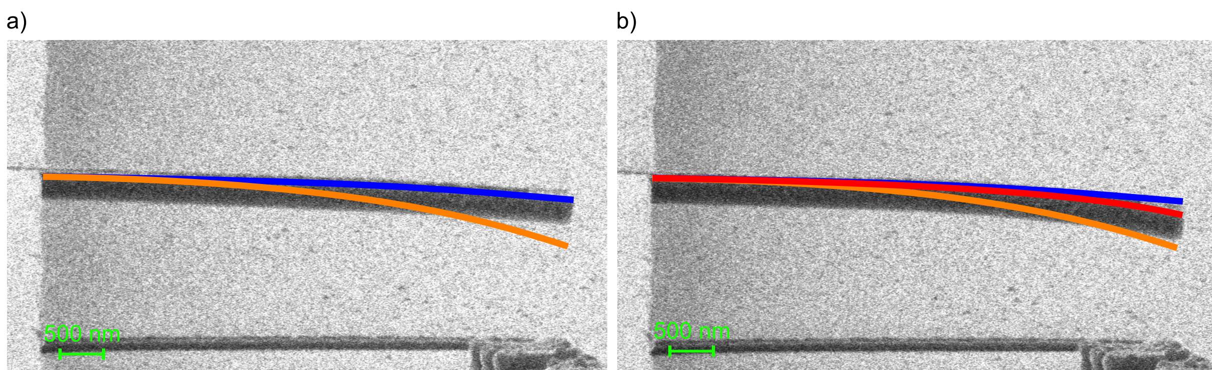

Figure 5.15: SEM micrographs of a NiMnGa nanoactuator with 150 nm beam width and 7 µm beam length at room temperature (a) and at maximum temperature (b). The lines indicate the deflection at room temperature (blue), $T_1 > RT$ (orange) and T_{max} at maximum heating power (red). The layer thicknesses of NiMnGa and Si are 200 nm and 100 nm, respectively.

In the following the results of the homogeneous temperature measurement are presented. The corresponding extraction of the deflection data out of the recorded SEM video files has been done manually, since the automatic marker tracking routine is not possible. The non-stationary drift of the nanoactuator of about 100 µm in x- and y-direction is much higher than the field of view. In comparison, upon Joule heating the drift is only approximately 100 nm. Due to these high displacements upon homogeneous heating, it is necessary to compensate for the drift while measuring. By moving the electron beam of the SEM, a drift of several µm can be compensated. However, the stage has to be moved several times during the measurement to guarantee, that the double-beam structure does not leave the field of view.

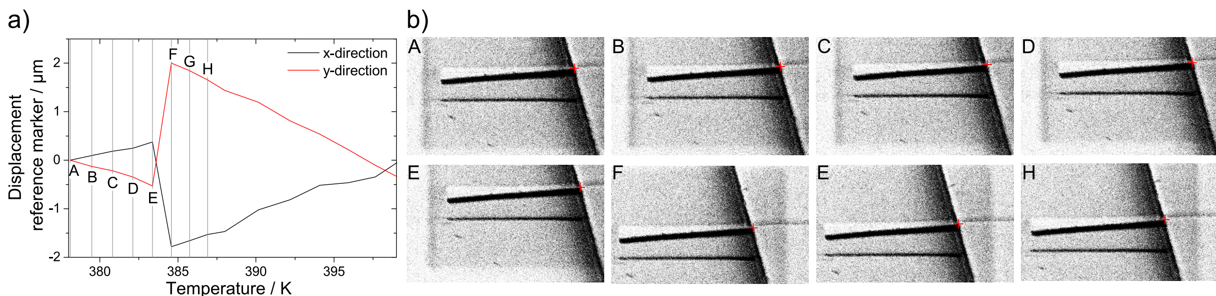

Figure 5.16: Illustration of the drift during the homogeneous temperature measurement. a) Displacement in x- and y-direction of the onset of the bimorph actuator upon heating and b) representative frames of the SEM micrograph video data.

In Figure 5.16 a) the movement of the onset of the bimorph actuator is shown together with SEM micrographs of 8 frames, named A to F. During frame A to E (and F to G) the structure drifts right- and upwards. To prevent the structure to leave the field of view, the stage is moved between frame E and F, which is visible in the displacement versus temperature curve by a sharp step. The marker tracking algorithm always searches for similar features in the next frame in the vicinity of the position in the prior frame. Thus large displacements cause a loss of the markers. The second reason for the manual extraction of the deflection data is the large change in brightness and contrast during the measurement. During the temperature increase and decrease, the brightness and contrast settings have to be modified several times. This results in significant changes of the double beam appearance. Thus the algorithm struggles with jumps due to the drift and changing features. The time consuming alternative is to position one reference marker at the onset of the beam and one marker at the beam tip by hand. The resulting temperature dependent deflection minus the deflection value at room temperature, normalized to the total beam length is shown in Figure 5.17 a). For comparison, the Joule heating resistance and deflection characteristics of the same bimorph actuator with a beam width of 150 nm, a beam length of 9 µm and layer thicknesses of NiMnGa and Si of 200 nm and 100 nm, are shown in Figure 5.17 b) .

The cooling has only been performed until about 40 °C due to the long cooling time. In absence of active cooling and the reduced convection at the low pressure inside the SEM (about 10^{-6} mbar), the cooling is solely enabled by temperature equalization by conductive heat transfer. Comparing the homogeneous deflection measurement (a) and the Joule heating measurement (b) the noise level is of the same order of magnitude.

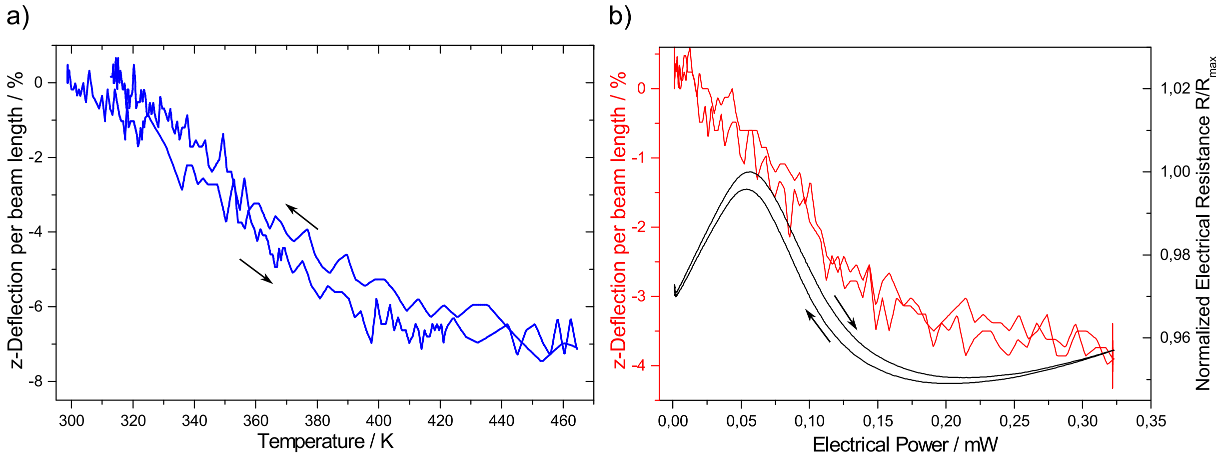

Figure 5.17: Normalized deflection characteristics upon homogeneous heating and cooling (a) and upon Joule heating (b) of a NiMnGa/Si bimorph nanoactuator with 150 nm beam width, 9 µm beam length and layer thicknesses of NiMnGa and Si of 200 nm and 100 nm.

Furthermore, the total deflection is roughly 8 % in (a) and below 4 % in (b). For the deflection the temperature of the onset of the bimorph plays an important role. As shown in Figure 5.5 in Chapter 5.1 the onset of the double stays quite cold even at high tip temperatures upon Joule heating. Thus the higher deflection value observed by homogeneous heating goes in line with the expectations. Besides, the observed slope of the deflection behavior in (a) and (b) agree very well. In Figure 5.18 the deflection characteristics of five other NiMnGa/Si bimorph actuators of the second chip generation are shown, having layer thicknesses of NiMnGa and Si of 200 nm and 100 nm, respectively.

For all bimorph structures of the second generation, a more or less distinct kink between 375 K and 400 K is observed. According to the results presented in Chapter 5.1, the austenite start temperature is evaluated to be below 360 K, thus significantly below the observed kink in the deflection characteristics.

Summing up, for the layer thicknesses of NiMnGa and Si of 200 nm and 340 nm, respectively, a downward deflection behavior upon Joule heating is observed, showing a small decrease of $|\frac{\Delta z}{\Delta P}|$ for temperatures exceeding A_S, "left turn". Also for a the layer thicknesses of NiMnGa and Si of 200 nm and 100 nm, respectively, a downward deflection is observed upon Joule heating. However, above a certain electrical power value, the deflection change is strongly reduced and even reverses for the smallest beam width of 50 nm. The deflection upon homogeneous temperature variation is about twice as much as the corresponding deflection upon Joule heating. Furthermore, the deflection change already visible in the Joule heating experiment is also observed for homogeneous heating. The corresponding kink is observed at a measured temperature of about 375 K - 400 K. For the interpretation of these results, the expected deflection behaviour of a SMA/Si bimorph cantilever from Chapter 4.1 is shown again in Figure 5.19 to illustrate the effect of differential thermal expansion.

In the first section, the cantilever is expected to deflect downwards due to the bigger thermal expansion coefficient of NiMnGa (martensite) compared to Si. When reaching

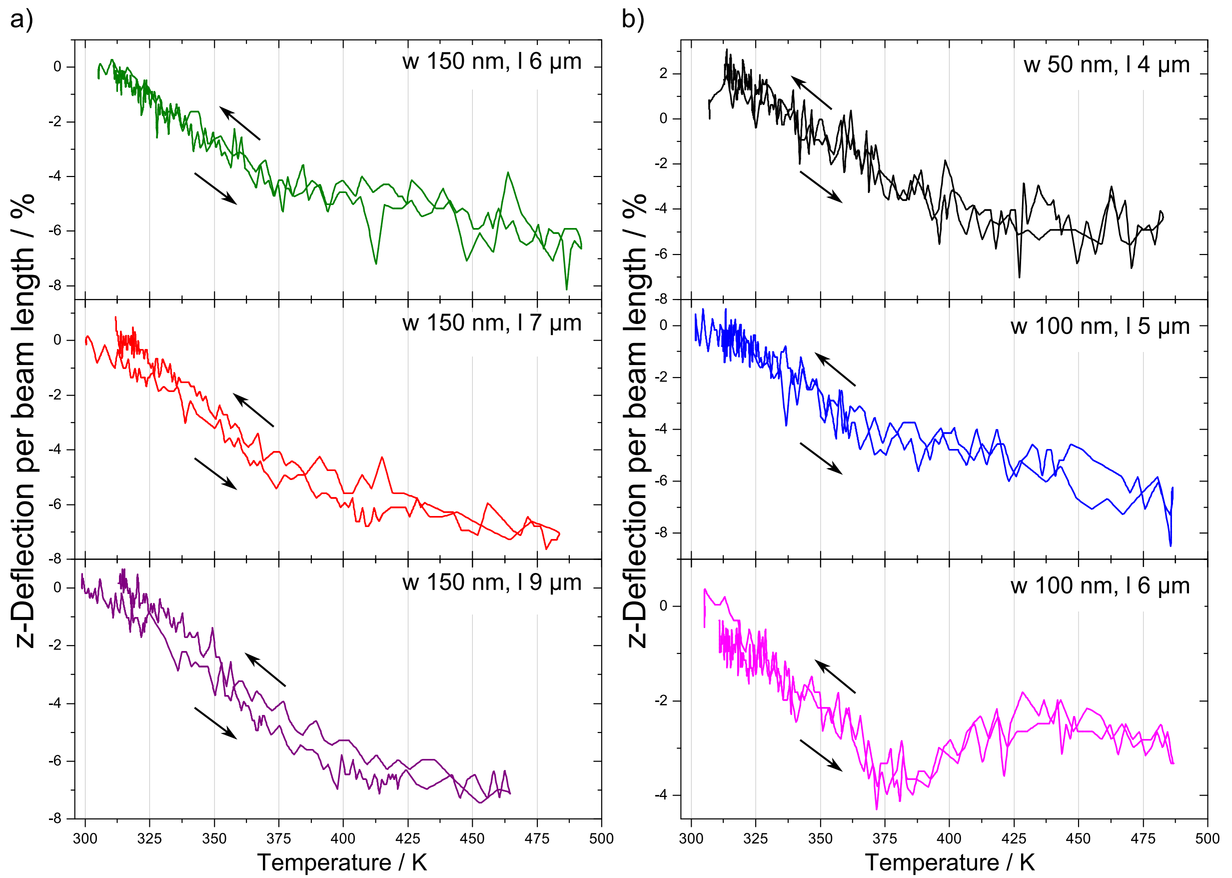

Figure 5.18: Normalized deflection in z-direction upon homogeneous temperature variation. The beam widths w and beam lengths l of the various nanoactuators are indicated and the layer thicknesses of NiMnGa and Si are 200 nm and 100 nm, respectively.

A_S, the stress relaxation upon martensitic transformation can reset the deformation, resulting in upwards deflection. For temperatures above A_F, downward deflection occurs again, according to the thermal expansion coefficients of NiMnGa (austenite) and Si. There is an optimum layer ratio for maximum upwards deflection in regime II, due to the necessary yield stress for the martensitic transformation. This layer ratio of NiMnGa / Si has been determined to be 90 nm / 250 nm or 1 / $2.\overline{7}$ as evaluated in Chapter 4.1.

The deflection upon Joule heating and the deflection upon homogeneous heating show a significant slope change. Even upwards deflection direction for the second chip generation having layer thicknesses of NiMnGa and Si of 200 nm and 100 nm, respectively, is observed. This is the first time, the SME is observed in the mechanical characteristics of nanostructured bimorph devices.

A more detailed look on the results reveals, that the temperatures associated with the kink, do not coincide with the austenite start temperature but are rather shifted to higher values. Referring to the Joule heating experiments, a shift of the austenite start temperatures to lower values, compared to the thin film, was observed. To explain this shift, there are two possible interpretation approaches. On the one hand, the

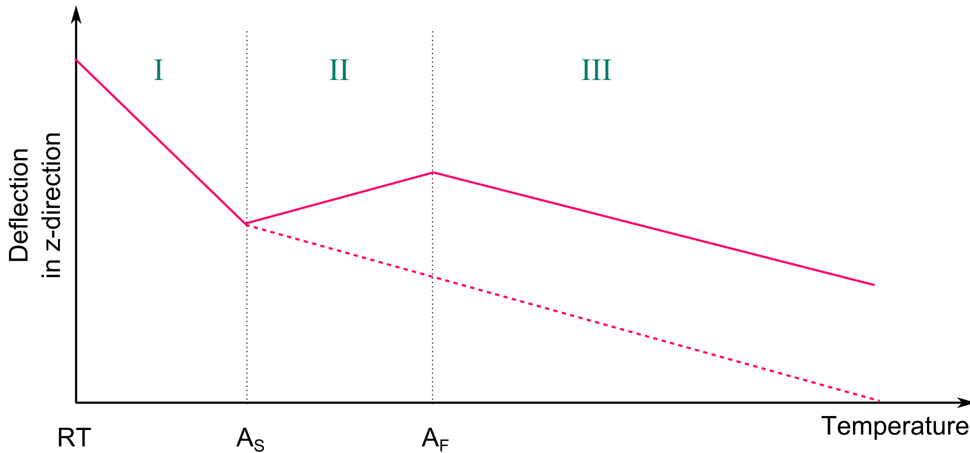

Figure 5.19: Schematic of the deflection behavior of a NiMnGa-Si bimorph nanoactuator as a function of temperature. In regimes I and III, the different thermal expansion coefficients of NiMnGa and Si lead to downward deflection for increasing temperature, whereas in regime II, the phase transformation may lead to resetting, i.e. upward deflection (solid line). However, if the stress is below a critical yield stress limit, the phase transformation in regime II does not lead to stress relaxation (dashed line).

transformation start could be at lower than observed temperatures or right at the start of the homogeneous-temperature-deflection measurement. The corresponding kink therefore would indicate the austenite finish temperature. This possibility is indicated in Figure 5.20 a) by $\overline{A_S}$ and $\overline{A_F}$ in green. This reasoning can not explain the significant slope change and especially not the upward deflection though. A more probable, and before considered, interpretation (denoted by A_S and A_F in blue) is that the temperature measurement is deficient, and has an error of approximately 10 %. As shown in the photograph in Figure 5.20 b) the temperature sensor is positioned at the backside of the sample, right next to the heating resistor. If there is a systematic error caused by the assembly, it would raise the measured temperature, shifting the figures temperature scale towards the same direction as the observed one, $T_{real} < T_{meas}$. Furthermore, the measurements depicted have not been done in thermal equilibrium, which would enhance the effect of higher measured temperatures for the evaluations even more. The significant slope change, showing high agreement with the simulated behavior, combined with the most likely systematic error, we are aware of, reveals a significant shape memory effect for the NiMnGa/Si bimorph nanoactuators.

According to the simulation model presented in Chapter 4.1 the amplification of the SME does not coincide with the simulated optimum layer thickness ratio. The optimum layer ratio of NiMnGa / Si has been determined to be about $1 \;/\; 2.\overline{7}$. Hence, the layer ratio of the first chip generation, 200 nm / 340 nm = 1 / 1.7, approaches the optimum value better than the second generation, 200 nm / 100 nm = 2 / 1 . Therefore, the enhanced SME of the second generation should be caused by another mechanism not yet considered in the simulation model or the simulation prediction fails for unknown reasons. So far we can only speculate on the possible root cause.

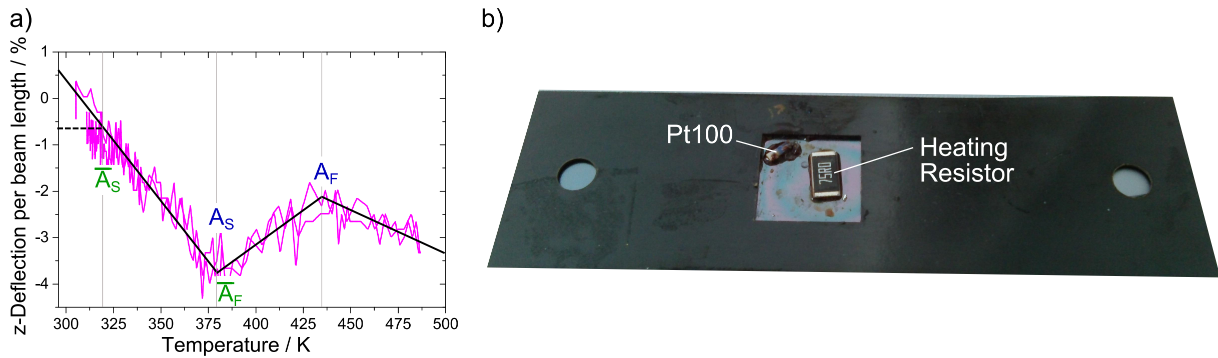

Figure 5.20: a) Normalized deflection in z-direction upon homogeneous temperature variation of a NiMnGa/Si bimorph nanoactuator with 100 nm beam width, 6 µm beam length and layer thicknesses of NiMnGa and Si of 200 nm and 100 nm, respectively. Both interpretation possibilities with the associated austenite start and finish temperatures are indicated. b) Photograph of the backside of the used sample transfer plate with the heating resistor and the PT100 temperature sensor glued to the backside of the chip.

The different room temperature deflection states might be an indicator for the deviation of the deflection behavior of the two nanoactuator generations. The initial deflection is due to a certain stress profile at room temperature. The upward room-temperature deflection (first generation) is a direct consequence of the hot sputter deposition of the NiMnGa layer and the ratio of thermal expansion coefficients of NiMnGa compared to Si. The initial downward deflection can be caused by additional intrinsic stress in the SOI device layer. This assumption is substantiated by the observation, that some of the freestanding Si cantilevers did point downwards prior to the NiMnGa deposition. During manufacturing of the SOI wafer, e.g. doping, intrinsic stress can be built in. Possibly, the dopant concentration is not constant over the Si device layer thickness resulting in varying intrinsic stress for different Si layer thicknesses. Another possibility might be the effect of thinning the Si device layer. Even though, no macroscopic changes of the surface roughness due to the additional reactive ion etching step have detected, there might be some structural or even chemical modification of the surface. The Si modifications might result, in combination with the hot sputter deposition, in a stress level sufficient for stress-induced martensitic transformation. The resulting stress profile might produce a significant volume fraction of NiMnGa exceeding the yield stress needed to undergo the martensitic transformation.

5.3 Optical Performance

A series of experiments on NiMnGa/Si nanoactuator structures is made to investigate whether the developed SMA-based NEMS actuator fulfills the requirements on actuation stroke for a sufficient on-off ratio of an optical switch device. The power-deflection characteristic is measured by recording SEM videos of the actuator deflection upon Joule heating. The measurement routine is similar to the one described in 3.1. Figure 5.21 shows in-situ electrical resistance and deflection characteristics determined simultaneously. An out-of plane deflection due to the bimorph effect of 100 nm is observed for about 64.5 µW and 200 nm for 118 µW. With respect to the simulation result presented in 4.3 the movement of the actuator is significantly higher than the minimum values for 3 dB on-off ratio. Besides the actuation cycles can be performed repeatedly with high reversibility.

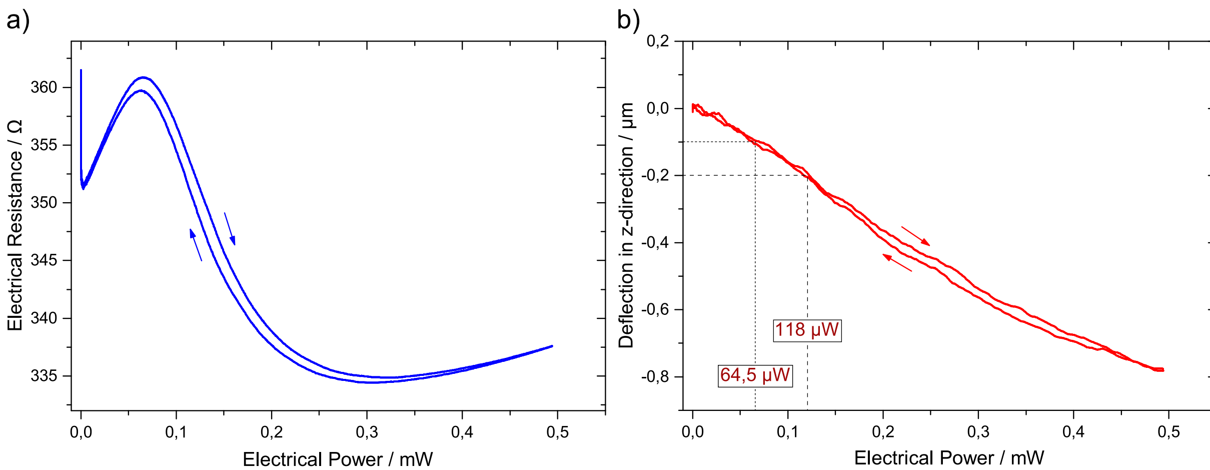

Figure 5.21: a) Electrical resistance and b) deflection characteristics upon Joule heating of a NEMS bimorph actuator with 150 nm beam width, 14 µm beam length and the nominal layer thickness of a NiMnGa to Si of 200 nm to 340 nm.

The optical performance is determined by measuring the transmission of test devices with different gap size between waveguide and cantilever. The measurement routine is described in detail in Chapter 3.3. The observed dependencies of optical transmission on gap size for two different waveguide widths w are shown in Figure 5.22 together with the simulation result. The modeled signal is normalized to experimental reference intensity for large gap distances to show the simulation and experimental data in the same plot. The optical losses are underestimated in the simulation, since the model only takes back-scatting due to tapering and coupling into the SiO_2 and air surrounding into account. For an actual slab waveguide device, the surface roughness is crucial for the transmission signal. Especially in the tapered waveguide section, the surface quality of the Si-waveguide plays a major role. Besides, any contamination, e.g. dust, in the vicinity of the tapered waveguide will reduce the transmission significantly, since the underlying principle of high sensitivity on the actuator position is directly associated with a high sensitivity to any inhomogenity in the area of interest. Therefore, the FEM transmission signal is corrected by the experimentally determined value for high gap distances by subtracting a constant value. The measured optical transmission values in 5.22 are completed by error bars

indicating the standard deviation between different test devices. The error in gap-size is $\pm$ 5 nm.

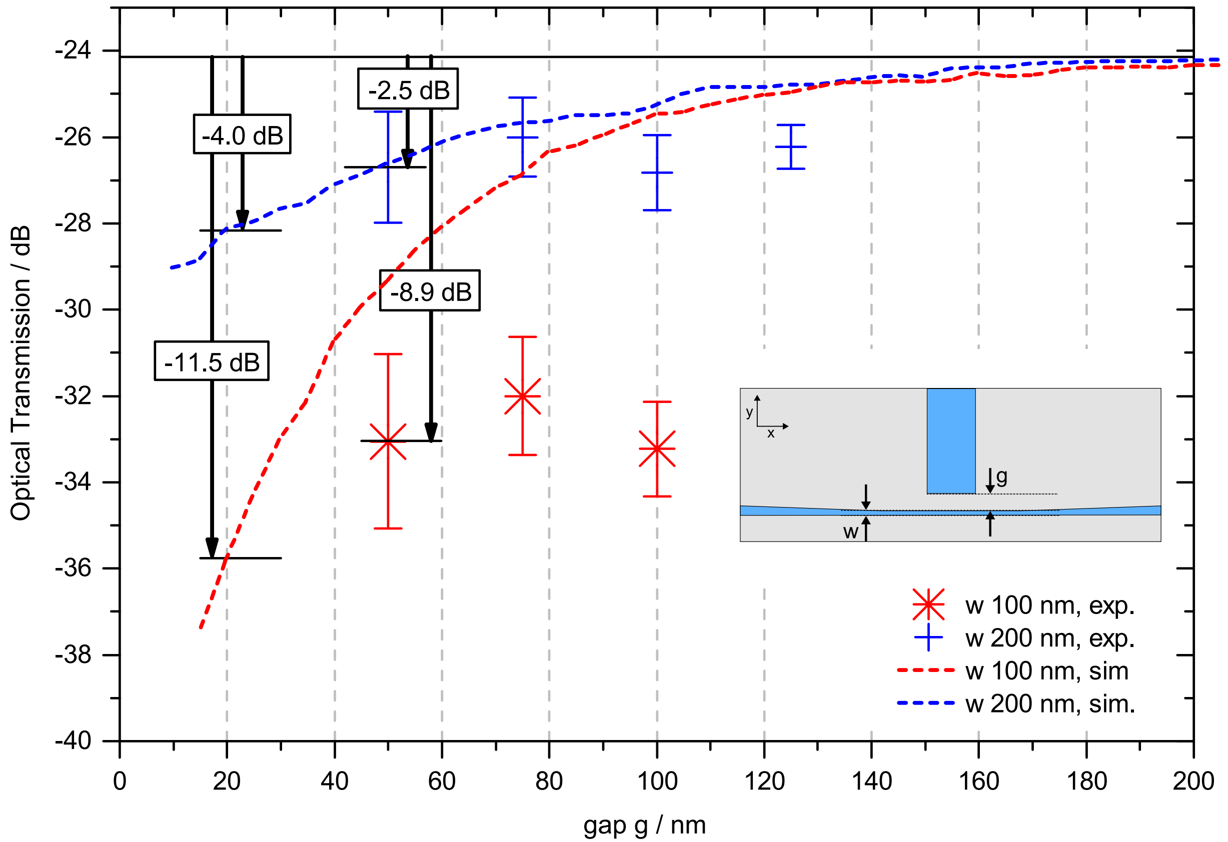

Figure 5.22: Optical transmission as a function of gap size between actuator structure and tapered waveguide. Transmission changes are indicated for gap size of g = 50 nm and g = 20 nm. Figure reproduced in accordance to [105].

Concerning the effect of waveguide width the optical transmission measurements are in line with the FEM model. For a gap size of $g = 50$ nm and waveguide width in the tapered section of $w = 200$ nm, a transmission change in the range of 1.2 dB to 3.8 dB is observed, matching the simulated behavior. For a minimal waveguide width of $w = 100$ nm the transmission signal drop is more pronounced and of the order of 9 dB. The simulation result is just not included in the scatter due to fabrication tolerances, but follows the same trend. For the same reason, no clear gap size dependency is observed for waveguide width of $w = 200$ nm and $w = 100$ nm. Most likely, the variation between two passive test chip devices is higher than the gap size dependency.

Within this chapter, not only the fabrication of complex functional nanostructures revealing the characteristic behavior of SMA down to the nanoscale is presented, but also a possible application as a optical on-chip device is fabricated and measured for the first time.

6 Conclusion & Outlook

SMAs provide highest energy density values and a favorable scaling behavior, which makes them promising candidates for applications in the nanoscale. This chapter concludes the thesis on the development of SMA/Si bimorph nanoactuators by a short summary and gives an outlook on opportunities for further optimization.

6.1 Conclusion

Within this thesis, NiMnGa/Si nanoactuators are fabricated and characterized and the results are compared with numerical calculations. Furthermore, the applicability of a NiMnGa-based NEMS actuator for an optical switch is investigated.

At first, a process for fabrication of nanometer-sized NiMnGa/Si actuators is developed. Therefore, the Si is structured prior to the NiMnGa deposition, creating free-standing Si double beam structures with critical dimensions down to 50 nm beam width at a beam length of 9 μm. Already the fabrication of Si nanostructures reveals an influence of the Si layer thickness on the room temperature deflection state. While cantilevers with Si layer thicknesses of more than 150 nm are flat after the process routine, structures with a thinner Si layer thickness tend to deflect slightly downwards.

In the following, NiMnGa is deposited onto the pre-structured Si substrate and the behavior of the NiMnGa/Si bimorph actuators is characterized in-situ upon Joule heating and homogeneous heating. For this purpose, an existing measurement setup is extended, to enable setting the temperature of the whole chip while monitoring the deflection of individual double beam structures. Upon Joule heating, the deflection and the electrical resistance versus electrical power of the bimorph double beam structures are recorded in parallel. By homogeneous temperature variation, the temperature-dependent deflection is measured. These setups enable the systematic study of the thermo-mechanical behavior of NiMnGa/Si bimorph actuators with varying dimension down to the nanometer regime. The corresponding noise level of the electrical resistance data is negligible and the variation of the deflection is approximately 10 nm.

A continuous finite element model capable of approximating the thermo-mechanical and electro-mechanical properties of SMAs is developed to compare with the experimental data to identify size effects. A good approximation of simulation and experimental data is found for devices with various beam widths and lengths and layer thicknesses of NiMnGa and Si of 200 nm and 340 nm, respectively. Based on the model, no unambiguous size-effect is identified, since the observed size-dependent variations of the characteristic transition temperatures are within the measurement accuracy. In first approximation the

deflection shows a linear behavior for devices with a layer thickness ratio of NiMnGa / Si of 200 nm / 340 nm. This reveals a dominance of the bimorph-effect, caused by the differential thermal expansion of NiMnGa and Si, compared to the shape memory effect.

In contrast, a significant deviation between FEM simulation and the behavior of devices with layer thickness ratio of NiMnGa / Si of 200 nm / 100 nm is observed. The electrical resistance is shifted towards lower heating power values for decreasing beam width, which indicates a size-effect. Moreover, the hysteresis is reduced to almost zero in temperature-dependent electrical resistance measurements. The hysteresis, caused by the martensitic transformation depicts one of the major drawbacks of SMAs concerning applications as sensors. Therefore, a reduction of the hysteresis width with decreasing specimen size gives rise to efficiency optimization and a variety of application opportunities especially in the field of sensing. In addition, the resistance drop associated with the martensitic transformation increases for decreasing beam width, which indicates a reduction of residual austenite. The amount of residual austenite refers to the volume faction of non-transforming shape memory material. With less residual austenite, hence more material participating in the phase transformation from martensite to austenite, the effect size increases which can lead to an overall performance enhancement.

Concerning the mechanical behavior, a significant deflection change and even a reversal of the deflection direction occurs for layer thicknesses of NiMnGa and Si of 200 nm and 100 nm. The latter is observed upon Joule heating and upon homogeneous heating and shows the influence of the SME. This is the first time the SME is observed in the mechanical response of nanostructured NiMnGa bimorph devices. The deviation between the characteristics determined for constant NiMnGa layer thickness of 200 nm and different Si layer thicknesses demonstrates the high impact of the substrate and, in particular, thickness of the substrate on the overall behavior of the specimen.

Besides the fundamental research on NiMnGa/Si nanoactuators, their suitability for an actual application is investigated based on the example of an electrically controlled integrated optical waveguide switch. For this purpose, an optical layout including waveguide and grating coupler structures is developed and integrated in the fabrication process of the SMA nanoactuators. In order to better understand the experimental observations, the dependency of the transmission signal on the actuator position has been approximated by three dimensional FEM simulations. These show that, the actuation stroke has to be in the order of 100 nm to achieve a on-off ratio of 3 dB, which is achieved by the SMA/Si nanoactuators fabricated during this work.

6.2 Outlook

The technology to fabricate, characterize and systematically study NiMnGa/Si nanoactuators reveals new opportunities to optimize functional nanoscale devices. Based on the combined results of simulation and experiment, different design rules can be formulated, whether aiming for maximized actuation stroke, maximal proportion of the SME effect compared to thermal bimorph effect or whether the investigation of size effects has priority.

In the following a selection of promising experiments, layouts and measurements setup extensions are presented.

Substrate constraints

One of the key results of this thesis is the high impact of the substrate and its thickness on the device characteristics. The investigation of the underlying physical mechanisms is not only of fundamental interest, but also paves the way for better actuation performance. Up to now, two different layer thicknesses of NiMnGa and Si have been compared. Therefore, a detailed study varying the Si thickness would lead to a deeper understanding of the intrinsic substrate constraints. Hereby, I suggest not only to reduce the thickness by an additional thinning step but also to compare wafers with different initial Si device layer thickness. This would help to investigate the impact of the thinning process. Moreover, a variation of the Cr thickness and its deposition temperature would further increase the knowledge about the initial stress state. The deposition of Cr at room temperature leads to a deflected state of the Cr/Si bimorph before reaching the sputtering temperature of NiMnGa. Finally, the complete removal of the base layer material, Si and Cr, would lead to an easier interpretation of measurement results. Si is the material of choice concerning future integration with other functional units, e.g. waveguide structures. Even though, the range of application by using Si substrates is the broadest, a substrate material which does not provide a size-dependent thermal conductivity is favorable for the determination of material parameters and the identification of size effects. For example, a noble metal, e.g. Ag, could be used. This would have the disadvantage of influencing the resistance signal of the NiMnGa/Ag bimorph double beam, but the thermal conductivity is less temperature dependent and does not decrease with decreasing specimen size.

SMA properties

Besides the study on substrate properties, a variation of the NiMnGa thin film properties could enable performance optimization. Figure 6.1 shows a SEM micrograph of one of the smallest NiMnGa/Si bimorph double beam structures. As can be clearly seen the grain size on the double beam cantilevers approaches the beam width of approximately 200 nm. By reducing the number of grains an oligo-crystalline structure is approached. The associated decrease of grain boundaries results in a decrease of defects impeding the reorientation. Therefore, a further reduction of the Si support structure dimensions, and therefore of the beam width, is expected to result in an enhanced SME.

As an alternative, a thin layer of MgO could be deposited on pre-structured Si samples by Atomic Layer Deposition (ALD) to enable epitaxial growth. The systematic study of epitaxial grown, hence single-crystalline SMA nanostructures, might reveal new insights in size-dependent characteristics of the SME.

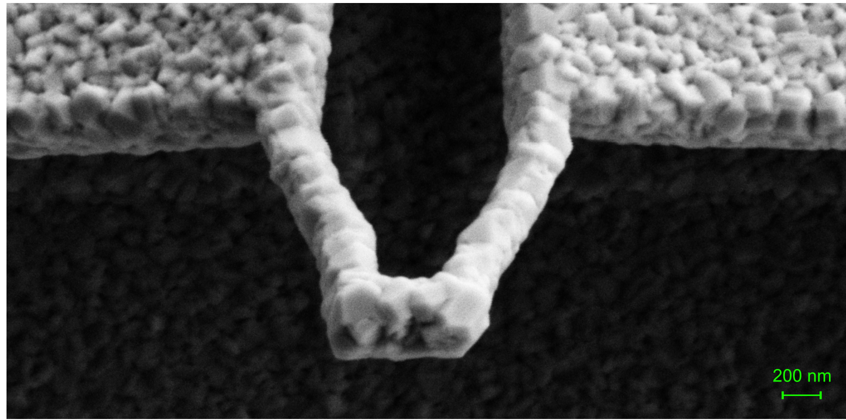

Figure 6.1: SEM micrograph of a NiMnGa/Si bimorph with the layer thicknesses of NiMnGa, Cr and Si of 200 nm, 50 nm and 200 nm, respectively. The grain size on the double beam structure is significantly larger than on the thin film and approaches the beam width of 200 nm.

Characterization

The access to more observables would be favourable with regard to direct observation of material characteristics of nanostructures but also in terms of optimizing the FEM model. The development of an in-situ tensile test platform, as it is shown schematically in Figure 6.2, would allow for the acquisition of stress-strain diagrams of NiMnGa cantilevers with nanometer-sized beam widths. Hence, the Young's Modulus of nanostructured SMA devices could be obtained, which would not only be of general scientific interest, but would also improve the FEM model. A major drawback concerning the simulation is the broad range of material parameters, since these have to be estimated based on bulk data. Furthermore the combination of SEM with Digital Image Correlation (DIC) post-processing could be used to generate local strain maps. The design of an in-situ mechanical test-platform is shown in figure 6.2 including possible results. Based on this design first tests are already performed within the group [109].

Besides the further improvement of the in-situ measurement setup, the development of an intrinsic deflection sensor is of high interest. Due to the small device geometries the deflection measurement is relying on the SEM. However, in the SEM no magnetic fields can be applied, since they would divert the scanning electrons. A measurement of magnetic field dependent deflection therefore has not been done yet. A possible realization would be, to use the developed optical switch as an intrinsic deflection sensor. The deflection of the actuator can be used to tune the transmission through a waveguide structure, however the transmission signal can also operate as an indicator for the deflection of the actuator. In the actual design the actuator is electrically driven, but with a similar layout the deflection while varying the magnetic field could be examined.

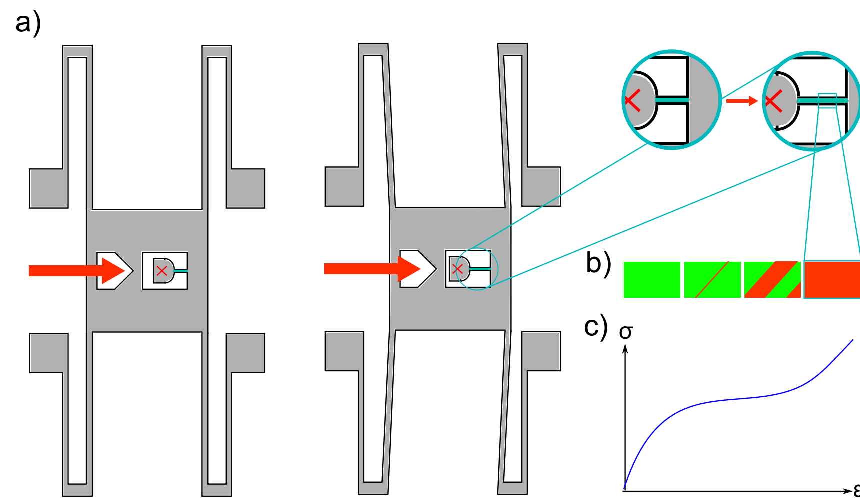

Figure 6.2: a) Layout and illustration of an mechanical test platform for in-situ force measurement of SMA nanostructures. By applying a force, the test platform is moved, leading to stretching of the nanostructure. The layout allows acquisition of stress-strain diagrams (c) or even local strain maps, evaluated by SEM-DIC, revealing strain band formation at the nanoscale (c).

Application

The combination of high temperature sputter deposited SMA structures with Si waveguides is one of the main achievements of this thesis. As a first application, an SMA-based actuator for optical switching is presented. The actual device is designed as an on-off switching device, however, it is capable of tuning the transmission signal in a continuous way. With an alternative design it is possible to create not only a 1×1 switch, but a 1×2 or 1×n relation of input to output ports. In such a design, the actuator guides the light from one output to another instead of just "disturbing" the signal.

In this work, electric, mechanical, and optical performance are discussed with regard to potential applications but also in terms of fundamental research. The achievements show the high potential of SMA nanoactuators as smart components for complex systems. However, it is of major interest to continue the research on nanoscale SMA devices to get a deeper understanding of the underlying physical mechanisms and to use this knowledge to integrate this technology into actual devices.

Appendix

A - Characteristics of NiMnGa/Si bimorph actuators

	w / nm	**l / µm**	t_{SMA} **/ nm**	t_{Si} **/ nm**
1st chip generation				
	500	**20**	**200**	**340**
	400	18	200	340
	300	20	200	340
	250	18	200	340
	200	18	200	340
	150	14	200	340
	150	10	200	340
	100	12	200	340
	100	10	200	340
	100	9	200	340
	100	8	200	340
2nd chip generation				
	200	**7**	**200**	**100**
	150	9	200	100
	150	7	200	100
	150	6	200	100
	150	5	200	100
	100	8	200	100
	100	5	200	100
	50	5	200	100
	50	4	200	100

Table 1: Geometry characteristics of the double beam structures of the first and 2nd chip generation. Double beam width w, double length l, thickness of the SMA layer t_{SMA} and thickness of the silicon layer t_{Si} are indicated.

Bibliography

[1] T. Kudernac, N. Ruangsupapichat, M. Parschau, B. Macia, N. Katsonis, S. R. Harutyunyan, K. H. Ernst, and B. L. Feringa, "Electrically driven directional motion of a four-wheeled molecule on a metal surface", *Nature* **479**, 208–211 (2009). (p. 1)

[2] D. C. Lagoudas, *Shape Memory Alloys - Modeling and Engineering Applications* (Springer Science Business Media, LLC, 2008). (p. 1, 2, 5)

[3] K. Bhattacharya and R. D. James, "The Material Is the Machine", *Science* **307**, 53–54 (2009). (p. 1)

[4] M. Kohl, M. Gueltig, V. Pinneker, R. Yin, F. Wendler, and B. Krevet, "Magnetic Shape Memory Microactuators", *Micromachines* **5**, 1135–1169 (2014). (p. 1)

[5] W. M. Huang, Z. Ding, C. C. Wang, J. Wei, Y. Zhao, and H. Purnawali, "Shape Memory Materials", *MaterialsToday* **13**, 54–61 (2010). (p. 5)

[6] K. Bhattacharya, "Crystallographic Attributes of a Shape-Memory Alloy", *Journal of Engineering Materials and Technology* **121**, 93 (1999). (p. 6)

[7] J. Cui, Y. S. Chu, O. Famodu, Y. Furuya, J. Hattrick-Simpers, R. D. James, A. Ludwig, S. Thienhaus, M. Wuttig, Z. Zhang, and I. Takeuchi, "Combinatorial search of thermoelastic shape-memory alloys with extremely small hysteresis width", *Nature Materials* **5**, 286–290 (2006). (p. 6)

[8] T. Tadaki, K. Otsuka, and K. Shimizu, "Shape Memory Alloys", *Annual Review of Materials Science* **18**, 25–45 (1988). (p. 7)

[9] R. Zarnetta, R. Takahashi, M. L. Young, A. Savan, Y. Furuya, S. Thienhaus, B. Maaß, M. Rahim, J. Frenzel, H. Brunken, Y. S. Chu, V. Srivastava, R. D. James, I. Takeuchi, G. Eggeler, and A. Ludwig, "Identification of Quaternary Shape Memory Alloys with Near-Zero Thermal Hysteresis and Unprecedented Functional Stability", *Advanced Functional Materials* **20**, 1917–1923 (2010). (p. 7)

[10] Y. I. Yoo and J. J. Lee, "Two-way Shape Memory Effect of NiTi under Compressive Loading Cycles", *Physics Procedia* **22**, 449–454 (2011). (p. 8)

[11] F. Heusler, "Beitrage zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung", *Verh. Deutsche Physikalische Gesellschaft* **5**, 219 ff (1903). (p. 9)

[12] J. Pons, V. A. Chernenko, E. Cesari, and V. A. L'vov, "Stress-strain temperature behaviour for martensitic transformation in Ni-Mn-Ga single crystals compressed along 001 and 110 axes", *Journal de Physique IV France* **112**, 939–942 (2003). (p. 9)

[13] O. Heczko, L. Straka, and K. Ullakko, "Relation between structure, magnetization process and magnetic shape memory effect of various martensites occuring in Ni-Mn-Ga alloys", *Journal de Physique IV France* **112**, 959–962 (2003). (p. 9)

[14] A. A. Likhachev, A. Sozinov, and K. Ullakko, "Different modeling concepts of magnetic shape memory and their comparison with some experimental results obtained in Ni-Mn-Ga", *Material Science & Engineering A* **378**, 513–518 (2004). (p. 9)

[15] C. W. Lay, *Herstellung und In-situ-Charakterisierung von Formgedächtnis-Bimorph-Nanoaktoren* (Dissertation Karlsruher Instituts für Technologie (KIT), 2015). (p. 9, 38, 50, 70)

[16] A. N. Vasil'ev, A. D. Bozhko, V. V. Khovailo, I. E. Dikshtein, V. G. Shavrov, V. D. Buchelnikov, M. Matsumoto, S. Suzuki, T. Takagi, and J. Tani, "Structural and magnetic phase transitions in shape-memory alloys $Ni_{2+x}Mn_{1-x}Ga$", *Physical Review B* **59**, 1113 1119 (1999). (p. 10)

[17] J. Tellinen, A. Suorsa, Jääskeläinen, I. Aaltio, and K. Ullakko, "Basic properties of magnetic shape memory actuators", *Proceedings of the 8th International Conference ACTUATOR 2002* **59**, 566–569 (2002). (p. 11)

[18] A. Sozinov, N. Lanska, A. Soroka, and W. Zou, "12 magnetic field-induced strain in Ni-Mn-Ga-based non-modulated martensite", *Applied Physics Letters* **102**, 021 902 (2013). (p. 11)

[19] K. Ullakko, J. K. Huang, C. Kantner, R. C. O'Handley, and V. V. Kokorin, "Large magnetic field induced strains in Ni2MnGa single crystals", *Applied Physics Letters* **69**, 1966–1968 (1996). (p. 11)

[20] R. A. Kellogg, A. B. Flatau, A. E. Clark, M. Wun-Fogle, and T. A. Lograsso, "Temperature and stress dependencies of the magnetic and magnetostrictive properties of Fe0.81Ga0.19", *Journal of Applied Physics* **91**, 7821–7823 (2002). (p. 11)

[21] I. Chopra, "Review of state of art of smart structures and integrated systems", *AIAA Journal* **40**, 2145–2187 (2002). (p. 11)

[22] D. J. Bell, T. J. Lu, N. A. Fleck, and S. M. Spearing, "MEMS actuators and sensors: observations on their performance and selection for purpose", *Journal of Micromechanics and Microengineering* **15**, 153–164 (2005). (p. 12)

[23] P. Truitt, J. B. Hertzberg, C. C. Huang, K. L. Ekinci, and K. C. Schwab, "Efficient and Sensitive Capacitive Readout of Nanomechanical Resonator Arrays", *Nano Letters* **7**, 120–126 (2007). (p. 12)

[24] X. M. H. Huang, X. L. Feng, C. A. Zorman, M. Mehregany, and M. L. Roukes, "VHF, UHF and microwave frequency nanomechanical resonators", *New Journal of Physics* **7**, 247 (2005). (p. 12)

[25] J. Hartbaum, *Magnetisches Nanoaktorsystem* (Dissertation Karlsruher Instituts für Technologie (KIT), 2013). (p. 12)

[26] J. S. Lee, D. S. W. Park, A. K. Nallani, G. S. Lee, and J. B. Lee, "Sub-micron metallic electrothermal actuators", *Journal of Micromechanics and Microengineering* **15**, 322 (2005). (p. 12)

[27] D. Yan, A. Khajepour, and R. Mansour, "Design and modeling of a MEMS bidirectional vertical thermal actuator", *Journal of Micromechanics and Microengineering* **14**, 841–850 (2004). (p. 12)

[28] L. Yang, Y. Yu-Hsin Lin, Y. Hu, and C. Liu, "A large-displacement thermal actuator designed for MEMS pitch-tunable grating", *Journal of Micromechanics and Microengineering* **19**, 015 001 (2008). (p. 12)

[29] R. B. Karabalin, M. H. Matheny, X. Feng, E. Defay, G. Le Rhun, C. Marcoux, S. Hentz, P. Andreucci, and M. L. Roukes, "Piezoelectric nanoelectromechanical resonators based on aluminum nitride thin films", *Applied Physics Letters* **95**, 103 111 (2009). (p. 12)

[30] S. Chu, "Laser manipulation of atoms and particles", *Science* **253**, 861–866 (1991). (p. 12)

[31] M. Li, H. P. Pernice, C. Xiong, T. Baehr-Jones, M. Hochberg, and H. X. Tang, "Harnessing optical forces in integrated photonic circuits", *Nature* **456**, 480–484 (2008). (p. 12)

[32] J. Peirs, D. Reynaerts, , and H. Van Brussel, "Scale effects and thermal considerations for micro-actuators", *Proceedings. 1998 IEEE International Conference on Robotics and Automation* **2**, 1516–1521 (1998). (p. 12)

[33] J. R. Greer and J. T. De Hosson, "Plasticity in small-sized metallic systems: Intrinsic versus extrinsic size effect", *Progress in Materials Science* **56**, 654–724 (2011). (p. 12, 13, 14)

[34] B. Clark, D. Gianola, O. Kraft, and C. Frick, "Size independent shape memory behavior of nickel-titanium", *Advanced Engineering Materials* **12**, 808–815 (2010). (p. 13)

[35] C. Frick, S. Orso, and E. Arzt, "Loss of pseudoelasticity in nickel-titanium submicron compression pillars", *Acta Materialia* **55**, 3845–3855 (2007). (p. 13)

[36] C. Frick, B. Clark, S. Orso, A. Schneider, and E. Arzt, "Size effect on strength an strain hardening of small-scale [1 1 1] nickel compression pillars", *Material Science and Engineering: A* **489**, 319–329 (2008). (p. 13)

[37] C. P. Frick, B. G. Clark, A. S. Schneider, R. Maaß, S. Van Petegem, and H. Van Swygenhoven, “On the plasticity of small-scale nickel-titanium shape memory alloys”, *Scripta Materialia* **62**, 492–495 (2010). (p. 13)

[38] C. P. Frick, B. G. Clark, P. Sonnweber-Ribic, and E. Arzt, “Orientation-independent pseudoelasticity in small-scale NiTi compression pillars”, *Scripta Materialia* **59**, 7–10 (2008). (p. 13)

[39] J. Ye, R. K. Mishra, A. R. Pelton, and A. M. Minor, “Direct observation of the NiTi martensitic phase transformation in nanoscale volumes”, *Acta Materialia* **58**, 490–498 (2010). (p. 13)

[40] J. M. San Juan, M. L. Nó, and C. A. Schuh, “Superelasticity and shape memory in micro- and nanometer-scale pillars”, *Advanced Materials* **20**, 272–278 (2008). (p. 13)

[41] J. M. San Juan, M. L. Nó, and C. A. Schuh, “Nanoscale shape-memory alloys for ultrahigh mechanical damping”, *Nature Nanotechnology* **4** (2009). (p. 13)

[42] Y. Chen and C. A. Schuh, “Size effects in shape memory alloy microwires”, *Acta Materialia* **59**, 537–553 (2011). (p. 13)

[43] V. A. Chernenko, M. Ohtsuka, M. Kohl, V. V. Khovailo, and T. Takagi, “Transformation behavior of Ni-Mn-Ga thin films”, *Smart Materials and Structures* **14**, 245–252 (2005). (p. 13)

[44] D. König, P. J. S. Buenconsejo, D. Grochla, S. Hamann, J. Pfetzing-Micklich, and A. Ludwig, “Thickness-dependence of the B2-B19 martensitic transformation in nanoscale shape memory alloy thin films: Zero-hysteresis in 75 nm thick Ti51Ni38Cu11 thin films”, *Acta Materialia* **60**, 306–313 (2012). (p. 13)

[45] P. Lazpita, G. Rojo, J. Gutierrez, J. M. Barandiaran, and R. C. O’Handley, “Correlation Between Magnetization and Deformation in a NiMnGa Shape Memory Alloy Polycrystalline Ribbon”, *Sensor Letters* **5**, 65–68 (2007). (p. 13)

[46] M. Chmielus, X. X. Zhang, C. Witherspoon, D. C. Dunand, and P. Müllner, “Giant magnetic-field-induced strains in polycrystalline Ni-Mn-Ga foams”, *Nature Materials* **8**, 863–866 (2009). (p. 14)

[47] Y. Ganor, d. Shilo, T. W. Shield, and R. D. James, “Breaching the work output limitation of ferromagnetic shape memory alloys”, *Applied Physics Letters* **93**, 122 509 (2008). (p. 14)

[48] M. Reinhold, D. Kiener, W. B. Knowlton, G. Dehm, and P. Müllner, “Deformation twinning in Ni-Mn-Ga micropillars with 10M martensite”, *Journal of Applied Physics* **106**, 053 906 (2009). (p. 14)

[49] C. A. Schuh, J. K. Masonanda, and C. Lind, “Quantitative insight into dislocation nucleation from high-temperature nanoindentation experiments”, *Nature Materials* **4**, 617–621 (2005). (p. 14)

[50] C. Cisse, W. Zaki, and T. B. Zineb, “A review of constitutive models and modeling techniques for shape memory alloys”, *International Journal of Plasticity* **76**, 244–284 (2016). (p. 15)

[51] E. Patoor, D. C. Lagoudas, P. B. Entchev, L. C. Brinson, and X. Gao, “Shape memory alloys, part I: general properties and modeling of single crystals”, *Mechanics of Materials* **38**, 391–429 (2006). (p. 15)

[52] E. Patoor, A. Eberhardt, and M. Berveiller, “Thermomechanical behaviour of shape memory alloy”, *Archives of Mechanics* **40**, 775–794 (1988). (p. 15)

[53] E. Patoor, A. Eberhardt, and M. Berveiller, “Micromechanical modelling of the shape behavior”, *Proceedings of ASME International Congress and Exposition* **189**, 23–37 (1994). (p. 15)

[54] K. Tanaka and S. Nagaki, “A thermomechanical description of materials with internal variables in the process of phase transitions”, *Ingenieur Archiv* **51**, 287–299 (1982). (p. 15)

[55] K. Tanaka, “A thermomechanical sketch of shape memory effect: one-dimensional tensile behavior”, *Res Mechanica* **18**, 251–263 (1986). (p. 15, 59)

[56] S. Seelecke and I. Mueller, “Shape memory alloy actuators in smart structures: Modeling and simulation”, *ASME. Applied Mechanical Reviews* **57**, 23–46 (2004). (p. 15)

[57] M. Forbes, J. Goulay, and M. Desmulliez, “Optically interconnected electronic chips: a tutorial and review of the technology”, *Electronics and Communication Engineering Journal* **13**, 221–232 (2001). (p. 17)

[58] R. A. Soref and B. R. Bennet, “Electrooptical effects in silicon”, *IEEE Journal of Quantum Electronics* **23**, 123–129 (1987). (p. 17)

[59] C. K. Tang and G. T. Reed, “Highly efficient optical phase modulator in SOI waveguides”, *Electronics Letters* **31**, 451–452 (1995). (p. 17)

[60] A. Liu, A. Lia, R. Jones, L. Liao, D. Samara-Rubio, D. Rubin, O. Cohen, R. Nicolaescu, and M. Paniccia, “A high-speed silicon optical modulator based on a metal-oxidesemiconductor capacitor”, *Nature* **427**, 651–618 (2004). (p. 17)

[61] A. Sciuto, S. Libertino, A. Alessandria, S. Coffa, and G. Coppola, “Design, fabrication, and testing of an integrated Si-based light modulator”, *Journal of Lightwave Technology* **21**, 228–235 (2003). (p. 17)

[62] Q. Xu, S. Manipatruni, B. Schmidt, J. Shakya, and M. Lipson, “12.5 Gbit/s carrier-injection-based silicon micro-ring silicon modulators”, *Optics Express* **15**, 430–436 (2007). (p. 18)

[63] Q. Xu, B. Schmidt, J. Shakya, and M. Lipson, “Cascaded silicon micro-ring modulators for WDM optical interconnection”, *Optics Express* **14**, 9430–9435 (2006). (p. 18)

[64] F. Chollet, "Devices Based on Co-Integrated MEMS Actuators and Optical Waveguide: A Review", *Micromachines* **7** (2016). (p. 18)

[65] K. Hogari and T. Matsumoto, "Electrostatically driven micromechanical 2x2 optical switch", *Applied Optics* **30**, 1253–1257 (1991). (p. 18)

[66] H. Bezzaoui and E. Voges, "Integrated optics combined with micromechanics on silicon", *Sensors and Actuators A: Physical* **29**, 219–223 (1991). (p. 18)

[67] Y. Kim, M. Allen, and N. Hartman, "Micromechanically based integrated optic modulators and switches", *SPIE Proceedings* **1793**, 183–189 (1992). (p. 18)

[68] W. Lukosz and P. Pliska, "Integrated optical interferometer as a light modulator and microphone", *Sensors and Actuators A: Physical* **26**, 337–340 (1991). (p. 18)

[69] D. Auernhammer, M. Kohl, B. Krevet, and M. Ohtsuka, "Intrinsic position sensing of a Ni-Mn-Ga microactuator", *Smart Materials and Structures* **18**, 265–271 (2009). (p. 19)

[70] E. Abbe, "The Relation of Aperture and Power in the Microscope (continued)", *Journal of the Royal Microscopical Society* **2**, 460–473 (1882). (p. 19)

[71] E. Abbe, "Beitrage zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung", *Archiv fur mikroskopische Anatomie* **9**, 413–468 (1873). (p. 19)

[72] M. Rothschild, "A roadmap for optical lithography", *Optics and Photonics News* **21**, 26–31 (2010). (p. 20)

[73] D. P. Sanders, "Advances in patterning materials for 193 nm immersion lithography", *Chemical Reviews* **110**, 321–360 (2010). (p. 20)

[74] B. Wu, "Extreme ultraviolet lithography: A review", *Journal of Vacuum Science and Technology B* **25**, 1743 (2007). (p. 20)

[75] R. Collins and B. Blott, "The adsorption and nucleation of zirconium on tungsten field emitters", *Surface Science* **10**, 349–368 (1968). (p. 20)

[76] Z. Szczudlo, A. Ciszewski, and Y. B. Losovyj, "Field electron emission study of Ti and Hf adsorption layers on W", *Applied Surface Science* **174**, 138–147 (2001). (p. 20)

[77] A. J. Speth, A. D. Wilson, A. Kern, and T. H. P. Chang, "Electronbeam lithography using vector scan techniques", *Journal of Vacuum Science and Technology* **12**, 1235–1239 (1975). (p. 21)

[78] P. Rai-Choudhury, *Handbook of microlithography, micromachining and microfabrication* (SPIE Optical Engineering Press, 1997). (p. 22)

[79] A. J. Antolak and W. Williamson, "Electron backscattering from bulk materials", *Journal of Applied Physics* **58**, 526–534 (1985). (p. 22)

[80] M. Parikh, "Corrections to proximity effects in electron beam lithography. I. Theory", *Journal of Applied Physics* **50**, 4371–4377 (1979). (p. 22)

[81] S. Y. Lee and B. D. Cook, "PYRAMID a hierarchical, rule-based approach toward proximity effect correction. I. Exposure estimation", *IEEE Transactions on Semiconductor Manufacturing* **11**, 108–116 (1998). (p. 22)

[82] G. Owen and P. Rissmann, "Proximity effect correction for electron beam lithography by equalization of background dose", *Journal of Applied Physics* **54**, 3573–3581 (1983). (p. 22)

[83] A. M. Barklund and H. O. Blom, "Influence of different etching mechanisms on the angular dependence of silicon nitride etching", *Journal of Vacuum Science & Technology A* **11**, 1226–1229 (1993). (p. 24)

[84] S. Franssila, *Introduction to Microfabrication* (John Wiley and Sons, 2010). (p. 24)

[85] F. Larmer and A. Schilp, "Verfahren zum anisotropen Ätzen von Silicium", *Patent DE 4241045* (May 1994). (p. 26)

[86] X. Mallhaoui, R. Dussart, T. Tillocher, P. Lefaucheux, P. Ranson, M. Boufnichel, and L. J. Overzet, "SiOxFy passivation layer in silicon cryoetching", *Journal of Applied Physics* **98**, 104 901 (2005). (p. 26)

[87] M. D. Henryy, *ICP Etching of Silicon for Micro and Nanoscale Devices* (Passadena, California, California Institute of Technology, Diss., 2010). (p. 26)

[88] L. Y. Tsou, "Highly Selective Reactive Ion Etching of Polysilicon with Hydrogen Bromide", *Journal of the Electrochemical Society* **136**, 3003–3006 (1989). (p. 26)

[89] K. Itoh, K. Iwameji, and Y. Sasaki, "Fabrication of Thin Silicon Wires by Anisotropic Wet Etching of SOI Structures", *Japanese Journal of Applied Physics* **30**, L1605 (1991). (p. 26)

[90] C. W. Dyck, J. H. Smith, S. L. Miller, E. M. Russick, and C. L. J. Adkins, "Supercritical carbon dioxide solvent extraction from surfacemicromachined micromechanical structures", *Micromachining and Microfabrication Process Technology II* **SPIE**, 2879 (1996). (p. 28)

[91] K. S. Harsha, "Principles of Physical Vapor Deposition of Thin Films", *Elsevier* **1** (2006). (p. 29)

[92] M. Ohring, "Materials Science of Thin Films - Deposition and Structure", *Academic Press* **2** (2002). (p. 31)

[93] F. Lambrecht, C. Lay, I. R. Aseguinolaza, V. Chernenko, and M. Kohl, "Ni-Mn-Ga/Si Shape Memory Bimorph Nanoactuation", *Shape Memory and Superelasticity* **2**, 347–359 (2016). (p. 55, 62, 63, 64, 65, 66, 88)

[94] F. Albertini, F. Canepa, S. Cirafici, E. A. Franceschi, M. Napoletano, A. Paoluzi, L. Pareti, and M. Solzi, “Composition dependence of magnetic and magnetothermal properties of Ni-Mn-Ga shape memory alloys”, *Journal of Magnetism and Magnetic Materials* **272-276**, 2111–2112 (2004). (p. 57)

[95] M. Kohl, A. Agarwal, V. A. Chernenko, M. Ohtsuka, and K. Seemann, “Shape memory effect and magnetostriction in polycrystalline Ni-Mn-Ga thin film microactuators”, *Materials Science and Engineering: A* **438-440**, 940–943 (2006). (p. 58)

[96] M. Ohtsuka, M. Matsumoto, and K. Itagaki, “Mechanical and shape memory properties of ferromagnetic Ni2MnGa sputter-deposited films”, *Journal de Physique IV France* **112**, 899–902 (2003). (p. 58)

[97] H. Watanabe, N. Yamada, and M. Okaji, “Linear Thermal Expansion Coefficient of Silicon from 293 to 1000 K”, *International Journal of Thermophysics* **25**, 221–236 (2004). (p. 58)

[98] I. R. Aseguinolaza, I. Reyes-Salazar, A. V. Svalov, K. Wilson, W. B. Knowlton, P. Müllner, J. M. Barandiarán, E. Villa, and V. A. Chernenko, “Transformation volume strain in Ni-Mn-Ga thin films”, *Applied Physics Letters* **101**, 241 912 (2012). (p. 58)

[99] I. R. Aseguinolaza, I. Orue, A. V. Svalov, K. Wilson, P. Müllner, J. M. Barandiarán, and V. A. Chernenko, “Martensitic transformation in Ni-Mn-Ga/Si(100) thin films”, *Thin Solid Films* **558**, 449–454 (2014). (p. 58)

[100] J. H. Jeong, S. H. Chung, S. H. Lee, and D. Kwon, “Evaluation of elastic properties and temperature effects in Si thin films using an electrostatic microresonator”, *Journal of Microelectromechanical Systems* **12**, 524–530 (2003). (p. 58)

[101] I. Aaltio, O. Söderberg, M. Friman, I. Glavatskyy, Y. Ge, N. Glavatska, and S. P. Hannula, “Determining the liquidus and ordering temperatures of the ternary Ni-Mn-Ga and quaternary Ni-Mn-Ga-Fe/Cu alloys”, *ESOMAT 2009* **04001** (2009). (p. 64)

[102] J. Rösler, H. Harders, and M. Bäker, *Mechanisches Verhalten der Werkstoffe* (Teubner, 2006). (p. 64)

[103] C. J. Glassbrenner and G. A. Slack, “Thermal Conductivity of Silicon and Germanium from 3K to the Melting Point”, *Physical Review* **134**, A1058–A1069 (1964). (p. 70, 82)

[104] C. G. Montgomery, R. H. Dicke, and E. M. Purcell, *Principles of microwave circuits*, no. 25 (IET, 1948). (p. 77)

[105] F. Lambrecht, I. R. Aseguinolaza, V. Chernenko, and M. Kohl, “Integrated SMA-based NEMS actuator for optical switching”, *Proceedings IEEE International Conference of Micro Electro Mechanical Systems (MEMS)* **29**, 79–82 (2016). (p. 78, 105)

[106] M. Asheghi, M. N. Touzelbaev, K. E. Goodson, Y. K. Leung, and S. S. Wong, "Temperature-Dependent Thermal Conductivity of Single-Crystal Silicon Layers in SOI Substrates", *Journal of Heat Transfer* **120**, 30–36 (1998). (p. 89)

[107] D. Li, Y. Wu, P. Kim, L. Shi, P. Yang, and A. Majumdar, "Thermal conductivity of individual silicon nanowires", *Applied Physics Letters* **83**, 2934–2936 (2003). (p. 89)

[108] Y. S. Ju, "Phonon heat transport in silicon nanostructures", *Applied Physics Letters* **87**, 153 106 (2005). (p. 89)

[109] R. Fechner, A. Muslija, and M. Kohl, "A micro test platform for in-situ mechanical and electrical characterization of nanostructured multiferroic materials", *Proceedings of the Micro- and Nano-Engineering Conference (MNE)* **accepted** (2016). (p. 110)