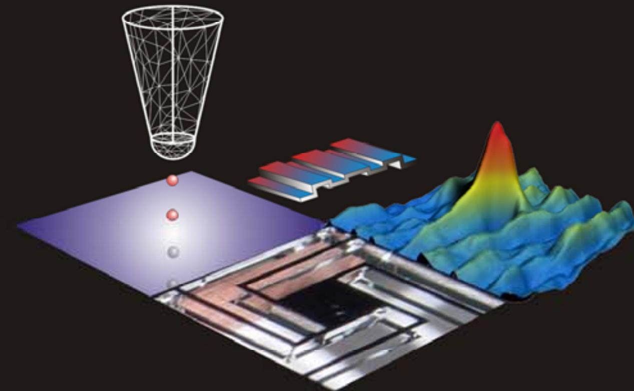
ISBN 978-90-365-2782-8

Capacitive MEMS-based sensors: Thermo-mechanical stability and charge trapping

Martin van Essen

Capacitive MEMS-based sensors: Thermo-mechanical stability and charge trapping



Martin van Essen

Uitnodiging

Graag nodig ik u uit voor het bijwonen van de openbare verdediging van mijn proefschrift, getiteld

Capacitive MEMS-based sensors: Thermo-mechanical stability and charge trapping

op vrijdag 23 januari 2009 om 13:15 uur in zaal 2 van gebouw De Spiegel van de Universiteit Twente, Enschede

Voorafgaand aan de verdediging geef ik om 13:00 uur een korte toelichting over mijn onderzoek.

Aansluitend aan de promotieplechtigheid is er een receptie ter plaatse.

Tenslotte heet ik u graag vanaf 21:00 uur welkom op het feest in *Café De Kater,* Oude Markt 5, Enschede

Martin van Essen
Boterstraat 6
7545 HJ Enschede
+31 6-14208071
mcvanessen@gmail.com

Paranimfen:

Johannes Pleikies
Bart Ormel

Capacitive MEMS-based sensors:

Thermo-mechanical stability and charge trapping

Ph.D. committee

Chairman and secretary

Prof. Dr. Ir. H.J.W. Zandvliet	University of Twente

Supervisor

Prof. Dr. H. Rogalla	University of Twente

Assistent supervisor

Dr. Ir. J. Flokstra	University of Twente

Members

Prof. Dr. Ir. M. Verhaegen	Delft University of Technology
Dr. J. Sesé	University of Zaragoza
Prof. Dr. Ir. A. de Boer	University of Twente
Prof. Dr. Ir. B. Poelsema	University of Twente
Prof. Dr. ing. B. van Eijk	University of Twente

Cover: South: MEMS (top) structure with double springs; West: Stylistic Finite Element mesh of a conducting AFM tip, string of mirror charges; North: Schematic MEMS structure with temperature gradient; East: Contact potential of a local trapped charge, measured by conducting AFM.

The research described in this thesis was performed in the Faculty of Science and Technology and the MESA+ Institute for Nanotechnology at the University of Twente. The project is part of the research program of the Dutch Foundation for Fundamental Research on Matter (FOM), financially supported by the Netherlands Organization for Scientific Research (NOW), and Philips Research.

M.C. van Essen
Capacitive MEMS-based sensors:
Thermo-mechanical stability and charge trapping
Ph.D. thesis University of Twente, Enschede, The Netherlands.

ISBN: 978-90-365-2782-8

Printed by PrintPartners Ipskamp, The Netherlands

© M.C. van Essen, 2009

CAPACITIVE MEMS-BASED SENSORS: THERMO-MECHANICAL STABILITY AND CHARGE TRAPPING

PROEFSCHRIFT

ter verkrijging van
de graad van doctor aan de Universiteit Twente,
op gezag van de rector magnificus,
prof. dr. H. Brinksma,
volgens besluit van het College voor Promoties
in het openbaar te verdedigen
op vrijdag 23 januari 2009 om 13.15 uur

door

Martin Cornelis van Essen

geboren op 14 juni 1980

te Middelburg

Dit proefschrift is goedgekeurd door:

Prof. Dr. H. Rogalla (promotor)
Dr. Ir. J. Flokstra (assistent-promotor)

“There are only two ways to live your life.
One is as though nothing is a miracle.
The other is as though everything is a miracle.”

Albert Einstein

CONTENTS

1	**Introduction**	**1**
1.1 Micro Electro Mechanical Systems (MEMS)		**2**
1.2 Capacitive MEMS		**6**
1.3 Motivation and addressed questions		**7**
1.4 Outline		**8**
2	**MEMS Theory**	**11**
2.1 Capacitance – voltage relation		**12**
	2.1.1 Energy and forces in capacitive MEMS structures	12
	2.1.2 Pull-in voltage and electromechanical coupling	15
	2.1.3 Charge control	17
	2.1.4 Dielectric layer	19
2.2 Beam deflection theory and stress		**23**
	2.2.1 Beam equation	23
	2.2.2 Electric load	24
	2.2.3 Bridge instability	27
	2.2.4 Beam with partial load and tensile stress	29
	2.2.5 Flank loaded beam	32
2.3 Built-in voltage and parasitic charges		**33**
	2.3.1 Sources of built-in voltage	34
	2.3.2 Types of charges in dielectric layers	40
	2.3.3 Charge transport	41
	2.3.4 Summary	45
2.4 Electromechanical response		**45**
	2.4.1 Equivalent circuits	46
	2.4.2 Characteristic frequencies and parameters	50
	2.4.3 Influence of bias voltage	53
	2.4.4 Damping	54
2.5 Considered side-effects		**54**
	2.5.1 Non-parallel electrodes	55
	2.5.2 Casimir effect	58
	2.5.3 Fringe forces, charge distribution	60
	2.5.4 Breakdown	61
3	**Measurements on EMMA samples**	**65**
3.1 Measurement principles and setup		**66**
	3.1.1 Impedance measurement	66
	3.1.2 Insert, flow cryostat	68
3.2 Silicon nitride bridges		**68**
	3.2.1 Devices	68
	3.2.2 Thermal dependence	70
	3.2.3 Built-in voltage in silicon nitride bridges	72
	3.2.4 $C(V)$ – asymmetry	72

CONTENTS

3.3 Aluminum bridges **76**
3.3.1 Devices 76
3.3.2 Influence of rapidly changing temperature on aluminum bridge characteristics 77
3.3.3 Influence of varied stabilized temperature on aluminum bridge characteristics 79
3.3.4 Built-in voltage 82
3.4 Conclusions and discussion **85**

4 Thermo-mechanically stable MEMS **87**

4.1 Design and fabrication **88**
4.1.1 General design description and considerations 88
4.1.2 Top structure design 89
4.1.3 Fabrication process top structures 92
4.1.4 Wet-etching of silicon 93
4.1.5 Wet-etching of top structures 94
4.1.5.1 Etch path 94
4.1.5.2 Etch rate monitoring 96
4.1.5.3 Prevention of corner under-etching 98
4.1.5.4 Results of etching process 99
4.1.6 Bottom mask design 100
4.1.7 Fabrication process bottom structures 101
4.2 Test top structure **102**
4.2.1 Experiment setup, phase shift interferometry 102
4.2.2 Results 103
4.2.3 Spring constant anomaly 105
4.3 Mechanical sample characterization **105**
4.3.1 Sample description 106
4.3.2 Single peak characteristics 107
4.3.3 Bias dependent resonance 108
4.3.4 Phase error 112
4.3.5 Damping 112
4.4 Temperature dependent experiments on device-level charging Effects through $C(V)$ - curves **114**
4.4.1 Temperature dependent sample characterization 115
4.4.2 Built-in voltage 117
4.5 High vacuum experiments **119**
4.5.1 Sample characterization 120
4.5.2 Built-in-voltage 120
4.6 Conclusions **121**

CONTENTS

5 Conducting AFM **123**

5.1 Principles of AFM **124**
5.1.1 The family of Scanning Probe Microscopy 124
5.1.2 Scanning Tunneling Microscopy 125
5.1.3 Atomic Force Microscopy 125
5.1.4 Frequency shift in FM AFM 130
5.2 Charge imaging on thin oxide films **132**
5.3 The electrostatic problem and earlier models **134**
5.3.1 The electrostatic problem 135
5.3.2 Earlier models 136
5.4 The Multi-Mirror Model **138**
5.4.1 Imaging in tip 138
5.4.2 Force and gradient calculation: convergence 139
5.4.3 Field line plots, electrostatic repulsion 141
5.4.4 Electrostatic repulsion 142
5.4.5 Critical behavior 143
5.4.6 Resolution depth and magnitude 144
5.4.6.1 Bias variation 144
5.4.6.2 Geometrically varied measurements 147
5.5 Finite Element Modeling **150**
5.5.1 Principles of Finite Element Modeling 150
5.5.2 Determination of force gradient 152
5.5.3 Conical tip 153
5.6 Model comparison **155**
5.6.1 Interaction parameters 155
5.6.2 Dependence on charge depth 156
5.7 Relation to empiricism **157**
5.8 Conclusions and discussion **160**

6 Trapped charges limiting capacitive MEMS-based sensors **163**

6.1 Charge trapping in a gravity gradiometer **164**
6.2 Charge trapping in a RF power sensor **176**
6.3 Discussion and conclusions **179**

Appendix 1 Fundamental mirroring principles and potentials **181**

A1.1 Basic mirroring principles **182**
A1.2 Potential functions **190**

Appendix 2 AFM electrostatics: Miscellaneous **197**

A2.1 Electrostatic energy of a system of conductors and reflection potential **197**
A2.2 Approximations for the electrostatic energy in the AFM system **201**
A2.3 Dielectric extensions **210**
A2.4 Calculation methods **214**

CONTENTS

Summary 217

Samenvatting (in dutch) 221

Dankwoord 227

CHAPTER 1

INTRODUCTION

1.1 MICRO ELECTRO MECHANICAL SYSTEMS (MEMS)

Miniaturization pervades present-day technology like once ether was thought to pervade all substances. Because of the 'general desire' for ever increasing functionality of compact pieces of technology, components have gradually been scaled down by factors of 1000 or much more, compared to equivalent components from the younger days of our grandparents. Machinery of fabulous accuracy and precision enables the fabrication of complete devices, in which relevant length scales recede far below the observation threshold of the human eye. Current processor industry involves transistors measuring a mere 45 nanometer, which means that a thousand or more are required to span the diameter of a human hair.

The history of miniaturization of *moving* parts can be followed along centuries of developing craftsmanship in the field of mechanical clocks, see Figs. 1.1A-F. In a time span of 1000 years, mechanical parts have been downsized by a factor of 1000. Let us briefly review a few highlights, starting with the water clock, or clepsydra ("stealer of water"), built by Su Song in 1088 [1,2], see Fig. 1A. The complete construction measures 40 feet (12 meters). The cited manuscript from 1092 A.D. is the earliest known mention of an endless chain transmission, which was part of the mechanics that converted the movement of falling water into indicating and striking time. In Europe, the oldest mentions of mechanical clocks are from the beginning of the 14th century. Because these clocks were driven by slowly falling weights, the earliest specimens were feasible in church towers only. The Salisbury Cathedral in England has the oldest known still working clock, from 1386 [3], see Fig. 1.1B. The mechanical part measures just over a cubic meter. Fig. 1.1C shows Christiaan Huygen's design of a pendulum clock [4]. He was the first to realize a working device in 1656. This achieved unprecedented accuracy within home-suitable size. Meanwhile, the technique with a wound spring had already taken a flight. It is claimed that the oldest surviving clock driven by a wound spring had been given to Herzog Philipps des Guten von Burgund around 1430. It can be seen in the Germanisches Nationalmuseum in Nuremberg, Germany [5]. The wound spring enabled drastic miniaturization. Peter Henlein fabricated a portable 'Taschenuhr' in the early 16th century. Classic pocket watches appear in the 17th century. Fig. 1.1D shows the famous "Marie Antoinette" by Breguet [6], inventor of the *tourbillion*. The Marie-Antoinette is perhaps the most expensive watch ever made. This masterpiece took 44 years to complete and consists of more than 800 hand-made pieces, contained in an 18K golden case.

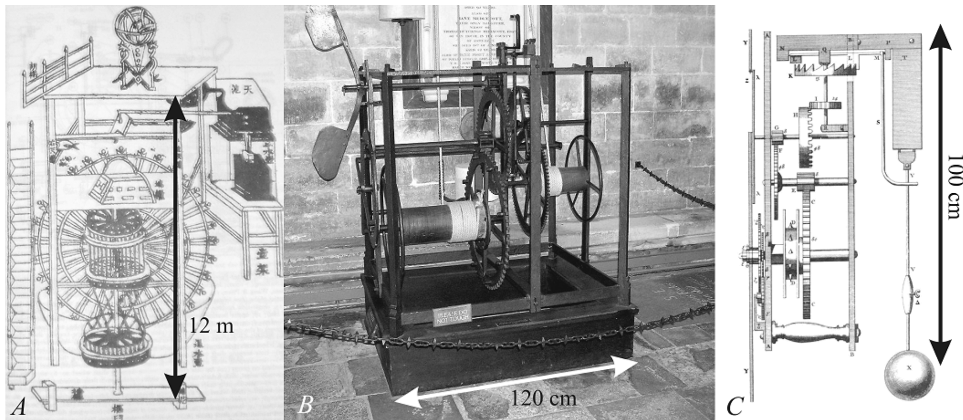


Fig. 1.1A. Chinese water clock by Su Song
Fig. 1.1B. The Salisbury Cathedral clock
Fig. 1.1C. Christiaan Huygens's design of a pendulum clock

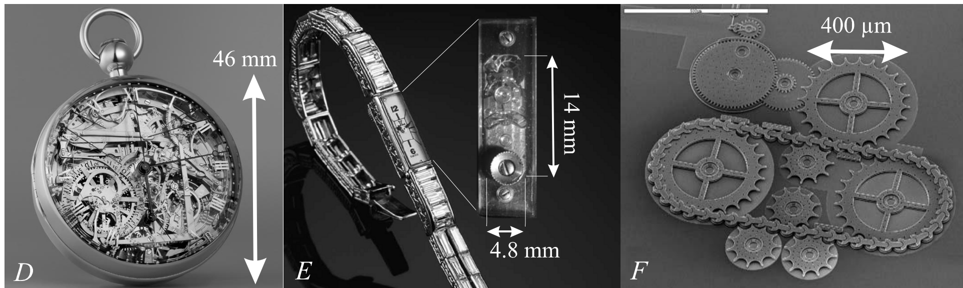


Fig. 1.1D. Breguet's "Marie-Antoinette" pocket watch (1827)
Fig. 1.1E. Smallest wristwatch ever, Calibre 101 by Jaeger – LeCoultre (1929).
Fig. 1.1F. MEMS transmission chain

In the early 19th century, progressed miniaturization shrunk mechanical clocks to sizes such that they could be worn around the wrist. Those wristwatches became popular only after circa 1900. The smallest mechanical wristwatch ever is the Calibre 101 by Jaeger – LeCoultre (1929) [7], which measures a mere $14 \times 4.8 \times 3.4$ mm^3, weighing 0.9 gram. On submillimeter scale, mechanical transmission becomes increasingly challenging. Shown are miniaturized silicon gears driven by a chain transmission [8], basically a miniature version of what Su Song used for his water clock. The diameter of largest gear is about 400 µm. It will only be matter of time until we have to read the time through a microscope.

The omnipresence of electronics has not forgotten to imprint its finger on the area of downscaled mechanical worlds. The merging of both miniaturized worlds resulted in a worldwide and rich flora of Micro Electro Mechanical Systems (MEMS), suitable for having assigned numerous applications.

From a purely scientific point of view, MEMS are interesting because it unites many topics of physics: Mechanics, electrostatics, fluid dynamics, optics and so on. Increased downscaling justified coining the category of NEMS, where 'N' associates the currently much coveted term 'nano' with the field of microscopic (or perhaps 'nanoscopic') machinery. This downscaling itself gives rise to issues peculiar for MEMS.

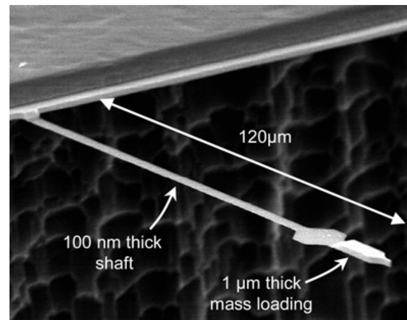


Fig. 1.2 IBM microcantilever used for Magnetic Resonance Force Microscopy (MRFM)

Quantities associated with the cubic dimension, like mass, change much more rapidly than quantities associated with surface and length. Cantilevers stretched out at long aspect ratios show no significant weight-induced deflection, whereas a model upscaled towards everyday life would not sustain the extra weight without further measures. The cantilever shown in Fig. 1.2 [9] has a length of 120 µm and a thickness of 100 nm. Accounts of what consequences downscaling has for various physical interactions and quantities are given in [10] and [11].
MEMS can be classified along various criteria. One is whether it concerns an actuating device (such as a micro pump) or a sensing device. Sensors equipped with a feed-back mechanism could be said to combine both. A second criterion concerns the physical principle that characterizes the device: for an actuating device this is the driving mechanism and for a sensor this is the read-out mechanism.
Within the world of miniaturized devices (microsystems), MEMS form a very pronounced and pluriform world. Fig. 1.3 could be a coarse map of the world of Microsystems plus the topology of some of its subsets, amongst others MEMS. Categorization of microdevices along this latter criterion is somewhat arbitrary, because many devices encompass various physical principles. A couple of them are summed below.

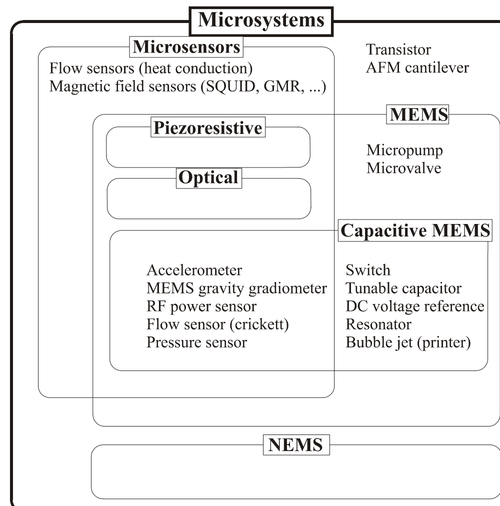


Fig. 1.3 Schematic of Microsystems, and MEMS with several sub-categories.

Piezo-electricity

If certain crystals are put under pressure (the Greek πιέζω means “to press, to squeeze”), they generate charge on the surface, proportional to the pressure. A literally household application of this *piezo-electric* effect is a lighter, as an alternative for flint stone. A small hammer strikes the crystal, which causes it to produce a spark. Lead zirconate titanate ($PbZrTiO_3$ or “PZT”) is a common piezo(electric) crystal, but also ordinary quartz. The charge separation leads to a voltage over the crystal that can be read out to be a measure for the applied pressure. In the opposite way, a voltage over the crystal will expand the crystal. This can be exploited to displace elements in actuating devices. Fig. 1.4 shows a schematic of a piezoelectrically driven micropump [12].

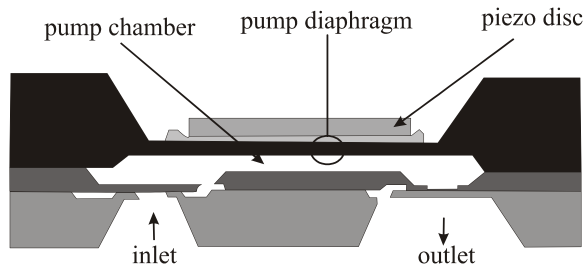


Fig. 1.4 Piezoelectrically driven micropump (Reproduced from [12])

Nowadays, ink jet printers make use of piezoelectric micropumps to fire tiny ink droplets on paper. Because piezocrystals expand only a tiny bit upon an applied voltage, and because of the very linear relation between the two, these crystals are eminently suitable for highly precise positioning. Sub-Ångström positioning is essential in for example Atomic Force Microscopy.

Piezo-resistive effect

The resistivity of some metals and especially semiconductors varies (linearly) with strain. Silicon is a noteworthy example. The piezo-resistive effect finds its most common use in strain gauges. A film of piezoresistive material is applied on a flexible geometry. If this geometry is deformed, the change in resistance in the film due to the strain is a measure for the pressure that deforms the geometry. Note that this physical principle can only be used for passive, sensing applications, unlike piezo-electricity, which can be used in an active sense as well.

Thermal conductivity

There are several methods available for sensing properties of fluid flows. These properties include amongst others mass flow, flow velocity, heat capacity. In [13] a sensor is demonstrated based on heat pulses traveling upstream and downstream. This relies on the thermal conductivity of the fluid and of the elements picking up the heat pulse. A displacement sensing device based on thermal conductivity is given in [14].

Optics

Very high displacement resolution can be obtained by optical readout of the sensor. Loh [15] used an interferometer to readout a cantilever displacement with a resolution of 3×10^{-3} Å. Using an array of interdigitated cantilevers, he developed an accelerometer with a resolution of 40 ng/√Hz (1 ng = 9.8×10^{-9} m/s²).
A different implantation of optic readout is reflection of a laser beam. Sharp timing of laser pulses hitting reflectors placed on the moon revealed that our natural satellite recedes from the earth at 3.8 cm/year. More into the world of microsensors is monitoring the position of a reflected laser beam on a position sensitive detector (PSD) to accurately determine the frequency of a vibrating Atomic Force Microscope cantilever (see also Chapter 5). Admittedly, also in the latter case the optical read-out part is not yet micro-machined itself.
The relatively young field of Micro Opto Electro Mechanical Systems (MOEMS) concerns with integrating more and more optic elements on the microscale. In actuating sense for example optical switches, and for sensing purposes the possibilities of micromachined optically pumped magnetometers are explored [16,17]. Optical excitation of atoms in a micromachined gas cell enables to obtain a very sensitive magnetic field sensor which due to its small dimensions is highly desired for space applications.

Electron Tunneling

Perhaps the finest resolution in displacement sensing can be achieved by electron tunneling, in the order of 10^{-4} Å/√Hz. This extremely high resolution is due to the exponential decay of tunneling current with distance. This way, Scanning Tunneling Microscope can be used to detect height differences in surfaces that measure a small fraction of an atom. Liu [16] used tunneling to readout the displacement of a moving element in an accelerometer to achieve 20 ng/√Hz.

Inductive coupling

Superconducting Quantum Interference Devices (SQUIDs) [18,19] are the most sensitive magnetometers around, capable of detecting fields weaker than 1 fT. In particular, they are very suitable to sense with extreme precision the magnetic fields produced by small currents running underneath the sensor. This way they are very suitable for displacement readout. The MiniGRAIL project [20,21 and references therein] aims to detect gravitational waves through tiny deformations of a massive sphere. The expected (vibrating) deformations are in the order of 10^{-20} meter. These deformations are amplified with two orders of magnitude, which means that the SQUIDs still need to notice deflections of 1 atto-meter, about a thousandth of a proton size.
Another example of inductive coupling is flux concentrators to locally amplify magnetic fields, which can then be sensed by a Hall element [22] or a GMR sensor [23].

Capacitive coupling

Capacitive MEMS are employed as actuating as well as sensing devices. Capacitive displacement sensing does not need cooling, unlike superconducting inductive displacement sensing. Still, it can still be done very accurately by AC impedance measurements. The capacitive coupling between two electrodes can also be exploited to

establish movement of flexible components in actuating devices. More on capacitive MEMS is given in the next section.

Resonance

Resonating elements are applied in very accurate measurements of numerous physical properties. For example, Non-contact AFM (see chapter 5) is able to scan many kinds of interactions between the tip of an AFM and a surface, through variations of the vibration frequency of a cantilever. These frequency variations can be detected with high precision. Of course, the resonating element itself must again be read-out, in order to transform the movement into an electric signal.

1.2 CAPACITIVE MEMS

One continent in the world of MEMS is centralized around the capacitive coupling between two electrodes: Capacitive MEMS. Its simplest appearance is the parallel plate configuration, see Fig. 1.5:

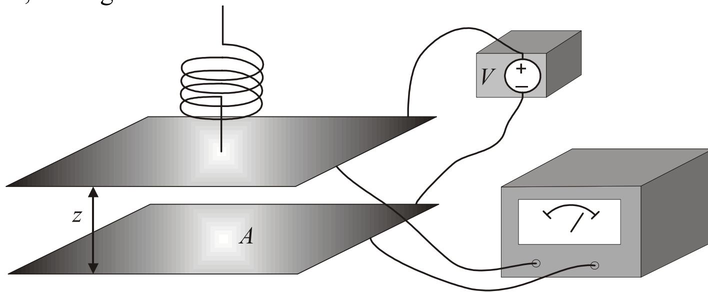


Fig. 1.5 Parallel plate capacitive MEMS. The drawing includes measurement electronics (bottom right) and a possible voltage control.

The capacitance of this configuration is proportional to the reciprocal separation distance, see Eq. (1.1):

$$C = \frac{\varepsilon_0 A}{z} . \tag{1.1}$$

Here A is the capacitive area and z the distance between the plates. ε_0 is the permittivity of vacuum, a constant of nature, with the value $8.854 \cdot 10^{-12}$ F/m (Farad per meter). If one electrode is attached to a flexible suspension, this separation distance and hence the capacitance becomes variable. This principle can be exploited to detect tiny displacements. A sensitivity of better than 10^{-3}Å/√Hz has been achieved [24]. Examples are capacitive readout of AFM cantilevers, inclination sensors and flow sensors, for instance the array of “cricket hairs”, see Fig. 1.6 [25, 26]: a fluid flow bends the hairs, which displaces capacitively coupled elements at the base of the hairs. The resulting change in capacitance is a measure for the strength of the flow. The four ‘feet’ make the sensor also sensitive to the direction of the flow. Accelerometers exist already since the early days of capacitive MEMS; see Fig. 1.7 for a typical schematic [27]. An accelerometer measures the acceleration of a surrounding body (vehicle, satellite)

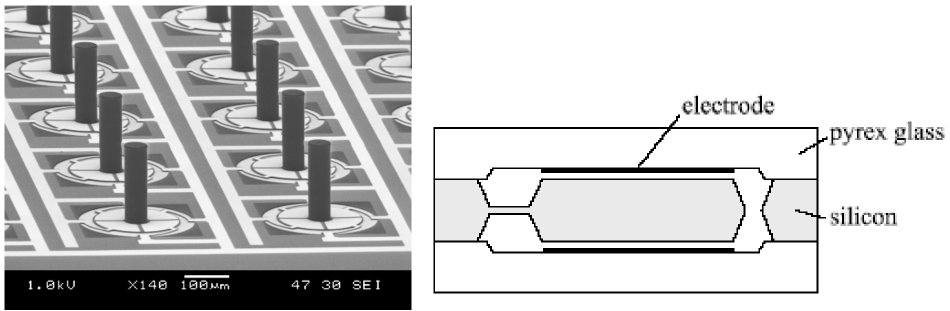


Fig. 1.6 (left) "Crickett hairs": an array of capacitive flow sensors.
Fig. 1.7 (right) Basic structure of a bulk micromachined accelerometer.

through the inertia of a suspended test mass. The mass contains electrodes that are capacitively coupled to fixed electrodes
During acceleration, the mass will be closer to one or the other fixed electrode, which results in a difference between the two capacitances. The development of accelerometers has nowadays become sufficiently mature to be well commercially active, for example by delivering sensors to example automotive industry. In metrology, the accelerometer finds important use in for instance the study of gravity fields, see Chapter 6.
Additionally, MEMS can be used as active components as well. If a voltage is applied across the electrodes, they will attract. For small voltages, the electric force will be balanced by the stiffness of the suspension. Resonators [28, 29] and tunable capacitors [30] make use of this property, as well as feedback designs: this way the sensing and actuation aspects of MEMS devices are combined. Feedback can be applied to keep the test mass of an accelerometer in its central position: the force (voltage), required to keep it there, is a measure of the experienced acceleration. An important advantage is that the operation range of the sensor is significantly enhanced, as contact between the electrodes is avoided. A final example of the application of voltage actuated MEMS is switches [30]: capacitive MEMS experience an instability for voltages above a certain threshold (see subsection 2.1.2): the spring force cannot balance the electric force anymore. Two electrodes are then used for attraction, bringing two different electrodes into a conductive contact. This way, very fast, miniaturized switches can be made.

1.3 MOTIVATION AND ADDRESSED QUESTIONS

Recent efforts in the development of high-precision capacitive MEMS-based sensors have involved challenging specifications concerning reliability and sensitivity. Depending on the specific design and application, there are always several factors limiting the performance of the device: mechanical and electric noise, challenges in and precision of fabrication. Sometimes it is not the device itself that forms the biggest limitations, but the readout-electronics. This thesis revolves around the often-encountered effect of charge trapping: unwanted (parasitic) electric charges, stuck in the bulk or at the surface of used dielectric materials, introduce additional forces that disturb the accuracy of the device. It is demonstrated in Chapter 6 for a few examples that very low charge concentrations are sufficient to degrade the sensor below the specifications. The fact that for none of these devices the design incorporates measures to minimize charge trapping

indicates that the influence of trapped charges is overlooked or at least underestimated. Initial concentrations may already be introduced by the fabrication process, and repeated operation causes further accumulation. While a sensor may be calibrated to compensate for an initial, stable distribution of parasitic charge, it is the hardly controllable instability (drift, hysteresis, noise) of charge accumulations during operation that provides extra challenges of meeting target specifications in the development of a particular sensor. This operation-induced accumulation is especially pronounced in micro-switches: each contact allows the exchange of significant loads of static charge, which after a certain number of cycles results in “sticking”: the switch does not open again after removal of the actuation voltage [31].

In this thesis the phenomenon of charge trapping is probed at two levels: device-level and fundamental level. It is investigated what influence parasitic charges have on the characteristics of capacitive MEMS-based sensors. These sensors sometimes have to operate under cryogenic circumstances, either intentionally cooled to reduce thermal noise, or the environmental conditions imply this, notably if employed in space. The corresponding question is how the influence of trapped charges changes when the environmental temperature is changed to the cryogenic regime. Can this change of influence from trapped charges be discerned from change of mechanical characteristics of the sensor? If not, what is needed to decouple these effects?

What processes guide the transport of charges through the material? What information can be retrieved about this from the two levels of research? In order to resolve the geometrical distribution of trapped charges in a dielectric layer, it is desirable if conducting AFM is able to resolve depth and magnitude of a localized charge. How can this be achieved?

The work presented in this thesis seeks to address these questions through analysis of measurements on capacitive MEMS sensing devices and a new developed model for the interaction of a conducting AFM tip with a localized charge, which is coupled to practice.

1.4 OUTLINE

This thesis can be divided in two main parts: MEMS-based research (chapters 2, 3, 4 and 6) and AFM-based research (chapter 5), with both parts in their own ways related to charge trapping.

Chapter 2 covers MEMS theory as far as found relevant in this thesis. In the first section the basic statics for general capacitive MEMS are provided. The pull-in effect is covered. Then a distinction between voltage control and charge control is made. Finally, the presence of dielectric layers in MEMS is included. Section 2 covers some beam deflection theory, important for the analysis of experimental work treated in Chapter 3. Then the influence of trapped charges, their origin and transport mechanisms are treated. There is a section devoted to theory used for the analysis of resonance measurements. The chapter finishes with the effects of non-parallel electrodes and the Casimir force.

Chapter 3 is devoted to experimental work performed on available RF MEMS devices, realized in a European project for the development of sensitive MEMS for metrology. Two classes of devices have been investigated for the effects of charge trapping at largely varying temperatures (4K-300K). The thermal variations turned out to be of dramatic influence on the performance or, rather, the characteristics of these devices. Deformation and build-up of thermo-mechanical stress affected capacitance and sensitivity to a large extent. For this reason, test structures were designed for better withstanding temperature

variations, which is the subject of Chapter 4. The designs and relevant considerations are provided there, as well as a detailed account of the novel fabrication process. After the structures passed initial tests with simultaneous interferometry and capacitance-voltage measurements, they were characterized in more detail, on which basis considerably improved thermal characteristics are demonstrated. Measurements of temperature dependent charge trapping effects are then discussed, together with some high vacuum tests. At this point, the MEMS part concludes.
Chapter 5 introduces the principles of Force Modulated Atomic Force Microscope. With this operation mode, the AFM cantilever is deliberately brought in oscillation close to its resonance frequency. Interactions between the tip and the sample modify the resonance frequency, which is kept constant during scanning. This technique enables on-situ study of trapped charges, now seen as *localized* entities, whereas in device-level study only their combined effects can be brought to light. Experimental work relevant of the research is discussed.
The larger part of this chapter is taken up by a theoretical account of the electrostatic interaction between a biased tip and sample. The latter consists of a dielectric layer on a metal electrode. Of special interest is a localized, parasitic charge in the dielectric. Models known from literature to describe the tip-sample interaction show some suspicious results and predictions. A new model has been developed that provides much different results; generally a dramatically larger interaction strength. On fundamental grounds, compared to finite element modeling and on the basis of empiricism the new model generally outperforms the existing models.
Chapter 6 concludes the main body of the thesis by considering what effects trapped charges can have on contemporary capacitive MEMS sensors. Considered here are an RF power sensor and a gravity gradiometer. It is demonstrated that demands on the performance of these sensors are such, that levels of charge trapping encountered typically in practice are clearly much higher than can be allowed. If the phenomenon of charge trapping is not considered or underestimated in the designs, performance demands will be impossible to meet.

REFERENCES

[1] Su Song, "Xinyi Xiangfayao" (1092)
[2] Needham, Joseph, "Science and Civilization in China: Volume 4, Physics and Physical Technology, Part 3, Civil Engineering and Nautics", Taipei: Caves Books, Ltd. (1986)
[3] http://www.salisburycathedral.org.uk/visitor.intro.php
[4] Chr. Huygens*, Horologium Oscillatorium sive de motu pendulorum* (1673)
[5] http://www.gnm.de/schausammlungen.html
[6] http://www.breguet.com
[7] Cologni Franco, Jaeger-leCoultre La Grande Maiso, Flammarion (2008)
[8] http://www.sandia.gov/media/NewsRel/NR2002/chain.htm
[9] H.J. Mamin et al., "Nuclear magnetic resonance imaging with 90-nm resolution", *Nature Nanotechnonoly* **2**, 301 (2007)
[10] M. J. Madou, "Fundamentals of microfabrication", 2nd edition, CRC Press, New York, USA (2002)
[11] M. Elwenspoek and R. Wiegerink, "Mechanical microsensors", Springer, Berlin, Germany (2001)

[12] R. Linnemann et al., “A self-priming and bubble-tolerant piezoelectric silicon micropump for liquids and gases”, *Proceedings of The Eleventh Annual International Workshop on Micro Electro Mechanical Systems*, Heidelberg, Germany (1998)
[13] J. van Kuijk et al., “Multi-parameter detection in fluid flows”, *Sensors and Actuators A* **46-47**, 369 (1995)
[14] Dauerstädt et al., “A new high temperature pressure sensor based on a thermal readout principle”, *Transducers*, pp. 525-530 (1999)
[15] N. C. Loh et al., “Sub-10 cm^3 interferometric accelerometer with nano-g resolution”, *Journal of Microelectromechanical Systems* **11**, No. 3, pp. 182 (2002)
[16] Schwindt et al., “Chip-scale atomic magnetometer”, *Applied Physics Letters* **85**, 6409 (2004)
[17] C. Liu and T. W. Kenny, “A high precision, wide-bandwidth micromachined tunneling accelerometer”, *Journal of Microelectromechanical Systems* **11** No. 3, (2001)
[18] J. R. Jaklevic, J. Lambe, A. H. Silver, and J. E. Mercereau, “Quantum interference effects in Josephson tunneling”, *Physics Review Letters* **12**, 159 (1964)
[19] J. M. Jaycox, and M. B. Ketchen, “Planar coupling scheme for ultra low noise dc SQUIDs”, *IEEE Trans. Magn.* **17**, 400 (1981)
[20] A. de Waard, L. Gottardi, J. van Houwelingen, A. Schumack, and G. Frossati, “MiniGRAIL, the first spherical detector”, *Classical and Quantum Gravity*, **20,** S144 (2003)
[21] J. Pleikies, O. Usenko, K.H. Kuit, J. Flokstra, A. de Waard and G. Frossati, “SQUID Developments for the Gravitational Wave Antenna MiniGRAIL”, *IEEE Trans. on Appl. Supercond.*, **17**, 764 (2007)
[22] F. Schmidt, S. Linzen, F. Schmidl, M. Mans, and P. Seidel, “Development of a thin-film high-temperature superconductor Hall magnetometer”, *Superconductor Science and Technology* **15**, 488 (2002)
[23] M. Pannetier, C. Fermon, G. Le Goff, J. Simola, and E. Kerr, "Femtotesla Magnetic Field Measurement with Magnetoresistive Sensors" **304**, 1648 (2004)
[24] X. Jiang, “An integrated surface micromachined capacitive lateral accelerometer with 2µG/√Hz resolution”, *Proceedings of Solid-State Sensor, Actuator and Microsystems Workshop*, Hilton Head Island, South Carolina, USA (2006)
[25] M. Dijkstra, J. van Baar et al., “Artificial sensory hairs based on the flow sensitive receptor hairs of crickets”, *Journal of Micromechanics and Microengineering* **15** S132 (2005)
[26] G. J. M. Krijnen et al., “MEMS based flow sensors as model systems for acoustic perception studies”, *Nanotechnology* **17** S84 (2006)
[27] F. Rudolf, “A micromechanical capacitive accelerometer with a two-point inertial-mass suspension”, *Sensors and Actuators* **4** 191 (1983)
[28] R. Legtenberg, and H. A. C. Tilmans “Electrostatically driven vacuum-encapsulated polysilicon resonators Part I. Design and fabrication”, *Sensors and Actuators A*, **45**, 57 (1994)
[29] H. A. C. Tilmans, and R. Legtenberg, “Electrostatically driven vacuum-encapsulated polysilicon resonators Part II. Theory and performance”, *Sensors and Actuators A*, **45**, 67 (1994)
[30] Th. G. S. M. Rijks et al., “MEMS Tunable capacitors and switches for RF applications”, *Proceedings of 24th International Conference on Microelectronics (MIEL),* Niš, Serbia-Montenegro (2004)
[31] M. van Spengen et al., “A comprehensive model to predict the charging and reliability of capacitive RF MEMS switches”, *Journal of Micromechanics and Microengineering* **14**, 514 (2004)

CHAPTER 2

MEMS THEORY

In this chapter some basic theory relevant in the context of capacitive MEMS will be discussed. Section 2.1 describes the mechanism of attraction of capacitively coupled electrodes. For small voltages, this force is quadratic to good approximation. The phenomenon of pull-in is covered. At some critical voltage the position of the flexible electrode will become unstable. Related to this is a short discussion of voltage control and charge control. As many practical devices involve dielectric layers, the effect of the presence of dielectrics on the so far established theory is digressed on. Section 2.2 is a treatment of the theory of the deflection of clamped-clamped beams, loaded with an electrostatic force. It will zoom in on how MEMS characteristics, especially the stiffness, depend on temperature. Section 2.3 will deal with how parasitic charges affect MEMS characteristics. The relation of parasitic charges with "built-in voltage" is laid. Various possible charge transport mechanisms through dielectrics are reviewed. The mechanical characterization of newly fabricated MEMS structures (Chapter 4) call for a coverage of resonance experiments in section 2.4. At last, in 2.5 a few effects (non-parallelity, Casimir force, fringe effect and breakdown) are shortly discussed.

2.1. CAPACITANCE – VOLTAGE RELATION

2.1.1 Energy and forces in capacitive MEMS structures

Capacitance C is defined as the proportionality constant between the amount of charge Q on a conductor and the potential at which the conductor is found:

$$Q = CV. \tag{2.1}$$

The capacitance indicates how much charge can be stored on the conductor given a certain potential. It is measured in *Farad*, which is a coulomb per volt. It is important to realize that the relation between potential and charge is always linear, irrespective of for example the shape of the conductor. It is worth stressing that capacitance is a purely geometric quantity, defined by sizes, shapes and dielectric constants if dielectrics are present, and does not intrinsically depend on electrostatic quantities such as charge, electric field and potential.
As potential is a relative concept, so is capacitance. It is possible though to define a potential and capacitance for a single conductor with respect to infinity, if potential is defined there, for example $V(|\mathbf{x}| \to \infty) \to 0$. The capacitance of a spherical conductor of radius R is $4\pi\varepsilon_0 R$ then[1]. If the sphere contains a charge Q, its potential is Q/C. In systems of many conductors, a capacitance between each pair of conductors can be defined.
In capacitive MEMS structures we have the usual configuration of two conductors, where the potential of one is defined with respect to the other conductor and thus the capacitance between them. Consider now a capacitor of arbitrary geometry with capacitance C and (intermediate) potential V. Now we take a small amount of charge dq from one electrode and transport it to the other electrode. In doing so, we have to overcome the potential difference v. This action will raise the potential by an amount $dv = dq/C$. We keep on transporting charge until a final amount of charge Q and a corresponding potential V is reached. The total electrostatic energy U_e stored in this capacitor thus becomes:

$$U_e = \int_0^Q v\,dq = \int_0^Q \frac{q}{C}\,dq = \frac{Q^2}{2C} = \frac{1}{2}CV^2 \tag{2.2}$$

Let us now consider a capacitive MEMS structure (Fig. 2.1) which consists of two capacitively coupled electrodes. The terminals are connected to a voltage source, maintaining the electrodes at a constant potential difference V, by storing an amount of charge Q on one of the electrodes and $-Q$ on the other. Peculiar of capacitive MEMS structures is that one of the electrodes has freedom of motion within a defined suspension, assumed of stiffness k which is constant for small displacements.

[1] Taking the earth's radius to be 6378 km, this means our planet's capacitance comes out at 710 µF, quite a 'household' value, as capacitances of this order are commonly used in power source stabilization in televisions and computers. Hence less than 9 electrons need to be scattered at each square meter to raise our planet's potential by one volt.

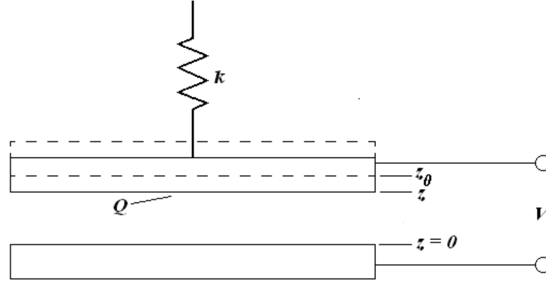


Fig. 2.1 Capacitive MEMS with stiffness k and a voltage V, which establishes a charge Q on the movable top electrode (and –Q on the bottom), causing the bridge to displace from the unactuated position z_0 to z.

Now if the top electrode is displaced a distance dz, the energy $U = U_e$ is changed by an amount of work done on the system dW_{in} minus work done by the system dW_{out}. The process takes place quasistatically, hence kinetic energy terms are neglected.

$$dU = dW_{in} - dW_{out}\ . \tag{2.3}$$

In this case, work is done by the voltage source. Under a constant voltage V, it transports an amount of charge dQ, contributing VdQ to dW_{in}. The change in mechanical energy dU_m through stretching of the spring may very well be subsumed in change of the internal energy dU, but in this case it is chosen to consider the electrostatic device to perform work dW_{out} on the spring[2]. Any other work done by the device, such as displacing fluids, would also be absorbed in this term. This is not considered here. The force on the spring F_s thus becomes:

$$F_S = \frac{dW_{out}}{dz} = -\frac{dU}{dz} + \frac{dW_{in}}{dz}\ . \tag{2.4}$$

In the present case, this force is entirely delivered by the electric force. With Eqs. (2.1) and (2.2):

$$F_{el} = -\frac{1}{2}V^2\frac{dC}{dz} + V\frac{dQ}{dz} = \frac{1}{2}V^2\frac{dC}{dz}\ . \tag{2.5}$$

It is seen that if the source work term dW_{in} is not taken into consideration for a *voltage controlled* device, the electric force would come out with the wrong sign. The case of a *charge controlled* device is treated in subsection 2.1.3. The electric force causes a displacement of the movable top electrode, which changes the capacitance. It is clear that the position of the bridge electrode and thus the position of the bridge electrode (not surprisingly) depends symmetrically on the voltage, i.e. *C(V) = C(–V)*.
The experiments presented in Chapters 3 and 4 are performed on samples on which in many cases the interpretation assumes a parallel plate configuration (Fig. 1), which in

[2] The reason for attributing the mechanical energy to W_{out} rather than U is to extract a separate formula for the electric force.

Chapter 4 can be considered exact[3]. The capacitance of a parallel plate capacitor with electrodes separated a distance z (see Fig. 2.1) is given by Eq. (2.6):

$$C = \frac{\varepsilon_0 A}{z} . \tag{2.6}$$

Here ε_0 is the permittivity of free space and A the effective capacitive area, i.e. the area that top and bottom electrodes overlap. It is assumed that the separation z is much smaller than the lateral dimensions, which allows for neglecting fringe fields. For a parallel plate capacitor, as in Fig. 1, the electric force Eq. (2.7) is found by inserting Eq. (2.6) into Eq. (2.5):

$$F_{el} = -\frac{\varepsilon_0 A}{2z^2} V^2 = -\frac{C}{2z} V^2 . \tag{2.7}$$

The structure is assumed to be elastically movable: the displacement is proportional to the applied force. The corresponding mechanical potential energy is given by:

$$W_m = \int_{z_0}^{z} -z' k(z') dz' = ½ k (z_0 - z)^2 . \tag{2.8}$$

Here z_0 is defined as the equilibrium position of the bridge structure in absence of any loading forces and z the actual position. The second equality sign in Eq. (2.8) applies in cases that deflections are sufficiently small for the spring constant to be independent of deflection. For appropriate definitions of stiffness and deflection, Eq. (2.8) is valid not only for parallel plate capacitors with a suspended bridge, but also in cases where deflection is a function of the lateral position, for example a cantilever beam or a clamped-clamped beam. In equilibrium, the electric force Eq. (2.5) and the spring force *on* the bridge $-F_s = -dW_m/dz$ balance:

$$-\frac{\varepsilon_0 A}{2z^2} V^2 + k(z - z_0) = 0 . \tag{2.9}$$

The first term represents the electric force and the second the mechanical force. Rewriting this expression yields an implicit relation between capacitance and voltage, see Eq. (2.10):

$$\frac{C_0^4}{C^2}\left(\frac{1}{C_0} - \frac{1}{C}\right) = \alpha V^2 . \tag{2.10}$$

Here $C_0 = C(V = 0) = \varepsilon_0 A / z_0$ is the capacitance at zero bias voltage. By definition,

[3] That is to say, the underlying design of the structures is such that the electrodes remain parallel during actuation, unlike for example the RF-sensors in Chapter 3. Those bridge electrodes deform to a curved shape when actuated.

$$\alpha \equiv \frac{\varepsilon_0^2 A^2}{2kz_0^4}. \tag{2.11}$$

There exists an explicit solution for C from Eq. (2.10) as a function of the bias voltage V. Because of its complicated form it is however common practice to take the Taylor expansion of the left-hand side of Eq. (2.10), which is valid for small voltages so that C is close to C_0. This reduces the capacitance – voltage relation to a simple quadratic expression:

$$C = \alpha V^2 + C_0. \tag{2.12}$$

In the next subsection a notion of 'small' voltages will be developed, after which the accuracy of this approximation can be estimated. When voltages increase, or if a still better approximation is desired, one can make a quadratic approximation of the left-hand side of Eq. (2.10) around $C = C_0$. Equating this with αV^2 and solving for C leaves still a relatively tractable explicit relation:

$$C = \frac{1}{6} C_0 \left(7 - \sqrt{1 - \frac{6C_0}{kz_0^2} V^2} \right). \tag{2.13}$$

Needless to say, for small voltages this reduces exactly to the parabolic approximation Eq. (2.12). The disadvantage of this *elliptic* form is that for data-analyses it involves the somewhat more cumbersome method of parameter-fitting (providing values for C_0 and the coefficient of V^2), which requires reasonable preliminary estimates of the coefficients as starting values. Furthermore, the form and its coefficients are intuitively less readily associable to a series of measurement data. In contrast, the *parabolic* form of Eq. (2.12) involves only basic polynomial fitting and the thus estimated values for C_0 and α are immediately recognizable as zero-voltage capacitance and parabolic curvature, which is immediately associated with sensitivity.

A final point of attention concerns the remark made in the beginning of the subsection that capacitance does not depend on electrostatic quantities such as voltage, while in Eq. (2.12) and Eq. (2.13) such relationships are explicitly formulated. This conceptual paradox is easily resolved by recognizing that the particular geometry, which determines the capacitance, can be influenced by external forces. They can be of any nature, and in the case of capacitive MEMS they happen to be electrostatic.

2.1.2 Pull-in voltage and electromechanical coupling

For higher voltages, the resulting non-linear electrical force can not be balanced anymore by the linear working of the mechanical force. Indeed, there exists a point at which an infinitesimally small increase in voltage would generate an infinitely large displacement of the bridge, which implies $dz/dV \rightarrow \infty$. For the derivation of the voltage at which this instability, called *pull-in*, occurs, Eq. (2.9) can be invoked, because the forces are just still

balancing. Solving this expression for V and equating the derivative $dV/dz = (dz/dV)^{-1} = 0$ provides for a configuration in which the plates remain parallel the following result:

$$V_{pull-in} = \sqrt{\frac{8kz_0^3}{27\varepsilon_0 A}}\ . \tag{2.14}$$

For this configuration, pull-in occurs when the bridge has lowered to 2/3 of the zero-bias state or, equivalently, the capacitance has increased by a factor of 1½ compared to the unactuated state. At this point, the spring force ceases to prevent collapse of the top electrode. Pull-in limits the operation range of a capacitive MEMS sensor for displacement or electromagnetic related quantities such as the power through a coplanar waveguide, although this can be resolved by equipping the device with a feedback circuit that keeps the bridge in place. If a MEMS device is deployed as a tunable capacitor, there are some methods available to considerably stretch the 'tuning ratio' (C_{max} / C_{min}) of 1½. Tuning ratios of 20 have been reported in [1]. On the other hand, not only does the pull-in effect serve useful purposes in switches, it also constitutes a characteristic voltage for a particular design, for which reason it has been proposed to use it as a standard voltage reference [2-4], concurring with Zener diodes in stability.
The pull-in instability is not yet reflected in the parabolic approximation Eq. (2.12), but the elliptic approximation Eq. (2.13) already contains this strong non-linear behavior. Here $dz/dV \rightarrow \infty$ occurs when the expression underneath the square root vanishes, which predicts instability for a voltage that is exactly a factor ¾ lower than the 'correct' expression Eq. (2.14).
The following remark is worth making: Assuming the device under test is a parallel plate capacitor and the measurement is being carried out well (no important unknown parasitic capacitances for example), a parabolic fit through a set of data at small voltages already suffices for estimating pull-in, without requiring further knowledge of the sample, such as for example the gap height z_0. Because in the parabolic approximation Eq. (2.12) naturally the same coefficients (C_0, α) appear as in the original force balance Eq. (2.10) from which it was derived, the pull-in voltage, Eq. (2.14) can be expressed in these coefficients alone:

$$V_{pull-in} = \sqrt{\frac{4C_0}{27\alpha}}\ . \tag{2.15}$$

The existence of a characteristic voltage such as pull-in voltage allows for a notion of 'small' voltages: For $V << V_{pull-in}$ the parabolic approximation Eq. (2.12) provides an accurate description of the dependence of the capacitance on the voltage. More concretely, for $V < 0.2V_{pull-in}$, the difference between the implicit function and its quadratic approximation is practically zero, while for $V = 0.5V_{pull-in}$ the approximation predicts a capacitance that is about 0.5% too low. This may be considered a limit below which the parabolic approximation is very well applicable. For larger voltage, the difference rapidly grows to about 1% for $V = 0.6V_{pull-in}$ and to 23.5% for $V = V_{pull-in}$. These numbers are independent of the electrode area or the initial capacitive gap. The above-mentioned derivation of the pull-in voltage is tightly linked to the notion of *electromechanical coupling*, with which is meant how mechanical properties, in

particular the spring constant, can be affected when under influence of electrostatic interaction. This is digressed on along a short discussion that stays intuitively close to the physics that are involved. Let us reconsider Eq. (2.3). Suppose we exert an additional external force F_{ext} on the bridge, for example by pushing it down with a finger or putting a weight on it. The input work then receives, apart from the source, an extra contribution $F_{\mathrm{ext}}dz$. The total balance now looks:

$$\tfrac{1}{2}V^2 dC = VdQ + F_{ext}dz - k(z - z_0)dz \,. \tag{2.16}$$

Again, the left hand side term is the increase of internal (electric) energy of the MEMS device. The right hand side terms are the source and external work done *on* the MEMS and the last term the work done *by* the MEMS. Dividing by *dz* and rearranging gives:

$$F_{ext} = k(z - z_0) - \tfrac{1}{2}V^2 \frac{dC}{dz} \,. \tag{2.17}$$

As said, the external force can be of any nature in principle, but in particular an (oscillating) electric force is of our interest, when dealing with resonant behavior, see section 2.4. The electromechanical stiffness is defined as the reluctance with which the bridge responds with displacement to an external force. We have:

$$K_{EM} \equiv \frac{dF_{ext}}{dz} = k + k_{el} \,. \tag{2.18}$$

Where we have defined

$$k_{el} \equiv -\tfrac{1}{2}V^2 \frac{d^2C}{dz^2} = -\frac{\varepsilon_0 A}{z^3}V^2 \,. \tag{2.19}$$

as the electric stiffness, which offers a *negative,* voltage dependent contribution to the effective, or electromechanical stiffness K_{EM}. The equality sign in Eq. (2.19) holds in case of a parallel plate capacitor. We now have a better founded argument in the derivation of the pull-in voltage in this type of device, which follows actually the same route as the one to Eq. (2.14). Namely, it is now easily recognized that some point must exist at which the effective stiffness Eq. (2.18) will decrease to zero and the bridge is on the verge of collapsing. Combining Eq. (2.19) with $K_{EM} = 0$ in Eq. (2.18) and Eq. (2.9) again yields the pull-in voltage.

2.1.3 Charge control

As we have seen in the previous subsection, the bridge of a capacitive MEMS connected to a voltage source cannot be lowered for the complete gap in a stable way: below $z = 2z_0/3$ pull-in occurs. It is nevertheless possible to lower the bridge steadily over the complete gap; this requires disconnecting the bridge from the voltage source and employ *charge control* instead. In a “Gedankenexperiment”, one could pick a pair of tweezers to take charges from the fixed electrode and drop them on the bridge electrode. In this case:

$$F_e = -\frac{d}{dz}U_e = -\frac{d}{dz}\frac{Q^2}{2C} = \frac{Q^2}{2C^2}\frac{dC}{dz} = -\frac{Q^2}{2\varepsilon_0 A}\,. \qquad (2.20)$$

Again, the last identity holds exclusively for a parallel plate capacitor. The force is now independent of the height. In fact we have just a test charge Q (bridge) in the field of an infinitely wide sheet charge Q/A (fixed electrode). Balancing by the spring force $k(z_0 - z)$ yields the following dependence of the capacitance on the charge:

$$C_Q = \frac{\varepsilon_0 A}{z_0 - Q^2 / 2\varepsilon_0 kA}\,. \qquad (2.21)$$

Adding charges will lower the bridge and increase the capacitance. The bridge touches for $Q_{down} \equiv \sqrt{2\varepsilon_0 k z_0 A}$, when the capacitance Eq. (2.21) tends to infinity. Initially, adding charges will also increase the potential difference between the plates, but this is compensated for the fact that a lowered bridge decreases the voltage again. For the voltage we have:

$$V_Q = \frac{Q}{\varepsilon_0 A}\left(z_0 - \frac{Q^2}{2\varepsilon_0 kA} \right). \qquad (2.22)$$

This voltage has a maximum at $Q_{\max} = \frac{1}{\sqrt{3}} Q_{down} \approx 0.58 Q_{down}$, for which it attains a value

$$V_{\max} = \sqrt{\frac{8kz_0^3}{27\varepsilon_0 A}}\,, \qquad (2.23)$$

which is equal to the pull-in voltage Eq. (2.14) for a voltage-controlled MEMS bridge. Automatically, this maximum occurs then when the bridge is lowered $z_0/3$ downwards and the capacitance has increased to 1½ of its original value, as is easily checked. This voltage dependence is plotted in Fig. 2.2.

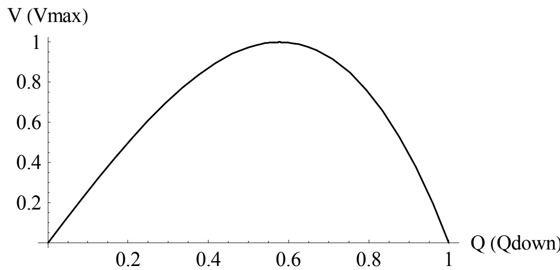


Fig. 2.2 Voltage dependence on charge in charge control

As can be seen from Eq. (2.22), the initial slope of this graph is $1/C_0$. Charge controlled operation has some interest in the field of current-driven MEMS switches [5,6].

2.1.4 Dielectric layer

In many cases capacitive MEMS structures contain dielectric layers covering at least one of the electrodes. Either intended, for example a layer protecting against damage or shortcut resulting from pull-in, or 'half-intended', as in the case of the fist generation HF power sensors described in Chapter 3, where a bridge made of dielectric material (silicon nitride) functions as a carrier for the electrodes which are placed on top. Here the dielectric layer is not applied in order to have dielectrics there, but is a necessity imposed by design. At last, dielectrics come 'along the way', unintended, for example the ultra-thin layer of native oxide that forms on aluminum electrodes, as described in the second generation HF power sensors in Chapter 3 and the double beam structures in Chapter 4. Any modifications to the relevant expressions in the previous subsections are restricted to parallel plate capacitors. A schematic representation is given in Fig. 2.3:

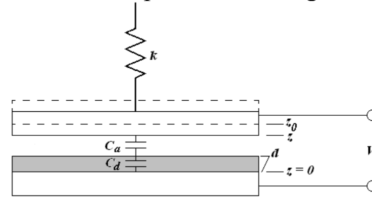


Fig. 2.3. Parallel plate capacitor with a dielectric layer of thickness d, forming a series connection of a dielectric capacitance C_d with air (or actuated) capacitance C_a. The voltage V of the top electrode is with respect to the bottom electrode.

A single dielectric layer is assumed on top of the bottom electrode. However, the relations in this subsection apply equally in case of two dielectric layers, one on each electrode, whose individual thicknesses add up to *d*. In this particular configuration of a parallel plate capacitor, the surface of the dielectric constitutes an equipotential. Rather by chance geometry coincides with the electrostatic 'landscape'. Hence the original capacitance divides into two series capacitances; a variable one (C_a) over the air gap and a constant one over the dielectric (C_d). The total applied voltage $V_t = V_a + V_d$ is distributed over the two capacitors as:

$$V_a = \frac{C_d}{C_a + C_d} V_t \quad \text{and} \quad V_d = \frac{C_a}{C_a + C_d} V_t \,. \tag{2.24}$$

In a measurement, it is the total voltage we apply and a total series capacitance that is returned. The total capacitance C_t combines the two series capacitors following:

$$C_t = \frac{C_a C_d}{C_a + C_d} \,, \tag{2.25}$$

where each of the single series capacitors has a larger capacitance value than the total series capacitance. If Eq. (2.25) is solved for C_a and is together with Eq. (2.24) inserted for the quantities *C* and *V* respectively in Eq. (2.12), the parabolic relation for the series configuration is obtained:

$$C_t(V) = \alpha_t V_t^2 + C_{t0} \,. \tag{2.26}$$

Here

$$C_{t0} = \left(\frac{C_d}{C_{a0} + C_d} \right) C_{a0} \tag{2.27}$$

is the total series capacitance in absence of actuation, where C_0 denotes the unbiased capacitance over the air gap alone. Furthermore,

$$\alpha_t = \left(\frac{C_d}{C_0 + C_d} \right)^4 \alpha_a , \tag{2.28}$$

in which α_a, the "gap sensitivity", is defined as in Eq. (2.11) but with z_0 replaced by $z_0 - d$. However, the purpose was to compare with the absence of a dielectric layer, i.e. with Fig. 2.1 and corresponding expressions. For this, we assume a dielectric layer with permittivity ε of thickness d. Defining $\eta \equiv d/z_0$ the portion of the unactuated gap that is taken up by the dielectric layer, it is easily verifiable that the ratio r of the capacitance including dielectric layer to the capacitance without becomes

$$r \equiv \frac{C_{t0}}{C_0} = \frac{1}{1 - \eta(1 - 1/\varepsilon)} \geq 1 . \tag{2.29}$$

The presence of a dielectric layer increases the capacitance, as is well known from standard electrostatics. It is insightful to check the limiting cases: For $\eta \rightarrow 0$ and $\eta \rightarrow 1$ (no dielectric and completely filled with dielectric) we have $r \rightarrow 1$ and $r \rightarrow \varepsilon$ respectively. For $\varepsilon \rightarrow 1$ (no dielectric) we have again $r \rightarrow 1$. At last, $\varepsilon \rightarrow \infty$ lets the dielectric grow to become a conductor in a pure electrostatic sense (infinite polarizability), which in effect reduces the original capacitive gap, which is reflected in $r \rightarrow (1 - \eta)^{-1}$. Now, the force balance on the bridge is given by:

$$\frac{1}{C_a^2} \left(\frac{1}{C_{ao}} - \frac{1}{C_a} \right) = u V_a^2 , \tag{2.30}$$

in which $u \equiv (2k\varepsilon_0^2 A^2)^{-1}$. Using Eqs. (2.24), (2.25) and (2.27) to substitute C_a, C_{a0} and V_a in Eq. (2.30), we are left with an expression for the force balance in terms of overall total capacitances and voltages:

$$\frac{1}{C_t^3} \left(\frac{C_t - C_{t0}}{C_{t0}} \right) = u V_t^2 . \tag{2.31}$$

C_d drops out. In the linear approximation for C_t close to C_{t0} this expression reduces to

$$C_t(V) = \alpha_t V_t^2 + C_{t0} . \tag{2.32}$$

Here $\alpha_t \equiv u C_{t0}^4$, so as a result:

$$C_t(V) = r^4 \alpha V_t^2 + r C_0 , \tag{2.33}$$

where α is defined as in Eq. (2.11). The conclusion is that the zero capacitance increases by a factor r as defined in Eq. (2.29) upon inserting a dielectric layer, as expected. The parabolic coefficient, the 'sensitivity', increases by a factor of r^4, which is not surprising too when considering that $\alpha = uC_0^4$ contains four factors of original zero-capacitances.
A logical consideration is how the pull-in voltage alters by the insertion of a dielectric slab. First, the capacitance is

$$C = \frac{\varepsilon_0 A}{z - d_{eff}}, \qquad d_{eff} \equiv d\left(1 - \frac{1}{\varepsilon}\right) > 0, \tag{2.34}$$

where an 'effective gap reduction' has been defined. It describes how much the gap is effectively reduced by the presence of the dielectricum. We can write $r = z_0/(z_0 - d_{\text{eff}})$. The magnitude of the electrostatic force in a capacitor with dielectric layers having a *combined* thickness d is given by:

$$F_{el} = \frac{\varepsilon_0 A}{2(z_0 - d_{eff})^2} V^2 . \tag{2.35}$$

This, combined with Eqs. (2.33) and (2.20) in which $K_{\text{EM}} = 0$, yields the pull-in voltage Eq. (2.36),

$$V_{pull-in}(d_{eff}) = \sqrt{\frac{8k(z_0 - d_{eff})^3}{27\varepsilon_0 A}} < V_{pull-in}(d_{eff} = 0), \tag{2.36}$$

which is smaller than the case without a dielectric due to the enhanced electric coupling. The critical height and capacitance at which the bridge becomes unstable is in this case:

$$z_{pull-in} = \tfrac{2}{3} z_0 + \tfrac{1}{3} d_{eff}, \tag{2.37}$$

$$C_{pull-in} = C(z = z_{pull-in}) = \frac{3}{2} \frac{\varepsilon_0 A}{(z_0 - d_{eff})} . \tag{2.38}$$

Even with a dielectric layer the pull-in capacitance is 1½ times that of the uactuated capacitance. It is possible to insert a dielectric layer sufficiently thick so that the bridge never arrives at an unstable point within the region in which the bridge can freely move. In other words, the bridge can be brought in contact with the dielectric surface over a controlled trajectory. Of course a dielectric that fills the complete airgap radically prevents pull-in – and any movement whatsoever. But more in general, the above is satisfied whenever

$$\eta = \frac{d}{z_0} \geq \frac{1}{1 + 1/(2\varepsilon)} . \tag{2.39}$$

In the context of dielectric layers it is meaningful to address the topic of *pull-out voltage*. Consider a parallel plate capacitive MEMS with at least one dielectric layer. If the voltage crosses the threshold of pull-in, the bridge collapses. In this case however, the capacitance will remain finite after collapse, what we will call *down capacitance*, where

on purpose the term “pull-in capacitance” is declined, which is rather associated with the capacitance at which pull-in is *about to occur*. In this case, the down-capacitance coincides with the dielectric capacitance[4]:

$$C_{down} = C_d = \frac{\varepsilon\varepsilon_0 A}{d}. \tag{2.40}$$

To release a pulled-in bridge, the voltage knob has to be turned back so that the electric and mechanical forces cancel for $z = d$:

$$k(z_0 - d) = \frac{\varepsilon\varepsilon_0 A}{2d^2} V^2, \tag{2.41}$$

so that

$$V_{pull-out} = \sqrt{\frac{2kd^2(z_0 - d)}{\varepsilon\varepsilon_0 A}}. \tag{2.42}$$

The whole process is illustrated in Fig. 2.4.

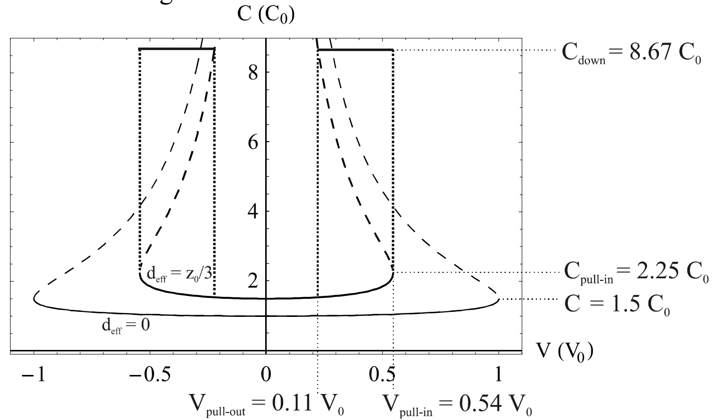


Fig. 2.4 Theoretical C(V)-graphs, showing the influence of a dielectric layer (thick curve).It is compared to the situation without dielectric (thin curve), which is normalized to its unactuated capacitance C_0 and its pull-in voltage V_{pi}. Assumed is $d_{eff} = 1/3$, for example S_iO_2 (ε = 3.9) filling 45% of the gap. This increases the capacitance by a factor r = 1½. The solid parts of the curves represent the physically accessible and stable states of the corresponding devices. The long-dashed parts of the curves correspond to continued (V,C) - solution set to the implicit relations Eq. (2.10) (thin) and Eq. (2.31) (thick). Arrows indicate an operation cycle of increasing the voltage starting from V = 0. At pull-in voltage, the bridge follows the short-dashed path. It collapses to C_{down},, which value is determined by the thickness(es) and the relative permittivity of the dielectric layer(s). Decreasing the voltage then will not have any effect on the bridge until it subsides below a threshold voltage (pull-out). The innermost short-dashed path is followed and the bridge is pulled upward then.

[4] This is because the bottom face of the movable top electrode is always parallel to the top face of the underlying dielectric layer, so that full contact will be established upon pull-in. For a clamped-clamped beam for example, a pulled-in top electrode has only a very small contact area in the centre, while on both off-centre areas a narrow air gap is maintained, that keeps the down-capacitance below the dielectric capacitance.

2.2 BEAM DEFLECTION AND STRESS

In many cases, MEMS structures are not readily described by the simple parallel plate configuration dealt with in the previous section. Rather, a natural consequence of the commonly applied surface micromachining techniques (patterning and stacking of layers) is that the movable part of the MEMS devices consists of a flexible membrane, floating above an electrode and clamped at one side (a cantilever), two sides or even four sides (also called 'diaphragm')[5]. This membrane may be the electrode itself (it is made of, for example, aluminum), or be the carrier of an electrode on it (silicon nitride is a suitable and widely used material). An accurate description requires the theory of the bending of plates and beams when subjected to a certain load. A renowned standard work in this field is [7]. This section will concentrate on clamped-clamped beams only, because of its relevance to analysis of RF sensors, in particular stress (section 3.3).

2.2.1 Beam equation

Consider now a beam[6] of length *L,* width *b*, and thickness *h*, subjected to a transverse load[7] $P(x)$ and an axial (tensile) force *S* (*s*ee Fig. 2.5):

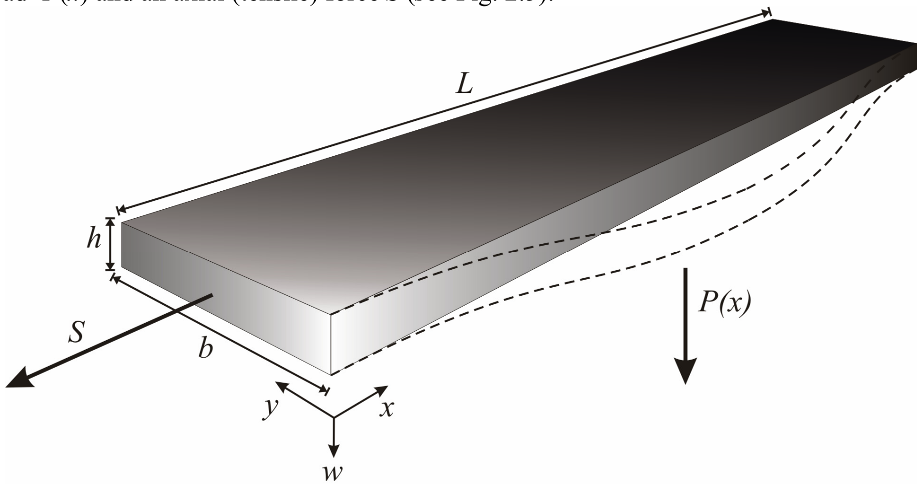


Fig. 2.5 Loaded clamped-clamped beam

Small deflections can generally be described by the Euler-Bernoulli equation [8]:

$$\frac{\partial^2}{\partial x^2}\left(EI\frac{\partial^2 w}{\partial x^2}\right) - S\frac{\partial^2 w}{\partial x^2} + \rho bh\frac{\partial^2 w}{\partial t^2} = P(x)\,. \tag{2.43}$$

---

[5] Assuming the membrane is rectangular in all these cases.

[6] Generally one speaks of a beam when its length L is considerably larger than its width b: $L > 5b$. To be more to the point, a beam is by definition a one-dimensional problem, where the deflection $w(x)$ depends on a single coordinate; unlike plates or diaphragms, which deflection $w(x,y)$ depends on two coordinates.

[7] Load, also known as distributed force, is measured in Newton per meter.

Here E is the plate modulus, which is equal to Young's modulus Y modified by a factor containing Poisson's ratio[8] ν via $E = Y / (1 - \nu^2)$. I is the area moment of inertia[9]. For a rectangle, I is given by $bh^3/12$. The product EI is known as *flexural rigidity* and is a measure of the beam's resistance against bending caused by transverse forces. ρ is the mass density of the material involved; ρbh is hence a line mass density. Incidentally, $EI = 0$ in Eq. (2.43) describes a vibrating string. In this context, we are merely interested in (quasi)statics, hence the kinetic term is set to zero: $\partial^2 w/\partial t^2 = 0$. The partial differential equation reduces to an ordinary one. A further simplification is brought by the assumption that the flexural rigidity EI is constant over the length of the beam[10], resulting in Eq. (2.44).

$$EI\frac{d^4 w}{dx^4} - S\frac{d^2 w}{dx^2} = P(x)\ . \tag{2.44}$$

Solving a concrete problem requires four additional boundary conditions, as Eq. (2.44) is a fourth order differential equation. The two most abundant boundary conditions are "clamped" (position w and angle (or slope; dw/dx) fixed; often zero, but not necessarily so), "free end" ($d^2w/dx^2 = 0$: zero curvature, and $d^3w/dx^3 = 0$: zero shear stress). These boundary conditions are both present in the case of a cantilever: it has one clamped end and one free end. There are more types of boundary conditions, which we will not deal with further. Finally, it should be remarked that in many texts, a *downward* deflection is considered *positive*; as a consequence, also downward forces must be taken positive.

2.2.2 Electric load

The situation focused on in this section is a clamped-clamped beam of length L, which in our case has the most natural conditions $w = 0$ and $dw/dx = 0$ at its two ends. This specific subsection aims at attaining some more insight in the problem at hand. Negligible tensile force ($S = 0$) is assumed, and a force on all points on the beam, to have the load function $P(x)$ continuous along the whole length of the beam. Later, a partial load will be considered.

The simplest cases are point loads (concentrated forces), and constant loads. An example of the latter is a bridge bending under its own weight. In the current context, the effect of weight of a micromachined beam can in most cases be neglected[11]. We are interested in the electric force, caused by a voltage V between the fixed and the flexible electrodes (separated a distance z_0 without actuation), see Fig. 2.6:

[8] Poisson's ratio, a material constant, accounts for the fact that a piece of material is slendered upon stretching.
[9] Alternatively called 'second moment of inertia'.
[10] A counterexample is triangular cantilevers for atomic force microscopy, which are commercially available. Their area moment of area is a function of position.
[11] Downscaling from the household meter to a micrometer decreases volume (hence mass) with a factor 10^{18}.

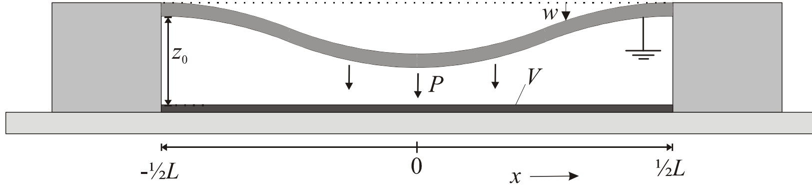


Fig. 2.6 Clamped-clamped beam under a distributed electric load.

The electric force causes a load

$$P = \frac{\varepsilon_0 b V^2}{2(z_0 - w)^2} . \tag{2.45}$$

This however makes the resulting differential equation Eq. (2.44) nonlinear. For low voltages, so that $w << z_0$, it is a reasonable approximation to solve just the uniform-load equation

$$\frac{d^4 w}{dx^4} = \xi , \qquad \xi \equiv \frac{\varepsilon_0 b V^2}{2 z_0^2 EI} . \tag{2.46}$$

With boundary conditions

$$w(-\tfrac{1}{2}L) = w(+\tfrac{1}{2}L) = 0 , \tag{2.47a}$$

$$\left.\frac{dw}{dx}\right|_{x=-\frac{1}{2}L} = \left.\frac{dw}{dx}\right|_{x=+\frac{1}{2}L} = 0 , \tag{2.47b}$$

this problem has as explicit solution

$$w(x) = \frac{\xi}{24}(x^2 - \tfrac{1}{4}L^2)^2 . \tag{2.48}$$

As soon as the displacement function is known, a capacitance-voltage relation can be established by

$$C(V) = \int_{-\frac{1}{2}L}^{\frac{1}{2}L} \frac{\varepsilon_0 b}{z_0 - w(x,V)} dx , \tag{2.49}$$

in which the voltage dependence of the displacement has been made explicit. For Eq. (2.48), the integral Eq. (2.49) can be solved exactly in terms of arctangent functions, but a bit more illuminating is if we approximate:

$$C(V) \approx \int_{-\frac{1}{2}L}^{\frac{1}{2}L} \frac{\varepsilon_0 b}{z_0}\left(1 + \frac{w(x,V)}{z_0}\right) dx = \frac{\varepsilon_0 A}{z_0} + \frac{\varepsilon_0^2 A^2 L^3}{1440 EI z_0^4} V^2 , \tag{2.50}$$

in which $A = bL$ is the total area. If we compare Eq. (2.50) to a parallel plate $C(V)$-relation (Eqs. (2.11) and (2.12)), we could associate an equivalent stiffness

$k_{eq}L^3 = 720EI = 60Ebt_m^3$, with t_m the thickness of the membrane. Different interpretations of stiffness are viable however, for example the ratio of the total electric force PL to the deflection of the center $w(x = 0)$: K_{eff}, as in subsection 2.2.4, which in this situation (Fig. 2.6) would read $K_{\text{eff}}L^3 = 384EI$. Note that both are quite larger than the stiffness for a point-loaded beam (See Chapter 4, Eq. (4.7)).
An enhancement in accuracy, especially at slightly higher voltages, is achieved if the load function Eq. (2.45) is linearized, so that the load at least retains a deflection-, or position-dependent character. The solution to the resulting equation

$$\frac{d^4 w}{dx^4} = \xi + \frac{2\xi}{z_0} w \tag{2.51}$$

reads:

$$w(x) = G\cosh ax + H\cos ax - ½z_0\,, \qquad a \equiv \sqrt[4]{2\xi / z_0} \tag{2.52}$$

with

$$\frac{G}{\sin(½aL)} = \frac{H}{\sinh(½aL)} = \frac{½d_0}{\cosh(½aL)\sin(½aL) + \cos(½aL)\sinh(½aL)}\,. \tag{2.53}$$

For larger voltages, the predictions for bridge deflection are for the uniform and linearized load functions significantly different. Let $P^{(n)}$ be the nth order expansion of the load function around $w = 0$ and $w_0 \equiv w(P^{(0)}, x = 0)$ and $w_1 \equiv w(P^{(1)}, x = 0)$ the deflection of the central point of the beam for the 'zeroth' and first order load function expansions. The ratio w_1/w_0 is then set out in Fig. 2.7 for different voltages in order to visualize in when their difference (or ratio) becomes appreciable. The two approaches are compared for parameters typical for structures in Chapter 3 (L = 300 μm, b = 1800 μm, z_0 = 1 μm, h = 1 μm, Y = 69 GPa, ν = 0.35, E = 78.65 GPa). The upper limit voltage 3. V is in the order of estimated pull-in.[12]

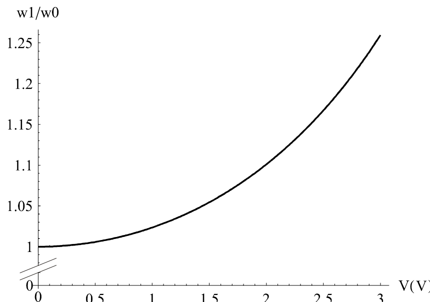


Fig. 2.7 Voltage dependence of w_1/w_0 at $x = 0$

Deflection profiles of these two approximations are plotted together with a third in the next subsection (2.2.3).

[12] At this moment, only rough guesses can be made for $V_{\text{pull-in}}$. Based on the parabolic coefficients of Eq. (2.50) and Eq. (2.15), $V_{\text{pull-in}}$ would be 4.4 V for the given parameters. For $k = K_{\text{eff}}$, and using Eq. (2.14) $V_{\text{pull-in}}$ = 3.2 V. It is stressed however, that Eqs. (2.14) and (2.15) apply for a strict parallel plate MEMS only. For the case at hand, the parallel plate configuration is distorted during actuation.

2.2.3 Bridge instability

The solutions Eqs. (2.48) and (2.52) both conceal important critical behavior of the bridge. As for the parallel plate configuration, cf. Eq. (2.12), a sufficiently high voltage will result in collapse of the bridge.
The stressless ($S = 0$), electrostatically loaded beam equation

$$\frac{d^4 w}{dx^4} = \frac{\xi d_0^2}{(z_0 - w)^2} \tag{2.54}$$

has no closed form solution with desired properties[13]. Instead, one can proceed by expanding the right-hand side of Eq. (2.54) around $w = 0$ by

$$\frac{1}{(z_0 - w)^2} = \frac{1}{z_0^2} + \frac{2}{z_0^4} w + \frac{3}{z_0^4} w^2 + \ldots = \sum_{i=0}^{\infty} \frac{i+1}{z_0^{i+2}} w^i \; . \tag{2.55}$$

Then, we insert for $w(x)$ a power series

$$w(x) \equiv \sum_{n=0}^{\infty} a_n x^n \; . \tag{2.56}$$

This has as a fourth derivative

$$\frac{d^4 w}{dx^4} = \sum_{n=4}^{\infty} n(n-1)(n-2)(n-3) a_n x^{n-4} \; . \tag{2.57}$$

The solution has to be symmetrical around $x = 0$, which leads us to drop the odd terms in Eqs. (2.56) and (2.57). Combining this with Eqs. (2.54) and (2.55) yields the relation

$$24a_4 + 360a_6 x^2 + 1680a_8 x^4 + \ldots = \sum_{i=0}^{\infty} (i+1) \frac{\xi}{z_0^i} \left(a_0 + a_2 x^2 + a_4 x^4 + a_6 x^6 + \ldots \right)^i \; . \tag{2.58}$$

Note the expansion on the right-hand side is to the ith power. Equating like powers of x now yields expressions for the coefficients a_4, a_6, a_8,… in terms of a_0 (= $w(0)$, the deflection of the central point) and a_2. For $x = 0$ we have immediately

$$a_4 = \frac{\xi z_0^2}{24(z_0 - a_0)^2} \; , \tag{2.59}$$

---

[13] In fact, there exists a solution to this equation: $w(x) = z_0 + \sqrt[3]{\tfrac{81}{40} \xi z_0^2 (x - A)^4}$, with A an arbitrary constant, but this function cannot come close to what is desired. Symmetry, $w(x) = w(-x)$, requires $A = 0$, while the boundary condition $w'(½L) = 0$ requires $A = ½L$. The constriction $w(½L) = 0$ asks for a complex A, rendering the whole function complex. Last but not least, for zero voltage ($\xi = 0$), we get $w = z_0$, where we would wish $w = 0$.

while equating quadratic and quartic terms in x gives:

$$a_6 = \frac{\xi}{180 z_0} a_2 , \tag{2.60}$$

$$a_8 = \frac{\xi}{840 z_0} a_4 + \frac{\xi}{560 z_0^2} (a_2^2 + 2 a_0 a_4) . \tag{2.61}$$

For calculating a_8 with Eq. (2.58), both $i = 1$ and $i = 2$ deliver terms. Here, a_4 can be eliminated by Eq. (2.59). Higher order terms become increasingly complicated. Calculations have to be done numerically then. At last, a_0 and a_2 are determined by the boundary conditions $w(½L) = 0$ and $w'(½L) = 0$. For demonstrating the instability of the bridge it suffices to determine the a_n until $n = 6$. Then, the boundary conditions dictate:

$$a_0 + a_2 (\tfrac{1}{2} L)^2 + a_4 (\tfrac{1}{2} L)^4 + a_6 (\tfrac{1}{2} L)^6 = 0 , \tag{2.62a}$$

$$a_2 (\tfrac{1}{2} L)^2 + 2 a_4 (\tfrac{1}{2} L)^4 + 3 a_6 (\tfrac{1}{2} L)^6 = 0 . \tag{2.62b}$$

It follows that a_0 has to satisfy

$$a_0 (z_0 - a_0)^2 = \frac{z_0^2 (\tfrac{1}{2} L)^4 \xi}{72} \left(\frac{180 z_0 - (\tfrac{1}{2} L)^4 \xi}{60 z_0 + (\tfrac{1}{2} L)^4 \xi} \right). \tag{2.63}$$

The maximum value of the left-hand side of Eq. (2.63) occurs when the center of the bridge is lowered by one-third of the gap: $a_{0,\max} = ⅓ z_0$, which gives a maximum value of $\tfrac{4}{27} z_0^3$. The right-hand side becomes maximal for a voltage corresponding with $(\tfrac{1}{2} L)^4 \xi = 60 z_0$, which counts in for a maximal value of $\tfrac{5}{6} z_0^3 > \tfrac{4}{27} z_0^3$. This demonstrates that for sufficiently high voltages, there are no real and physical solutions for Eq. (2.63), which signals the instability. According to this crude approximation up to 6th power in x, pull-in thus occurs when the left-hand side is maximal. Remembering the second part of Eq. (2.46), we have then

$$V_{pi} = \sqrt{\frac{64\left(127 - \sqrt{14689}\right) E I z_0^3}{3 \varepsilon_0 b L^4}} \approx 11.1 \sqrt{\frac{E I z_0^3}{\varepsilon_0 b L^4}} . \tag{2.64}$$

Using the parameters mentioned in the previous subsection (2.2.2), the deflection patterns of the three thus far encountered approximations are compared in Fig. 2.8 for a voltage of 3.30V, which according to Eq. (2.64) is at the verge of pull-in.

The three methods differ for factors of up to 2 for the deflection of the center when close to pull-in. To get a feel of the quality of the latter approximation, the left-hand side of the differential equation Eq. (2.54), a quadratic function, has been divided by the right-hand side, as a function of x, Fig. 2.9.

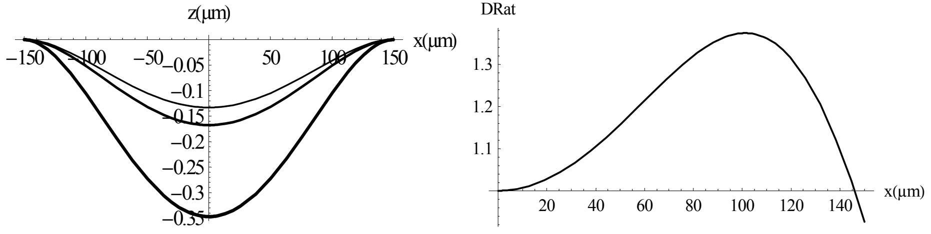


Fig. 2.8 (Left) Stressless bridge deflection for V = 3.30V for a uniform load (upper, thin), linearized load (middle, half-thick) and nonlinearized load, 6th order approximation in x (lower, thick).
Fig. 2.9 (Right) Left-hand side of Eq. (2.54) divided by right-hand side ("DRat") as a function of position

In the ideal case, this ratio "DRat" = $w''''(z_0 - w)^2/(\xi z_0^2)$ is unity everywhere. In the flank regions, the discrepancy is up to 30%, which is perhaps not too bad for describing the deflection profile so very close to a critical point (pull-in). The centre of the beam has automatically a unity DRat. This does not mean however, that the estimated value for the center deflection is exact: including more terms modifies Eq. (2.63), from which a_0 is to be solved. Instead, a measure for the average discrepancy could be calculated by $\frac{2}{L}\int_0^{L/2} DRat(x)^2\,dx$.

2.2.4. Beam with partial load and tensile stress

Now, consider a beam with a uniform partial load P_0 and a tensile force S, see Fig. 2.10. This case has been described in [9], based on [10]. In surface micromachining this is important because film stress is a common phenomenon. Additionally, in surroundings with variable temperature, differences in thermal expansion coefficients make themselves felt.

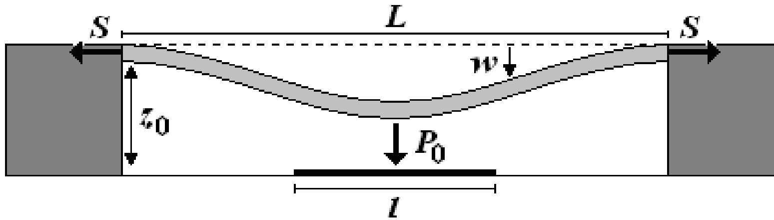


Fig. 2.10 Beam with partial electrostatic load and tensile stress

The maximum deflection w_{max} of a clamped-clamped beam caused by a transverse load distributed centrally and symmetrically around its centre point follows the relation

$$w_{\max} = \frac{P_0 l_c}{K_{eff}}, \tag{2.65}$$

with P_0 a uniform force density and l_c the length over which this force density is applied. The constant of proportionality linking deflection and applied transverse load K_{eff} is

sought after. If the thickness of the beam is much smaller than its length, the stiffness K_{eff} , taking into account axial forces, is well described by Eq. (2.65):

$$K_{eff} = \frac{-8k^2 L\lambda_r S \cosh\left(\frac{kL}{4}\right)}{(8+k^2L^2(\lambda_r-2)\lambda_r)\cosh\left(\frac{kL}{4}\right) - 8\cosh\left(\frac{kL}{4}(1-2\lambda_r)\right) + 4kL\lambda_r \sinh\left(\frac{kL}{4}\right)} . \quad (2.66)$$

In here, L is the length of the beam and $\lambda_r = l_c / L$ the portion of the beam that is subjected to the transverse load. k is defined as

$$k = \sqrt{\frac{12S}{Ebh^3}} . \quad (2.67)$$

S in (2.67) is the axial force, i.e. any force parallel to the beam attempting to stretch it. It is given by:

$$S = (\sigma_{NL} + \sigma_{other})bh , \qquad \sigma_{other} = \hat{\sigma}_0 + \sigma_T . \quad (2.68)$$

Here b and h are the width and the thickness of the beam respectively. Three stress contributions are taken into account. Firstly, σ_{NL} is the axial stress resulting from beam stretching when the beam is (non-linearly) deflected in transverse direction. It is estimated by $\sigma_{NL} = \pi^2 E w_{max}^2 / 4L^2$. Secondly, the fabrication of the beam may yield a residual film stress $\hat{\sigma}_0 = \sigma_0 (1 - v)$, with σ_0 the biaxial stress. Its value is to be taken at a fixed temperature T_0, for example room temperature. Thirdly at last, there is thermomechanical stress. In many cases, the bridge electrode and the substrate are composed of different materials, for example an aluminum bridge on a boron glass substrate. Variations in temperature will try to expand or contract both parts in unequal amounts. As the boron glass substrate is very much thicker (500 µm) than the aluminum bridge (1 µm) and much wider also, the forces governing the wafer's contraction dominate. The aluminum bridge has to fully comply to the stretching behavior of the glass. This gives rise to considerable stress in the bridge if temperature variations are large. It is assumed that this thermomechanical stress σ_T is given by

$$\sigma_T = E\frac{\Delta L}{L} = E(\alpha_2 - \alpha_1)(T - T_0) . \quad (2.69)$$

In Eq. (2.67), ΔL is the distance the bridge is stretched or compressed compared to the length it would have attained if the bridge were allowed to contract or expand freely upon changes in temperature. L is the beam length at actual temperature $T = T_0$. α_1 and α_2 are the linear thermal coefficients of top and bottom material. E applies for the aluminum bridge.

If the deflection of the beam is not greater than roughly its thickness, which condition is fulfilled for MEMS structures investigated here, K_{eff} can be assumed independent on w_{max} and approximated by neglecting σ_{NL}. This results in:

$$K_{eff} = -\frac{64Ebh^3}{3L^3}\frac{\eta^4\lambda_r\cosh(\eta)}{(1-2\eta^2\lambda_r(2-\lambda_r))\cosh(\eta)-\cosh(\eta(1-2\lambda_r))+2\eta\lambda_r\sinh(\eta)}, \quad (2.70)$$

$$\eta \equiv \frac{L}{2h}\sqrt{\frac{3\sigma_{other}}{E}}. \quad (2.71)$$

For example, the non-linear bending stress σ_{NL} is for a maximal deflection of 100 nm in case of a $L \times b \times h = 300 \times 3600 \times 1\ \mu m^3$ membrane about 19 kPa, much smaller than typical residual film stresses, that are in the order of tens of megapascals. The applicability of Eq. (2.68) as an approximation is extra supported by that the first correction on Eq. (2.70) in the power expansion of Eq. (2.66) in w_{max} is quadratic, not linear, which term is zero.
For negligible tensile stresses (the limit in which η approaches zero), Eq. (2.70) can be approximated further by:

$$K_{eff} = \frac{32Ebh^3}{L^3(2-\lambda_r^2(2-\lambda_r))}. \quad (2.72)$$

How the stiffness depends on the load distribution is shown in Fig. 2.11 for a particular geometry.

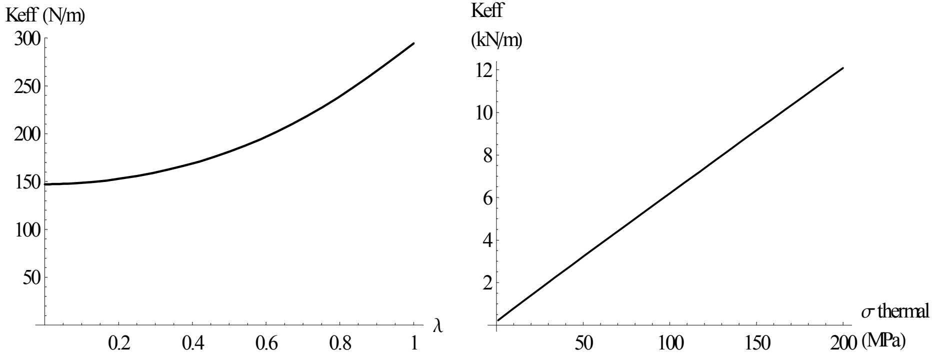


Fig. 2.11 (left) K_{eff} vs. λ_r (= l_c/L) for bLh = 3600 × 300 × 1 μm^3 bridge for zero axial load.
Fig. 2.12 (right) K_{eff} vs. σ_{other} for 3600 × 300 × 1 μm^3 aluminum bridge on silicon with λ_r = 1/3.

In case of $\lambda_r = 0$, Eq.(2.72) reduces to

$$K_{eff} = \frac{16Ebh^3}{L^3}, \quad (2.73)$$

the familiar expression for the beam stiffness for a centralized point load. If stresses get very large, the term with $\eta^2\cosh(\eta)$ will start to dominate the denominator in Eq. (2.70), which will result in K_{eff} linearly depending on the stress. Fig. 2.12 displays the dependence on the stress of the stiffness K_{eff}, calculated with Eqs. (2.70) and (2.71).

Fig, 2.12 clearly illustrates the abovementioned linear dependence of the stiffness on the stress and thus on the temperature variation.
It is mentioned once again that the above analyses assumes the load P is uniform over the actuation area l_c. As seen in subsection 2.2.2, more accuracy can be introduced by using a load linear in height. In Fig. 2.10, a load function as used in Eq. (2.51) applies at the beam section above the electrode. The general solution for the beam deflection at this actuation area reads:

$$w_{act}(x) = A\cosh g_1 x + B\cos g_2 x - \tfrac{1}{2} z_0 . \tag{2.74}$$

At the flanks, $P = 0$ applies.

$$w_{null}(x) = C\sinh(\tilde{k}x) + D\cosh(\tilde{k}x) + Ex + F ; \tag{2.75}$$

$\tilde{k}$, g_1 and g_2 are in this particular context defined as

$$g_1^2 = \tilde{k}^2 + g_2^2 \equiv \tfrac{1}{2}\tilde{k}^2 + \tfrac{1}{2}\sqrt{\tilde{k}^4 + 4a^4} , \qquad \tilde{k} \equiv \sqrt{S/EI} . \tag{2.76}$$

Here, a is the voltage parameter as defined in subsection 2.2.2. The coefficients A, …, F are found by applying the boundary conditions $w_{\text{null}}(x = ½L) = 0$ and $w_{\text{null}}'(x = ½L) = 0$ and furthermore demanding continuity of $w(x)$ and its first, second and third derivative at the boundary of the actuation zone ($x = ½l_c$). The very complicated form of these coefficients is not reproduced here.

2.2.5 Flank-loaded beam

A variation is how a beam deflects if it is loaded by two parallel strips width s that are not connected to each other; the middle region is unactuated, see Fig. 2.13. This is representative for the silicon nitride samples described in section 3.2.

The simplest way is the following: Suppose a beam is loaded downwards with force density P_0 over a distance l_c. The trick is to further load the beam *upwards* with the same force density over the middle part of width $l_c - 2s$. The overlap cancels and the two strips do the trick. Note that the effective stiffnesses are not similar in both cases for they depend on l_c and $l_c - 2s$ respectively. In a formula:

$$y_{center} = \frac{P_0 l_c}{K_{eff1}} - \frac{P_0(l_c - 2s)}{K_{eff2}} . \tag{2.77}$$

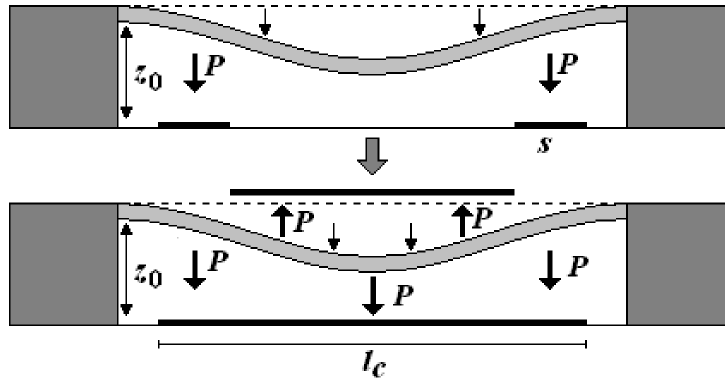


Fig.2.13 Flank-loaded beam, approximated by a two-sided centrally loaded beam

Also here applies: more precise is dividing the beam in actuated and unactuated regions and solving the beam equation for each of the regions. The solutions are then connected by applying proper constrictions.

A few summarizing remarks: For an accurate description of MEMS structures with a flexible bridge electrode, like the ones that will be described in section 3.3, the parallel-plate theory in section 2.1 should be replaced by beam deflection theory (see subsection 2.2.1). Clamped-clamped beam theory is firstly applied for an electric load spanning the whole beam. Various approximations to this nonlinear load are compared (subsection 2.2.2). Due to this nonlinearity, high voltages will result in unstable behavior for this bridge. An analytical estimate is made of this 'pull-in voltage' (subsection 2.2.3). To further accommodate to the experimental situation, axial stress is introduced. Variations in temperature are directly related to how they influence the beam stiffness. At last, the influence of partial loads (i.e. not spanning the entire beam length) is investigated.

2.3 BUILT-IN VOLTAGE AND PARASITIC CHARGES

Most of the in this thesis encountered capacitive MEMS, when characterized for bias voltage dependent characteristics (in particular capacitance), exhibit at least some offset with respect to the $V_{\text{bias}} = 0$ state. This offset, in many sources known as 'built-in voltage', ranges from less than 0.1 V up to several volts, depending on the design of the structure and the material composition. During experiments on the silicon nitride EMMA-samples (section 3.2), a few cases were encountered where smooth $C(V)$-curves ($-40 < V < 40$) were offset by more than 20 V. The first subsection (2.3.1) in the current section will discuss a few important sources for a voltage offset. Subsection 2.3.2 sums the main types of parasitic charges encountered in a dielectric layer. Subsection 2.3.3 mentions typical charge transport mechanisms that occur in dielectrics. Finally, subsection 2.3.4 summarizes the important points of this section.

2.3.1 Sources of built-in voltage

Because of the omnipresence of small built-in voltages, $C(V)$ – data have usually been fit with a quadratic expression allowing for this offset:

$$C(V) = \alpha(V - V_{bi})^2 + C_0\,, \tag{2.78}$$

where V_{bi} is the built-in, or offset voltage. Enlisted below are a few causes that can give rise to a built-in voltage. It is noted already that the 0.3 V amplitude, 800 kHz AC readout signal of the capacitance does not contribute to a built-in voltage; being far above resonance frequency (typically in the 1-10 kHz range), this means only the root-mean-square (rms) value matters. Taking the time average of the total voltage over the capacitor V_{tot} over many oscillations of its ac-component V_{AC} gives

$$\left\langle V_{tot}^2 \right\rangle = \left\langle (V_{DC} + V_{AC})^2 \right\rangle = V_{DC}^2 + 2V_{DC}\left\langle V_{AC} \right\rangle + \left\langle V_{AC}^2 \right\rangle.$$

As $\left\langle V_{AC} \right\rangle = 0$, we see that for $V_{\mathrm{DC}} = V$ and $\left\langle V_{AC}^2 \right\rangle \equiv V_{rms}^2$ we are left with Eq. (2.79):

$$C(V) = \alpha(V^2 + V_{rms}^2) + C_0\,, \tag{2.79}$$

in which we observe that the rms-voltage has the effect of adding a small, constant amount of αV_{rms}^2 to the capacitance, rather than producing an offset in the voltage.

1) Difference in work function.

If the electrodes are of dissimilar materials, their difference in work-function (also known as surface potential; the energy required to remove an electron from the surface) will give rise to an offset in the voltage, as demonstrated for various material combinations by VTT. These offsets range typically from 0 to 1 V and are consequently a factor to reckon with in the design of sensitive capacitive MEMS structures, which have to operate accurately. The offset this produces is however constant in time and in many practical cases a constant offset is acceptable, as the structure can be calibrated for it. Except for the EMMA silicon nitride structures, the devices investigated in this thesis all have similar electrode materials (aluminum).

2) Static parasitic charges

In many occasions the electrodes are covered by a dielectric material, either on purpose (usually relative thick, intentionally grown layers of at least a few tens of nanometers), or 'by accident' (usually the thin native oxide, for example 2-3 nm. Al_2O_3 covering all aluminum that has shortly been exposed to air). These dielectric layers provide sites ('states') in which charges (mostly electrons) get trapped. Those trapping states are most abundant at the surface, where the lattice of the dielectric is terminated and the unsatisfied bonds and many other defects are capable of hosting charges that either cross

the oxide or are caught from any alien material (gas, water, contamination) in contact with the exposed surface of the dielectric. Trapping states and transport will be more digressed on in subsection 2.3.2 and 2.3.3.
Consider now a MEMS layer stack with a surface charge on one of the electrodes:

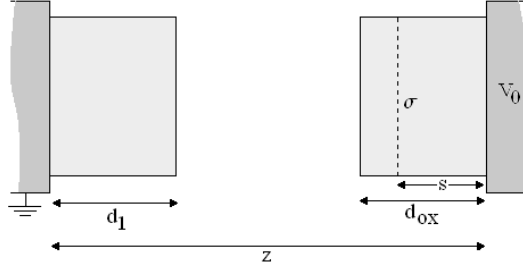


Fig.2.14 Metal electrodes (far left and right) with dielectric layers (thicknesses d_1 and d_{ox}, not necessarily equal) and a surface charge σ inside one of the layers, spaced s from the right electrode, put on a voltage V_0 with respect to the left electrode (grounded).

The electric field anywhere in the device can be calculated by solving the one dimensional Poisson equation Eq. (2.80):

$$\frac{d^2V}{dx^2} = -\frac{\rho}{\varepsilon\varepsilon_0}, \tag{2.80}$$

The contribution to ρ due to the surface charge is represented by $\sigma\delta(z - z_\sigma)$. The field just above the surface of the left dielectric is then:

$$E_{d1} = \frac{V_0 + s\sigma/(\varepsilon\varepsilon_0)}{z_0 - (d_1 + d_{ox})(1 - 1/\varepsilon)}. \tag{2.81}$$

The force on the left side dielectric can be calculated with Maxwell's stress tensor, which in this static, one dimensional case is simply $\overleftrightarrow{T}_M \equiv T_M = ½\varepsilon_0 E_{d1}^2$, where Maxwell's initial has been attached to the tensor symbol to avoid confusion with different quantities that may be represented by T, such as temperature. This force is calculated to be

$$F = A\overleftrightarrow{T}_M = \frac{1}{2}\varepsilon_0 A\left(\frac{1}{z - (d_1 + d_{ox})(1 - 1/\varepsilon)}\right)^2 \left(V_0 - V_{bi}\right)^2, \tag{2.82}$$

Here d_{eff} from Eq. (2.34) is recognizable. In Eq. (2.82), the substitution

$$V_{bi} = \frac{s\sigma}{\varepsilon\varepsilon_0} \tag{2.83}$$

has taken place. This expression Eq. (2.83) [11] has a typical dipole character, for s is (half) the distance of this charge to its oppositely charged image in the metal electrode. In

many cases, like in this thesis all MEMS structures provided with aluminum electrodes, this voltage is therefore quite small. However, in the case of silicon nitride samples (section 3.2), the dielectric layer thicknesses are comparable to the overall gap dimensions, when charge on the surface will have a noticeable effect due to Eq. (2.83). The dependence of the dielectric thickness does not mathematically appear in Eq. (2.83), but its importance is reflected in what upper limit is put on s. For a $C(V)$-curve, this force Eq. (2.82) means immediately that the minimum capacitance occurs not at $V = 0$ but instead at $V = V_{bi}$. The curvature α is also affected, although this effect is less noticeable.

3) Non-uniform static parasitic charges

The built-in voltages and parasitic forces mentioned so far hinge upon the uniformity of the parasitic surface charge density and above all that it is a *continuous* entity. Typical parasitic charge densities in sensory capacitive MEMS are in the range $10^{-3} – 10^{-7}$ C/m². At the top of this scale we find devices with thick dielectric layers, such as the silicon nitride bridge HF-sensor (section 3.2). In the lower range of this scale, only a few charges per square micron are equivalent to these charge densities. Consequently, the lateral variations (the spacing between the charges) are in the order of the vertical scale (the capacitive gap), namely ~ 1 μm. This ratio of dimensions marks the border of continuity, i.e. for closer spacings, the charges can be regarded as a continuous distribution, while if the charges are mutually more remote, effects of their discrete nature will start to prevail. Consequently, if charges are distributed in clusters, packing together several unit charges on practically a single location, the local field can be quite different from the case in which the same number of charges (same charge density) is spread more evenly over a surface. A study of single charge clusters by AFM is presented in Chapter 5 of this thesis. The following situation is of interest, though quite basic in principle. Assume a distribution of localized charge clusters in the form of a rectangular grid, Fig. 2.15:

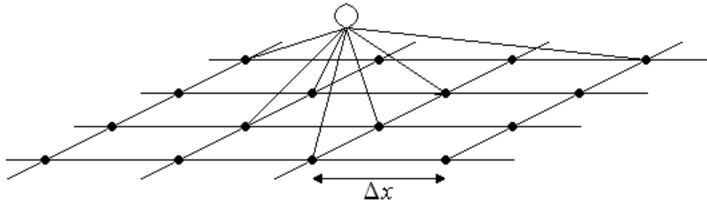


Fig. 2.15 Rectangular grid of identical charge clusters spaced Δx. The resulting electric field is probed at the position of the open circle (observer, test charge, Kelvin probe, etc), above the center of a square of four charge clusters.

The calculations for the electric field 1 μm above a grid of various charge cluster spacings, but constant charge density, are shown in Fig. 2.16. It is seen that if the charge clusters are spaced equal to the distance above which the field is probed, the discrete cluster distribution can just be considered 'continuous'. For larger spacings, the *local* field subsides. Obviously, the reverse is true if the local field is probed right above a charge cluster. With larger spacings, the number of charges per cluster grows in these calculations, to keep the average surface charge density constant. The field grows likewise. However, also in this case the transition to the continuous regime occurs at

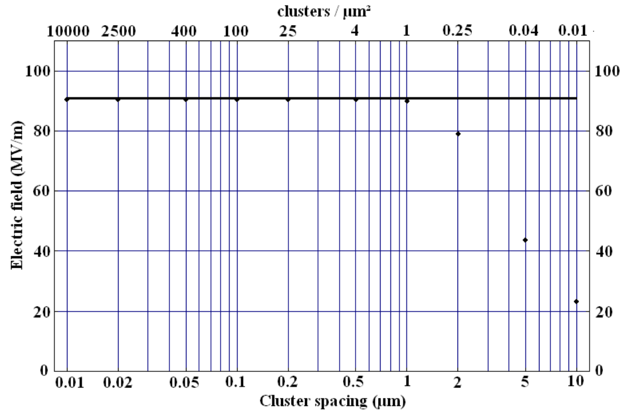


Fig.2.16 Electric field of a charge density of $1.6{\cdot}10^{-5}$ C/m² (100 unit charges per µm²) as a function of its distribution into clusters spaced 0.1 – 10 µm apart, containing 1 to 10^4 unit charges; The field is probed at the position indicated in Fig.2.15, 1 µm above the grid. The sample measured 2x2 mm², containing 400 to $4{\cdot}10^8$ clusters. The solid line indicates the field of a continuous surface charge $\sigma/2\varepsilon_0$.

spacings equal to the vertical observation distance. For closer spacings, the cluster size ceases to be of importance, neither the observation distance.

A perhaps surprising effect is that even if the net amount of parasitic charge is zero, a nonuniform distribution of charges is capable of producing parasitic forces. A simple schematic (Fig. 2.17) is given to clarify this:

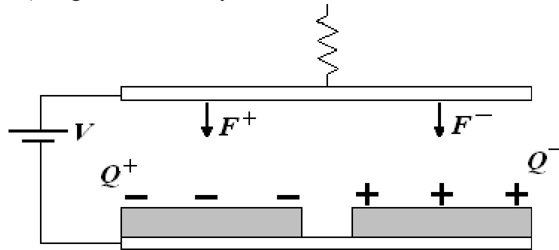


Fig.17 Capacitive MEMS with a dielectric layer, for clarity cut into two separate regions, with bipolar parasitic charging.

Both charged regions cause an attractive force on the top electrode, regardless their polarity. This force is present also with zero bias voltage, and with net *zero* charge ($Q^+ + Q^- = 0$). An expression for the resulting force is given in [12], based on [13]:

$$F = -\frac{\varepsilon_0 A}{2}\frac{(V-\overline{V}_{bi})^2 + \mathrm{var}_{Vbi}}{(d/\varepsilon + g)^2}. \tag{2.84}$$

Here $\overline{V}_{bi}$ the mean built-in voltage (averaged over a lateral position dependent built-in voltage).

var_{Vbi} is the variance[14] of the built-in voltage. d and ε are thickness and dielectric constant of the dielectric layer and g the gap between top electrode and the surface of the dielectric layer ($= z - d$ in many expressions in this thesis).It is remarked that for a nonzero variance, the built-in voltage can not be compensated by a bias voltage V, such that the electrostatic charge is force out, even though the net charge is zero.

4) Mobile charges

Until now only static features have been considered. Meanwhile, the measurement of a $C(V)$ relation is a dynamic, time-extended process, in which ideally only the position of the movable electrode varies, as a function of bias voltage that is tuned (linearly) in time. However, during the process, various parameters may change. At first, the static capacitance may change significantly during the process of a single curve measurement, because of drift in the electronics or mechanical drift. Fitting the data with a quadratic function will produce a (small) offset in the voltage, as is demonstrated in a simulation in subsection 3.2.4.

In this context, we consider what can happen if the amount and location of trapped charges in the dielectric layers is varying. In the first place, the average total trapped charge can drift slowly, on timescales of several curves up to many curves. Furthermore, within the duration of a single curve measurement, significant amounts of charge can trap and de-trap.

For a rough sketch of the phenomena, a simple simulation has been carried out, in which charge is allowed to accumulate on the surfaces of thin dielectrics through ohmic conduction through the dielectric. This accumulation and de-cumulation happens during a series of $C(V)$ curves with alternating bias voltage sweep directions. This kind of measurements is employed in practice as reported in Chapters 3 and 4.

The device consists of two parallel electrodes (for example aluminum) and a thin (native) aluminum oxide on top. Fig. 2.14 is a schematic, with $d_1 = d_{ox} = s = 2$ nm and a rest gap $z_0 = 1$ µm. For the dielectric constant $\varepsilon = 9.1$ and the spring constant $k = 500$ N/m have been taken. These values constitute a representative capacitive MEMS device. The assumed resistivity $\rho = 10^{10}$ Ωm is lower than reported values ($10^{12} - 10^{14}$ Ωm) in order to enhance the effects intended to demonstrate. Two reasons justify this kind of modification:

1) The thin film dielectric constant is expected to be lower than the bulk dielectric constant. Therefore, the field that trapped charges experience is less attenuated. The decreased resistivity compensates for the too high value of 9.1.

2) The dominant transport mechanism in films this thin is direct tunneling from the metal to the surface, not ohmic conduction. In a measurement, considerably more charge will accumulate on the surface than can be expected from ohmic conduction alone.

In the simulation presented here, zero initial trapped charge is assumed. Furthermore, it is assumed that all charge accumulates on the surface only. When a curve is measured, it is divided in time steps, corresponding to voltage increments. During a time step, a current density J flows through the dielectric, which causes charge to pile up on the surface. The trapped charge increment $\Delta\sigma$ during a time step Δt is calculated by Eq.(2.86):

[14] The common notation for variance, σ^2, has been abandoned to avoid confusion with surface charge density

$$\Delta\sigma = J\Delta t = \frac{1}{\rho} E\Delta t\,. \tag{2.85}$$

Two contributions have been taken into account for the electric field E a trapped charge in the dielectric experiences: The bias voltage (attracting charge towards the surface), and the trapped charge on the surface (repelling ‘new’ charge back and causing outflow of charge in absence of bias voltage). A result is shown in Figs. 2.18ab:

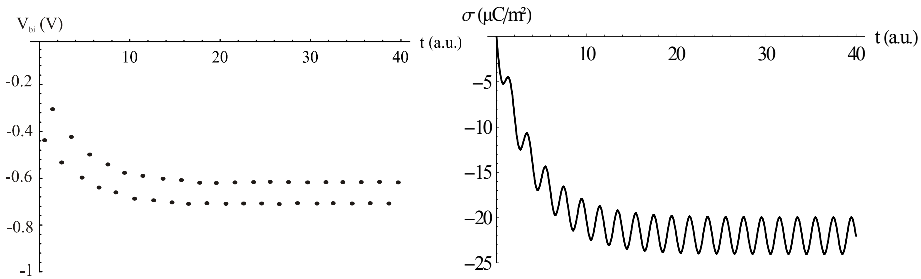


Fig. 2 18a (Left). Simulation of built-in voltage vs. time (alternating sweep-directions), with mobile parasitic charges; initial surface charge is zero. One point represents the built-in voltage, taken from a quadratic fit to C(V_{bias}) data for –5 V < V_{bias} < 0 V.
Fig. 2.18b (Right). The corresponding surface charge density

Summarizing the results of these simulations learn:

1) If the amount of parasitic charge changes significantly within one curve, a nonzero offset in the voltage is produced, even if initially no parasitic charge is present and even if the bias voltage range is symmetric around 0 (in Fig. 2.18 it is *a*symmetric).
2) For sufficiently mobile charge, a curve running backwards immediately after the first forward curve produces an offset on a different location.
3) The difference between forward offsets and backward offsets is constant in time.
4) The offset voltages and the surface charge density oscillate around an average value that stabilizes after a certain time.
5) Nonzero average values for offset voltage and surface charge result from asymmetric bias sweeps and net fixed initial charge
6) A voltage offset from a single curve is not a reliable measure for the amount of parasitic charge. Instead, a series of alternating forward-and backward biased curves reveals any net permanent charge, which is represented by the level around which the voltage offset oscillates. This is a lower limit, which may represent the charge if all were concentrated on the surface of the dielectric, where its influence is strongest. It is conceivable that in fact more charge is present in the dielectric, but stored in the bulk, where its influence is partly shielded by the surrounding dielectric and the proximity of their counterpolar images in the neighboring conductor.
7) A low spring constant tends to increase a voltage offset, as the device becomes more sensitive to forces resulting from parasitic charge.

The defects of the applied model are obvious and numerous: Firstly, though ohmic conduction does happen in dielectric layers, tunnel effects are much more important, of which a couple are mentioned in subsection 2.3.3. If the idea of an image charge is still applicable on the nanometer scale, it is clear that already a tunneling mechanism is necessary for a charge to enter the oxide, not only because of the natural barrier of the material interface, but also to ‘decouple’ a candidate parasitic charge from its image. After entering the oxide, the charge still needs to find his way to the surface, though its influence is already appreciable while still in the bulk of the oxide.
Still, the simulations are quantitatively not at all doing badly; values for built-in voltage and splitting are in fact quite typical for what is observed in measurements of $C(V)$ curves with symmetric bias. Much more importantly, they highlight qualitatively some important aspects that have qualitatively been observed in $C(V)$ measurements presented in chapters 3 and 4. They demonstrate the difficulty of associating a static concept like ‘built-in voltage’ to a dynamic and unstable quantity like ‘trapped charge’. Or, more precisely, assigning built-in voltage an instantaneous significance is valid, but a one-to-one association with the voltage-offset in a $C(V)$-curve is conceptually doubtful. From a single $C(V)$-curve it is therefore problematic to accurately estimate the amount of trapped charge.

2.3.2 Types of trapped charges in dielectric layers [14]

Charges that exist in a dielectric such as in a MEMS capacitor are of different nature, dependent on their origin. Generally, five groups of them have been distinguished in literature: a) surface charge, b) mobile ions, c) fixed dielectric charge, d) trapped bulk charge and e) trapped interface charge. They are depicted together in Fig. 2.19.

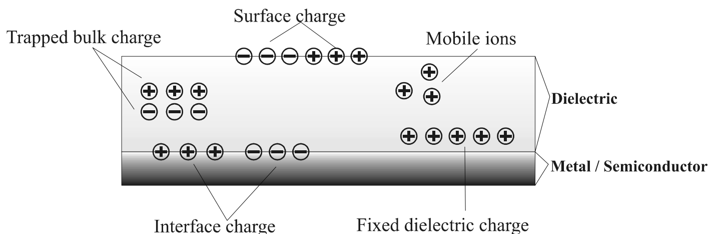


Fig.2.19 various types of trapped charges

The surface charge (a) on top of the dielectric can have many origins and is trapped by the impurities in the crystal lattice that are inherent to a surface. They can consist of holes, lattice imperfections, foreign atoms, etc. In the bulk material there can also be foreign, mobile ions (b) present that get there via imperfections in the fabrication process. Dedicated fabrication can reduce their number. The fixed dielectric charge (c) in the bulk of the dielectric can be present when the dielectric has been made by oxidation of the underlying electrode layer. It is located near the interface and are thought to originate from the quite abrupt ending of the oxidation process, when some ionized silicon atoms remain. The trapped bulk charge (d) sits firmly in lattice impurities of the dielectric. In the forbidden gap there are some states formed by the impurities, allowing a few charge carriers to enter the dielectric. The trapped interface charge (e) finally does not influence

the capacitor's behavior if the underlying electrode is a metal or highly doped semiconductor, for it then induces charges there so that the plate is an equipotential at a value that is brought about by the external voltage source.

2.3.3 Charge transport

As stated earlier, charge trapping starts to form a problem for accurate MEMS operation by being unstable, i.e. during operation a particular configuration of trapped charges throughout the device changes, which crosses the possibility of calibrating a device to include a (stable) parasitic electrostatic force. In order to create this instability, charges have to move throughout the dielectric. Many studies have been performed on Metal-Insulator-Semiconductor (MIS) – structures, more specifically MOS (O = Oxide). A few notable mechanisms that have been identified are mentioned below. Not all mechanisms as described here are immediately applicable, as a MEMS structure categorizes as a MAIS or MAIM (Metal – *Air* – Insulator – Semiconductor / Metal) stack, rather than MIS. If charges are transported from the conductor below the dielectric through the oxide, they end up in surface trap states rather than a metal. This has consequences for barrier heights and tunnel probabilities. The mechanisms themselves remain valid though.

Ohmic conduction

The most important difference between a semiconductor and an insulator is the width of the forbidden gap. For typical semiconductors like (undoped) silicon, the gap is in the order of 1 eV. At room temperature, a small but significant number of charges are excited from the valence band into the conduction band. For insulators, this band gap is at least several electron volts, which means a sharp decline in the number of available charge carriers. Though insulators do allow for tiny ohmic conduction ($\rho > 10^{12}$ Ωm for Al_2O_3), it claims only little terrain in the field of charge transport through insulators.

Schottky effect

The Schottky effect is one of the mechanisms describing the transition of the barrier between a metal and a semiconductor. It is based on the process of thermionic emission of charge carriers from the semiconductor into the metal or reversed. When a metal and a semiconductor are connected, band bending raises a barrier between the metal and the conduction band of the semiconductor. At sufficiently high temperatures (room temperature), some charge carriers (electrons for n-type semiconductors) possess the energy required to take the hurdle, which produces a current. In equilibrium, the forward current (forward as defined from semiconductor to metal) and the reverse current are equally large. A bias voltage lets one of the currents dominate, as an asymmetry arises in how high the barrier is perceived by the charge carriers on either side. In case of a forward bias, the Schottky current is given by

$$J = A_R T^2 e^{-e\phi_{Bn}/k_B T} \left(e^{eV/kT} - 1\right), \tag{2.86}$$

according to [15]. Here A_R denotes the *Richardson constant,* which depends on the effective mass of the dominant carrier type. T is the temperature in K, ϕ_{Bn} represents the barrier height, $k_B = 1.38 \cdot 10^{-23}$ J/K is Boltzmann's constant, e is the unit charge and V is the forward bias voltage. At room temperature, $kT \sim 0.026$ eV, so that for 1 V the '–1' term can be neglected. For conductivity *through* an oxide layer (rather than a M/S-barrier), other sources [16] provide for the Schottky current the following expression:

$$J = A_R T^2 e^{(\beta_{Sc}\sqrt{E} - e\phi_{Bn})/k_B T} . \tag{2.87}$$

Here, E is the electric field in the oxide. β_{Sc} is known as the *Schottky field attenuation factor*, given by

$$\beta_{Sc} \equiv \sqrt{\frac{e^3}{4\pi\varepsilon\varepsilon_0}} , \tag{2.88}$$

with ε the relative dielectric constant of the oxide.

Poole – Frenkel Transport

Closely related to Schottky conduction is Poole-Frenkel [17] conduction, see also [16]. This mechanism applies in oxide layers containing deep traps in the forbidden gap of the oxide, specifically for traps that are positively charged when empty and neutral when filled. Thermal excitation alone is not sufficient to excite a significant population of electrons to the conduction band, for the coulomb barrier, preventing the electron to be separated from the trap, is too high. However, a low electric field (typically 10^7 V/m, often written as 10^5 V/cm) lowers this coulomb barrier, so that the electrons can indeed be excited. The current due to Poole-Frenkel transport is very similar to Eq. (2.87):

$$J = J_0 e^{(\beta_{PF}\sqrt{E} - e\phi_{PF})/k_B T} . \tag{2.89}$$

Here, J_0 is the low-field current (*ohmic*, so proportional to E) of the oxide, ϕ_{PF} is the depth of the traps and the *Poole-Frenkel field attenuation factor* is simply given by $\beta_{PF} = 2\beta_{Sc}$. Because the charges drift along ohmic transport, Poole-Frenkel transport will likely be dominated by different mechanisms in case of for insulators with very high resistivity such as Al_2O_3.

Direct tunneling

When a metal or a semiconductor is covered with a very thin dielectric such as its native oxide (typically 2-3 nm), charge carriers can tunnel from the (semi)conductor, through the barrier layer to the surface, and back. In equilibrium, the forward and backward tunnel currents balance.

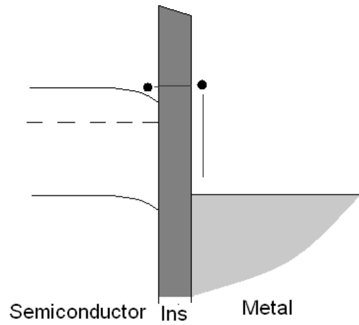


Fig. 2.20 Direct tunneling

This has however the effect that some net amount of charge will be permanently stored on the surface: negative for metals and n-type semiconductors and positive for p-type semiconductors [18].

Field emission: Fowler-Nordheim Tunneling

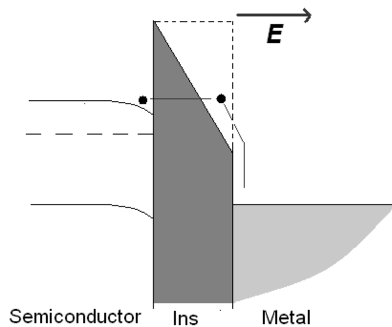


Fig.2. 21 Fowler-Nordheim tunneling

The theory of Field emission (FE) was first accurately described by Sir Ralph Fowler and Lothar Wolfgang Nordheim in 1928 [19]. It deals with the phenomenon of electrons being ejected from metal into vacuum, or into different media such as air and solid materials connected to the metal surface. The process is essentially different from thermionic emission, for here the electrons have to be excavated all the way from the Fermi level to the energy of the work function (in case of emission into vacuum) or barrier height (into any medium). A strong electric field (typically 10^8 V/m) is capable of doing so. Unlike thermionic emission, Fowler Nordheim transport is essentially a pure tunnel effect, capable of sending electrons right through an oxide (in [19], up to 13 nm has been studied) to an opposing metal or semiconductor. Furthermore, FN-tunneling is also present at cryogenic temperatures. In fact, the theoretical temperature dependence is essentially negligible between 0K and room temperature. In its original and simplest appearance, the Fowler-Nordheim current is described by:

$$J = \frac{e^3}{8\pi h \phi} E^2 \exp\left(-\frac{8\sqrt{2m}\phi^{3/2}}{3\hbar e} \frac{1}{E} \right), \tag{2.90}$$

ϕ is the barrier height. The electron mass is recognized in m and finally Planck's constant h and $\hbar \equiv h/2\pi$ are seen. Characteristic for Fowler-Nordheim tunneling is the strong dependence on the electric field E, which appears quadratically in the front factor and in the exponential as well. Note also the dependence on the barrier height.
An important modification arises if the image force of the electron in the oxide is taken into account, which attracts the electron towards the surface and lowers the barrier. The interested reader is referred to [20]. A second modification is encompassing temperature dependence. At nonzero temperatures, some electrons acquire energies above the Fermi level. Up to room temperature, this effect enhances the current by typically a few per cent only.

Hopping conduction

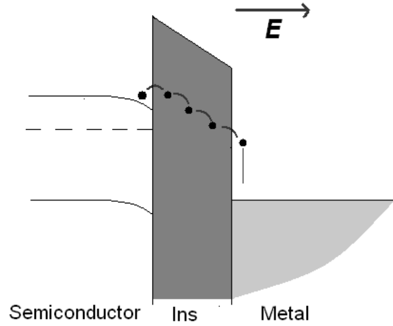


Fig. 2.22 Hopping conduction

Especially for thicker oxides, this mechanism is the main means of carrier transport through the bulk dielectric. Traps spread throughout the material, energies in the oxide forbidden gap, form 'stepping stones' for the charge carriers between which they move through tunneling. Mott [21] developed "Variable Range Hopping" to describe the dynamics of charge carriers in a bulk dielectric with traps. With "Range" the 'four-vector distance' Norm[(Δx, Δy, Δz, Δϕ)] is indicated. The space coordinate distance and the difference in trap depth energy from one trap to neighboring traps are treated on equal footing. For pure amorphous dielectrics, these quantities are theoretically independent. The temperature dependence of hopping conductivity S[15] depends on the the number of dimensions N_d in which the system is studied:

$$S = S_0 e^{-(T_0/T)^{\frac{1}{N_d+1}}} \quad . \tag{2.91}$$

For one-dimensional transport (along a wire), $N_d = 1$. Sheet transport gives $N_d = 2$, while volume currents, as in the case of bulk transport through dielectric layers, are characterized by $N_d = 3$.

[15] The commom symbol for conductivity σ is avoided here in a context dealing with surface charges.

2.3.4 Summary

Various sources of the offset voltage of a $C(V)$-curve have been mentioned: difference in work function, static parasitic charges, non-uniform parasitic charges, capacitance drift and mobile charges. It is well-founded to associate the concept of 'built-in voltage' with the first three (static) contributions. The latter two (dynamic), most notably the mobile charges, do not lend themselves readily to the static concept of a built-in voltage. The voltage-offset of a single $C(V)$-curve does generally not constitute a reliable value for the built-in voltage. Under stable operating conditions, alternately directed sweep series will split the built-in levels. When stabilized, their average does provide a workable value. Static trapped charges contribute to the built-in voltage in three ways, in order of importance: 1) interaction with charges on the opposite side of the capacitive gap, 2) with the electric bias field 3) by polarization of the opposing dielectric layer and interaction with it.
Trapped charge can be of varying nature and origin. Their abundance depends on the design of the device and used materials, and the fabrication process, in which are of influence on the number and depth of traps. The surface of a dielectric layer is an especially attractive region for trapped charge, as here the lattice is terminated and 'imperfections' are hence amply abound. However, especially amorphous materials also contain many bulk traps.
Of the mentioned mechanisms that transport charge to / through an insulating layer, direct tunneling for native oxides (section 3.3, chapter 4) and hopping conduction for thicker dielectrics (section 3.2, section 5.3) are on general physical grounds regarded as the leading processes for charging and discharging (sections refer to experimental situations, described in this thesis, to which they apply). In both cases, the electric field biases the charging of the surface. Ohmic conduction, Shottky and Poole Frenkel cannot be excluded to play a role however for slow variations in the built-in voltage. Fowler-Nordheim tunneling should be of minor importance, as because of the air gap (not present in MIS-structures), only a fraction of the total applied voltage spanning the oxide, and with it the electric field, while Fowler Nordheim tunneling starts to join the table only for high electric fields.

2.4 ELECTROMECHANICAL RESONANCE

Chapter 4 contains experiments of determining the (fundamental) resonance frequency of capacitive MEMS samples. This analysis serves the mechanical characterization of these samples, in particular the spring constant, which tells something about the sensitivity of the device. This has been done by electrostatic excitation of a mechanical vibration and the electronic readout of its response. It entails that a connection has to be laid between the electrostatic and the mechanical domain. Note that the term "electrostatic" is used to stress that the actuation is performed by voltage control: the applied force is due to static charge distributions. On the other hand, current control is associated with "electrodynamic": the applied force (magnetic) is due to moving charges (currents).

2.4.1 Equivalent circuits

In this subsection a simple and intuitive description of the electomechanical coupling will be provided. A rigorous treatment can be found in [22], in terms of the so-called "one-port desription", which covers the response of a resonator to an applied signal. The device under test (DUT) is characterized by a finite set of "state variables" q_i and communicated with through "ports". The communication through such a port is described by a pair of port variables, called "effort" (*e*) and "flow" (*f*). The "effort" is the signal the experimentalist applies to the DUT, which meets a response that is related to the effort. This response, or *flow f*, is defined as the rate at which the state of the DUT, at which the effort is aimed, is changing; in short, the time derivative dq_i/dt. We will consider here only the one-port description. Resonant behavior may be alternatively probed using a two-port description, which involves determining transfer functions of the system.

Let us first consider a driven mass-spring system with viscous damping[16] (Fig. 2.23), which motion is governed by (Eq. 2.92):

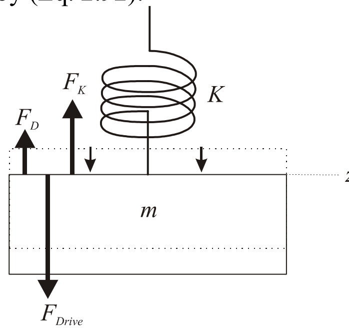


Fig. 2.23 Oscillator of mass m and stiffness K driven by a force F_{drive},, counteracted by a damping force F_D, added to which a spring force F_k that attempts to pull the bridge back into equilibrium.

$$m\frac{d^2z(t)}{dt^2} = -K(z_0 - z(t)) - D\frac{dz(t)}{dt} + F(t)\,. \tag{2.92}$$

Gravity is neglected here, although in some sensitive designs it should be taken into account. It would merely shift the equilibrium position, but if this is unwanted the sensor should be tilted so that gravity does not act in the sensitive direction (in Fig. 2.23 normal to the plate), or be equipped with a compensation mechanism (in which case the stiffness may be affected).
Here, $z(t)$ is the deflection of a harmonically suspended mass m with respect to its equilibrium position z = 0 and can be regarded as the relevant state variable of this system. The driving force F is the effort the experimentalist puts into the system. The spring, with stiffness K, attempts to affect the state of the system such that equilibrium is restored. The motion is damped with a constant D and is always opposed to the velocity;

[16] Other damping mechanisms are not considered here, such as drag force, which is proportional to the square of the velocity.

in other words, it invariably counteracts the flow. What is left from the driving force is the net rate at which the flow is increasing, or the acceleration.
In the special case without damping and driving force, upon sharp and short excitation the mass will oscillate at its free resonance angular frequency

$$\omega_0 = \sqrt{\frac{K}{m}} \qquad (D = F = 0). \tag{2.93}$$

Next we assume moderate viscous damping (under-damping), that is

$$0 < D < 2\sqrt{mK} \ . \tag{2.94}$$

For $D \geq 2\sqrt{mK}$, the system is said to be critically damped (=) or over-damped (>), in which cases no oscillation takes place and the mass slowly and asymptotically returns to its equilibrium position. Under-damping affects harmonic motion firstly by causing the amplitude to exponentially decay in time and secondly by lowering the angular frequency of free resonance:

$$\omega_D = \sqrt{\frac{K}{m} - \frac{D^2}{4m^2}} = \omega_0\sqrt{1 - \frac{D^2}{4Km}} \ . \tag{2.95}$$

In case of the application of a sinusoidal force finally, the oscillator will ultimately assume a harmonic motion with a frequency imposed by the driver, for the motion at own resonance frequency will be damped out in the long run. The phase difference between force and motion and the amplitude will depend on the applied frequency. The amplitude will be maximal for a force driving at the damped resonance frequency ω_D.
Next, we consider a seemingly totally different problem, namely that of the description of an electric circuit containing in series the passive, linear elements resistor *R*, capacitor *C* and inductor *L*. The current through the circuit obeys:

$$L\frac{dI(t)}{dt} + \frac{1}{C}\int_{t_0}^{t} I(t')dt' + RI(t) = V(t) \ . \tag{2.96}$$

For the purpose of assumed easier recognition it is chosen to write this equation in terms of the flow-variable, the electric current *I*, which is the time derivative of the state-variable electric charge *Q* via $I(t) = dQ(t)/dt$. The effort variable $V(t)$ distributes itself over the linear elements in the familiar way. When comparing the equations it is quickly recognized that the state variable (position and charge respectively) is influenced by similar time-dependencies. It lends itself in a natural way that the electric actuation and readout of a mechanical resonator can be descibed in pure electric terms, when apart from the state variables and their time derivatives, the following associations can be made:

$$L \rightarrow m, \qquad C \rightarrow 1/K, \qquad R \rightarrow D \tag{2.97}$$

These associations are not unique, but are the most natural ones, especially for a voltage-driven experiment. Care must be taken that no ‘=’-sign is read here. The proper translation between the electric and the mechanical domain is taken care of by a (bias voltage dependent) transduction factor Γ, which can be derived from establishing the actual connections between the port variables in the electrostatic domain and the mechanical domain. Let us start with comparing the effort variable, in concreto: what force is actually acting on the movable electrode as a function of applied voltage. This voltage consists of an AC and a constant DC part:

$$V = V_{DC} + V_{AC}(\omega t)\,. \tag{2.98}$$

This causes an electric force of magnitude:

$$F_{el} = \frac{\varepsilon A}{2z^2}\left(V_{DC}^2 + 2V_{DC}V_{AC}(\omega t) + V_{AC}^2(\omega t)\right). \tag{2.99}$$

The first term in this equation determines the position of the movable electrode z, which will constitute the equilibrium position around which the oscillatory motion takes place. The second term governs this actual harmonic motion at angular frequency ω. If the amplitude of the alternating voltage is much smaller than the applied bias voltage, the third term, driving the movable mass at 2ω, can be neglected in comparison with the former two. If we divide the force in a static and a motional part,

$$F_{el} = F_{stat} + F_{mot}\,, \tag{2.100}$$

we identify

$$F_{stat} = \frac{\varepsilon A}{2z^2}V_{DC}^2\,, \tag{2.101}$$

and, pulling through the aforementioned neglect of V_{AC} with respect to V_{DC},

$$F_{mot}(\omega t) \approx \frac{\varepsilon A}{z}V_{DC}V_{AC}(\omega t) = \Gamma V_{AC}(\omega t)\,. \tag{2.102}$$

The latter expression constitutes the driving force of the harmonic oscillation. The *transduction factor* Γ, coupling the mechanical domain to the electrostatic domain, is then defined as

$$\Gamma \equiv \frac{\varepsilon_0 A}{z^2}V_{DC}\,, \quad [\Gamma] = \mathrm{C/m}. \tag{2.103}$$

The flow variables can be connected by examining the current through the DUT. For this, we consider only the AC part of the current; the DC – part obviously very quickly dies out in time. The current I_{AC} (*flow*) is the time derivative of the amount of charge Q_{AC} (*state variable in the electrostatic domain*) that is stored on either capacitive electrode. Performing the derivation

$$I(t)=\frac{dQ(t)}{dt}=\frac{d(C_M(x,t)V(t))}{dt}=C_M(x,t)\left[\frac{dV_{AC}(\omega t)}{dt}+\underbrace{\frac{dV_{DC}(t)}{dt}}_{\to 0\,\forall t>>R_{\text{int}}C_M}\right]+V(t)\frac{dC_M(x,t)}{dt}$$

$$\equiv I_{AC}(t)+I_{mot}(t)\,, \tag{2.104}$$

we observe that part of the current I_{AC} stems from the ordinary AC probing of a static capacitor and part I_{mot} results from the motion of the suspended electrode that drives charges through the circuit. The dependence of the capacitance on the position of the top electrode has been made explicit. The time derivative of the DC voltage over the capacitor vanishes quickly; for typical devices with a capacitance of a couple of pF and a typical internal source resistance of 50 Ω, saturation is reached in less than nanosecond. If we unravel the motional current a bit further and again neglect the AC-component of the applied voltage ($V(t) \approx V_{DC}$), we get:

$$I_{mot}(t)\approx V_{DC}\frac{dC_M(z,t)}{dt}=V_{DC}\frac{dC_M(z,t)}{dz}\frac{dz}{dt}=V_{DC}\frac{\varepsilon_0 A}{z^2}\frac{dz}{dt}=\Gamma\frac{dz}{dt}\,. \tag{2.105}$$

The relation of the electrostatic flow variable I to the mechanical flow variable $v = dz/dt$ has now been established. By inserting $F(t) = F_{mot}(t) = \Gamma V_{AC}(\omega t)$ and $dz/dt = I_{mot}(t)/\Gamma$ into Eq. (2.92), we can further concretize the associations Eq. (2.97), arriving at:

$$L = m/\Gamma^2 \qquad C = \Gamma^2/K \qquad R = D/\Gamma^2. \tag{2.106}$$

The electric treatment of the mechanical resonance behavior of the DUT can be thought according to the following schematic (Fig. 2.24), known as *equivalent circuit*:

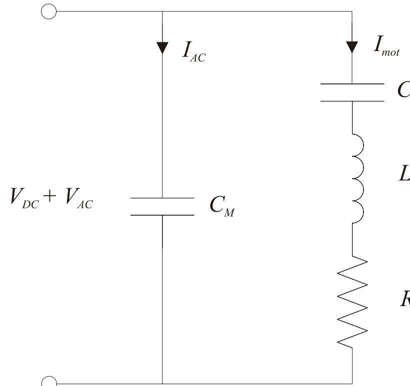


Fig. 2.24 Schematic of equivalent circuit.

Here C_M denotes the static MEMS capacitance, which is dependent on the bias voltage. Note that this circuit applies only at the fundamental resonance mode, and contributions of higher order resonances are assumed negligible. Higher resonance modes can be incorporated by adding appropriate RLC branches in parallel to the first one, which are usually much more impedant than the first mode.

It is convenient to consider the admittance $Y(i\omega) = I(i\omega) / V(i\omega)$ of the system, rather than the impedance Z. Starting from the total impedance of the parallel circuit (in which $Z_c = 1/i\omega C$, $Z_L = i\omega L$ and $Z_R = R$), the admittance $Y = 1 / Z$ comes out at:

$$Y(i\omega) = Y_M(i\omega) + Y_{mot}(i\omega) = i\omega C_M + \frac{i\omega C}{H_s(i\omega)}, \tag{2.107}$$

in which

$$H_s(i\omega) = 1 + \frac{1}{Q_s}\left(\frac{i\omega}{\omega_s}\right) + \left(\frac{i\omega}{\omega_s}\right)^2 \tag{2.108}$$

represents the second order frequency response function. In here,

$$\omega_s = \frac{1}{\sqrt{LC}} \qquad \text{and} \qquad Q_s = \frac{1}{R}\sqrt{\frac{L}{C}}. \tag{2.109}$$

are the series resonance frequency and the quality factor associated to this resonance. This topic will be returned to in the next subsection.

2.4.2 Characteristic frequencies and parameters

In this subsection important characteristics of the near-resonance behavior of the admittance will be discussed and how mechanical characteristics can be extracted from them. In an electromechanically coupled resonator, several characteristic frequencies can be distinguished. Firstly, there is the angular frequency ω_s that minimizes the impedance of the motional part (hence called *series resonance frequency*), or equivalently that maximizes the conductance G, the real part of the total admittance:

$$G(\omega) = \mathrm{Re}(Y(i\omega)) = \frac{R}{\left(\frac{1}{\omega C} - \omega L\right)^2 + R^2}, \tag{2.110}$$

so that

$$\omega_s = \frac{1}{\sqrt{LC}}, \tag{2.111}$$

and the peak height χ of the resonance becomes:

$$\chi = \max\left(G(\omega)\right) = \frac{1}{R}. \tag{2.112}$$

The width of the peak $\Delta\omega$ is defined as the spacing between the frequencies at which the conductance is half of its peak height. Solving for $G(\omega) = (2R)^{-1}$ yields four solutions for ω. The two positive angular frequencies are given by

$$\omega_{\pm} = \frac{\pm R + \sqrt{R^2 + 4L/C}}{2L} \tag{2.113}$$

so that

$$\Delta\omega \equiv \omega_{+} - \omega_{-} = \frac{R}{L} \ . \tag{2.114}$$

The quality factor of a resonance is defined as the 2π times the ratio of the energy dissipation in one oscillation period to the total energy in an oscillation. A good estimate to this is provided by

$$Q_s = \frac{\Delta\omega}{\omega_s} = \frac{1}{R}\sqrt{\frac{L}{C}} \ , \tag{2.115}$$

which is what is already encountered in Eq. (2.109).
Of interest is also the *parallel resonance frequency,* which is defined as the frequency that establishes constant charge at the terminals, i.e. (almost) no current is running in response to an oscillating voltage. It lies very close to the frequency that *minimizes* the absolute total admittance. It can be calculated by neglecting R^2 with respect to ωL, which is satisfied when $\omega \cong \omega_s$ (near-resonance) and $Q_s >> 1$. Then it follows from $\frac{d}{d\omega}\left|Y(i\omega)\right|_{R^2/\omega L \to 0} = 0$:

$$\omega_p \cong \omega_s \sqrt{1 + \frac{C}{C_M}} \ . \tag{2.116}$$

In the complex plane the locus of the admittance sweeps out a circle near resonance angular frequency, provided that the damping coefficient is independent of frequency so that R is a constant (otherwise the circle would become flattened). The resonant behavior for a low Q and a high Q resonator are provided in Figures 2.25, with parameters identical to the examples provided in [22] that are chosen such that a couple of characteristics of near-resonant behavior of a capacitive MEMS is well illustrated.
We observe that the radius of the circle is proportional to the *inverse* of the resistance R and thus, L and C kept constant (corresponding to identical mass and spring constant, implying identical resonators), proportional to the quality factor Q. Furthermore we observe that for high quality factors the resonance circle can intersect the real axis.

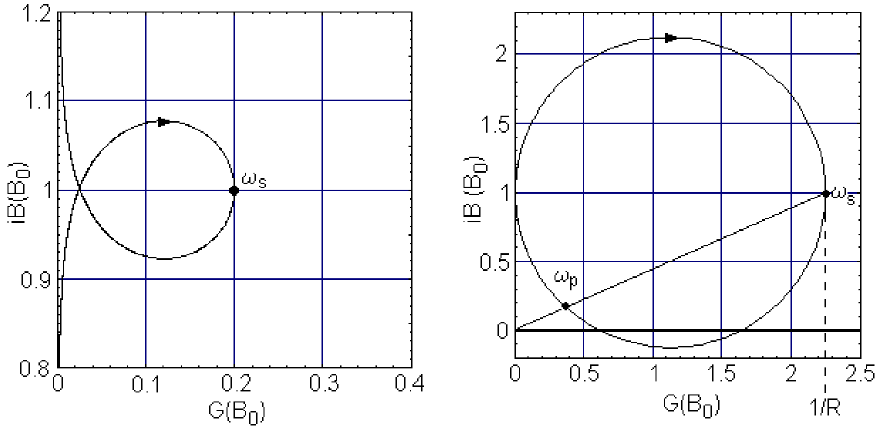


Fig. 2.25: Low-Q (Q = 20, left) and high-Q (Q = 225, right) locus plots of the near-resonance admittance for $C_M/C = 100$. The scale of conductance G and susceptance B is in units of $B_0 \equiv \omega_s C_M$. Arrows indicate the direction of increasing frequency. The series resonances ω_s are drawn on their exact spots, while for the high-Q plot the parallel resonance angular frequency ω_p is constructed very near to its actual location. Note the difference in scales between the two plots. Admittance points for successive fixed steps in frequency start densely packed around the origin and become more and more dilute near resonance, after which the point density quickly increases as soon as the 'circle' is left.

If the parameters *R*, *L* and *C* can be determined, information about the mass[17] and the stiffness can be retrieved. We may for example invoke the conductance, see Fig. 2.26:

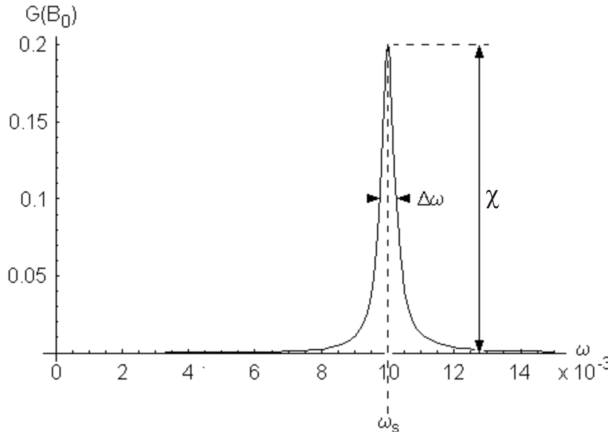


Fig. 2.26 Conductance for a resonator with $C_M/C = 100$ and Q = 20; for R, C and L the arbitrary choices R = 5, $L = 10^4$ and C = 1 have been made. ω_s, $\Delta\omega$ and χ are specified by Eqs. (2.111), (2.114) and (2.112)

[17] When the mass (density times volume) of the suspension can be neglected with respect to the resonating object, such as the double beam structures described in Chapter 4, the concept of mass does not present any difficulty or ambiguity. In case of resonating clamped-clamped beams and cantilevers, the inertial mass associated with resonance ("generalized mass"), though well-defined, may differ from the gravitational mass of the total structure.

A collection of data points will need a parameter fit for $G(\omega)$. This can be a cumbersome task, as the involved parameters can attain large ranges of values and additionally very unnatural compared to typical values of components used in electronics. Especially far from 'household practice' are values for a typical motional inductance L. In Chapter 4 L = 200 kH (kilo-Henry) is reported, while dedicated resonators with high quality factors can easily reach several MH

Providing useful starting values for the parameter fit can become problematic. Knowledge of the transduction factor Γ and approximate values for m and K may solve this problem, but we are more readily served if we rearrange a bit in the expression for the conductance:

$$G(\omega) = \frac{1/R}{1+\left(\frac{1}{\omega LC}-\omega\right)^2\left(\frac{L}{R}\right)^2}, \tag{2.117}$$

which, using Eqs. (2.111), (2.112), (2.114), can be written in terms of parameters that can easily be estimated from a plot of data points, in the approximation of angular frequency close to resonance ($\omega \approx \omega_s = (LC)^{-1/2}$):

$$G(\omega) = \frac{\chi}{1+4\left(\omega_s-\omega\right)^2\left(\frac{1}{\Delta\omega}\right)^2}. \tag{2.118}$$

Note the factor 4 in Eq. (2.118), resulting from quadratic expansion of the factor containing ω. Now, according to the relations Eq. (2.106), damping, mass and stiffness follow from the data and the coupling factor in the following manner:

$$D = \Gamma^2\frac{1}{\chi}, \qquad m = \Gamma^2\frac{1}{\chi\Delta\omega}, \qquad K = \Gamma^2\frac{\omega_s^2}{\chi\Delta\omega}. \tag{2.119}$$

In this thesis (Chapter 4), mainly the latter relation of these three will be found of interest.

2.4.3 Influence of bias voltage

The bias voltage plays a crucial part in the analysis of resonant behavior of MEMS resonators. In the first place, it determines how mass, stiffness and damping are to be translated in the equivalent lumped elements L, C and R through the transduction factor Γ, and is essential for the coupling between the electric and mechanical domain. Secondly, it is easy to understand that tuning the bias voltage will affect the parallel resonance frequency through

$$\omega_p = \omega_s\sqrt{1+\frac{C}{C_M(V)}} \tag{2.120}$$

in which the static MEMS capacitance C_M is obviously dependent on the bias voltage.

More concealed, but more pronounced in effect, is the fact that also the series resonance

$$\omega_s = \sqrt{\frac{K}{m}} \tag{2.121}$$

varies with bias voltage. The reason is that the stiffness K does not represent the intrinsic spring constant of the device, but rather the *electromechanical* stiffness as defined in (2.18) and (2.19), which is importantly affected by the bias voltage. For increased voltage this series resonance frequency follows a negative trend. At last, if we combine Eq. (2.115) with relations Eq. (2.106), we observe that the transduction factor drops out in the quality factor:

$$Q = \frac{1}{D}\sqrt{mK} \ , \tag{2.122}$$

which should obviously be the case as the quality factor is invariant in the electric and mechanical domains. However, it is precisely because of the characteristics of electromechanical stiffness that also the quality factor is dependent on bias voltage.

2.4.4 Damping

A detailed and thorough account of possible damping mechanisms is considered to be beyond this thesis' scope. In a subsection (4.3.5) in a section (4.3) on experiments on resonances a few words are devoted to this topic. From damping mechanisms such as viscous damping ($F_{damp} \propto v$),squeezed film damping, turbulent drag ($F_{damp} \propto v^2$) and energy losses via the springs, only the first two are thought of importance for the vacuum regime (10^{-4} – 60 mbar) in which the experiments have been carried out.

2.5 CONSIDERED SIDE-EFFECTS

This section deals with four effects that, on superficial reflection, could play a role in parallel plate capacitive MEMS. The first of these, non-parallelity of the electrodes, certainly needs consideration and is actually linked to an otherwise difficult-to-explain experimental case adopted in this thesis. The second, the Casimir-effect, is an exotic phenomenon in which considerable fundamental interest is taken. Programs have been set forward to employ capacitive MEMS in this field of research. For the structures encountered in this thesis however, the Casimir force is far too weak to play any role of importance. The third subsection touches coarsely the phenomenon of fringe fields present at the boundary of a parallel plate capacitor. For various aspect ratios the force is calculated with Finite Element Modeling and compared with standard theory. The last subsection very briefly considers the breakdown of air. Typical electric fields in capacitive MEMS in this thesis can be several MV/m, sufficiently high to generate air discharges. Nevertheless, because of the proximity of the electrodes they are not expected to occur even in experiments under ambient conditions.

2.5.1 Non-parallel electrodes

The theory encountered so far has mainly been concerning the ideal situation of a uniform capacitive gap, i.e. the spacing between the electrodes is constant everywhere within the capacitive area. For many practical cases this assumption provides a reasonable till good relation between gap and capacitance (in which one might speak of an 'effective' gap). An already encountered situation in which this assumption has released is the clamped-clamped beam electrode, where the actuating bottom electrode runs underneath a considerable portion of the span of the bridge electrode, so that the gap cannot be considered uniform. In this subsection the focus will be on plain electrodes (unlike the curved clamped-clamped beam), that are not exactly parallel. In subsection 4.3.3 a case is encountered in which the non-parallelity of the electrodes might provide an explanation for a disparity between capacitance and curvature of the C(V)-relation. While a parallel capacitance might clarify a curvature lagging behind on what could be expected on the basis of an observed capacitance, non-parallel electrodes could make the device more sensitive than what is to be expected from a certain capacitance value.

Let us consider a two-plate capacitor with capacitive area $A = wl$ ($l = b - a$) with minimal and maximal spacing z_a and $_b$ respectively, see Fig. 2.27:

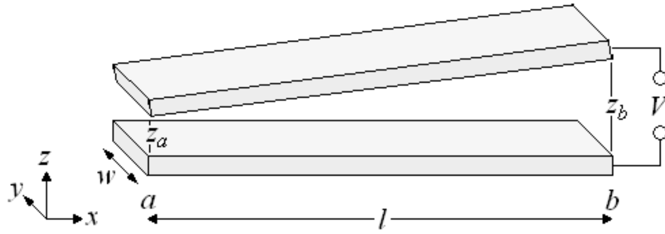


Fig. 2.27 Two-plate capacitor with non-parallel electrodes

In the following the assumption is made that the slope angle of the top electrode is so small (in practice not more than a few arc minutes) that the resulting decrease in capacitive area can be neglected. The capacitance of this structure is given by:

$$C = \int dC = \int_a^b \frac{\varepsilon_0 w}{z(x)} dx = \frac{\varepsilon_0 A}{z_{eq}}, z_{eq} \equiv \frac{z_b - z_a}{\ln(z_b / z_a)}, \quad (2.123)$$

where z_{eq} stands for "equivalent gap", the spacing between the electrodes of a parallel plate capacitor having the same capacitance. The limit $z_b \rightarrow z_a$ (parallel plates) is a useful check. One thing to note from (2.123) is that it is not so much the slope angle $\phi = \arctan[(z_b - z_a)/l]$ that determines the capacitance, but ratio and difference of the minimum and maximum separation. Consequently, the top electrodes in Fig. 2.28, both of area A, would give rise to identical capacitances.
It is seen further that if a rectangular top electrode is allowed to rotate about two axes intersecting in the centre (Fig. 2.29), the capacitance will be a minimum when top and bottom electrodes are parallel.

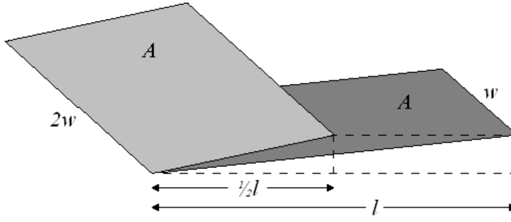


Fig. 2.28 Two sloped electrodes giving identical capacitances

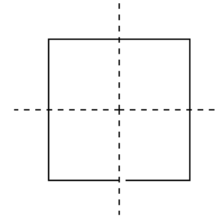


Fig. 2.29 rotation axes top electrode for parallellization

This is in fact a well-known procedure for parallelization of two plates and has been applied in for example [23] in which an experiment is described to measure the Casimir force using capacitive MEMS.
A potential difference between the plates pulls the top electrode down with a force

$$F = \frac{1}{2}V^2 \int \frac{dC}{z(x)} = -\frac{\varepsilon_0 A}{2z_a z_b}V^2 \ . \tag{2.124}$$

It is not difficult to show that $z_{eq}^2 \geq z_a z_b$ (equal if $z_a = z_b$). Consequently, the force in Eq. (2.124) is larger for a capacitor with a tilted plate than for a parallel plate capacitor. In order to proceed to $C(V)$ – characteristics, one needs to specify how the suspension (stiffness k) of the top electrode responds to the distribution of the force over the top electrode. Three cases will be compared. I) Parallel plate capacitor, II) Constant slope III) Free rotation.

I) Parallel plate capacitor.

This case has been discussed in section 2.1 of this chapter. For completeness it is recapitulated:

$$C_{par}(V) = \frac{\varepsilon_0^2 A^2}{2kz_0^4}V^2 + C_0 \, . \tag{2.125}$$

II) Constant slope (or completely suppressed rotation)

In this case it is assumed that the suspension is constructed such that the slope angle, or the difference $\Delta z = z_b - z_a$, remains constant during pulling down of the bridge. Assuming that the deflection of the leading edge of the top electrode is small compared to the gap, we find

$$C_{sl}(V) = \alpha_{sl}V^2 + \frac{\varepsilon_0 A}{\Delta z}\ln(1+\frac{\Delta z}{z_{a0}}), \qquad \alpha_{sl} \equiv \frac{\varepsilon_0^2 A^2}{2kz_{a0}^2(z_{a0}+\Delta z)^2} \geq \alpha_{eq} \, . \tag{2.126}$$

Here z_{a0} is the initial position of the lower edge of the plate. As expected from the force, the sloped curvature is larger than a parallel plate curvature α_{eq} featuring the equivalent gap.

III) Free rotation

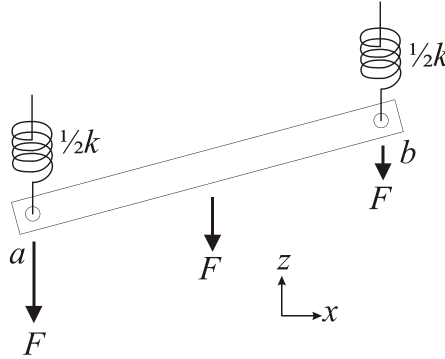


Fig. 2.30 Free rotatable, tilted top electrode with endpoint suspensions. Arrows indicate the force distribution on the plate.

The top electrode is assumed to be suspended on its end points on two identical springs, each with stiffness ½k (Fig. 2.30). Free rotation is possible at the points where the springs are attached to the plate. Of course many intermediate cases are conceivable, in which the pulled-down top electrode also has to work against torsion of the suspension. In fact, in II) the torsional resistance of the suspension is infinite.
Now, the springs each have to balance a moment. The moment generated by the electric force about the rotation axis through $x = a$ is given by

$$M_{el,a} = \int_{x=a}^{x=b} x \frac{dC}{z(x)} = \frac{\varepsilon_0 AlV^2}{2(z_b - z_a)} \left(\frac{1}{z_b} - \frac{\ln(z_b / z_a)}{z_b - z_a} \right), \tag{2.127}$$

for which the spring at $x = b$ has to compensate with a moment

$$M_{k,a} = -\tfrac{1}{2} kl(z_{b0} - z_b), \tag{2.128}$$

where the index of the moment refers to the rotation axis. The same relations apply *mutadis mutandum* to the rotation axis at b (in Eq. (2.127), replace the factor x in the integrand by $l - x$). The spring in a has to compensate for a larger electric moment and will consequently stretch further. Applying a voltage will consequently increase the slope of the top electrode. An analytical, explicit $C(V)$-relation for this third case is quite complicated in its appearance and not thought of sufficient interest. Instead, the three cases are compared numerically in Fig. 2.31.

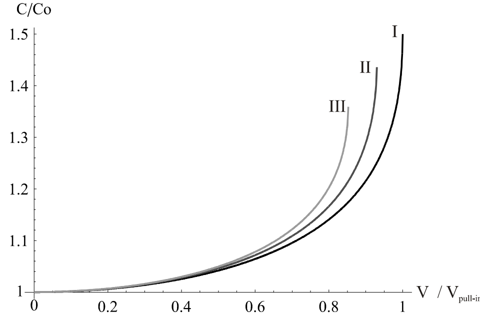


Fig. 2.31 Comparison of C(V) curves of I) parallel plate capacitor, II) constant sloped electrode III) free rotating sloped electrode. The parameters are chosen such that all initial capacitances are normalized on 1 and voltages to the pull-in voltage of the parallel plate capacitor. This means $\varepsilon_0 A = 1$, $z_0 = 1$ and $k = 27/8$. In cases II and III, the initial slope has been chosen such that $z_b = 2z_a$, which dictates $z_a = \ln 2$ to yield an initial equivalent gap of $z_{eq} = 1$. Curvatures are increasing for ascending case numbers. Furthermore, we observe further that the pull-in voltage decreases for I, II and III respectively and likewise the critical capacitance at which pull-in is about to occur.

Curvatures are ascending for the three respective cases. This demonstrates that for non-parallel plates, the structure can be more sensitive than what would be expected from the initial capacitance. For two ratios of z_b and z_a the curvatures are compared in Table 2.1 normalized on the parallel plate curvature ($\alpha_I = 0.149$).

	$z_b / z_a = 2$	$z_b / z_a = 3$
α_{II} / α_I	1.081	1.221
α_{III} / α_I	1.141	1.376

Table 2.1 curvature comparison, normalized on parallel plate curvature. For the two columns $z_a = \ln 2 \sim 0.69$ and $z_a = ½ \ln 3 \sim 0.55$ respectively.

The gap ratios given here are large but not implausible for what can be expected for the structures encountered in this thesis. Wafer thickness and geometric curvature can vary in the order of a few µm over the sample distance (a frame with lateral dimensions of a few cm), so that tangents for electrodes measuring a few mm become in the order of 0.001. Furthermore, despite efforts to minimize dust in the sample, only one particle is needed to tilt the frame a bit after which the electrodes become disparallel.

2.5.2 Casimir effect

In 1948 it was recognized by Casimir that even in absence of obvious electro-magnetic fields, non-gravitational forces could exist. He predicted that two grounded conducting plates, held parallel to each other at a very small distance in vacuum, would nevertheless attract each other. This is as a result of the non-emptiness of the vacuum, which is an important prediction of all relevant quantum field theories. Between the plates, only a small class of bounded vacuum states is allowed, as the wave functions have to vanish at te surface of the plate. Outside there is no restriction whatsoever and the density of states

there is much higher. This results in a pressure pushing the plates inward. The force as derived by Casimir is given by:

$$F_{Cas} = -K_{Cas}\frac{A}{z^4}, \qquad (2.129)$$

where the minus sign indicates the force to be attractive. A is the area of the plates and z their separation. The coefficient K_{Cas} is a numerical constant:

$$K_{Cas} \equiv \frac{\pi hc}{480} = 1.30 \cdot 10^{-27}\,\mathrm{Nm^2},$$

where h and c are Planck's constant and the vacuum speed of light respectively. One of the most notable aspects of the Casimir force is the very rapid drop off of the magnitude with distance. The effect is very small and only noticeable when the plates are very close to each other, in the range of 1 µm, but preferably much lower. This places high demands on the experimentalist, especially to ensure a very well kept parallelity between the plates. Because of this point experiments with one of the plates replaced by a sphere have been carried out. However, in January 2002 the first successful measurements on the particular parallel plate configuration haven been published [23] where use was made of a vibrating cantilever. A feedback circuit ensured optimal parallelity by tilting the relative position of the plates such that capacitance was extremized for a given separation distance (in the range of 0.5 – 3 µm). For the Casimir coefficient a value of $K_{\mathrm{Cas}} = (1.22 \pm 0.18) \cdot 10^{-27}$ Nm² was found, in good agreement with the theoretical prediction. A remarkable example of a quantum phenomenon manifesting itself on the 'macroscopic' level.

Capacitive MEMS encountered in this thesis (especially those in Chapter 3), have sizes not very different from the above-mentioned device, which could also be counted an example of a capacitive MEMS. A sample calculation for a typical device of A = 1 mm² and z = 1 µm would experience a Casimir force of 1.3 nN. This is a factor of 3400 smaller if a voltage of 1 Volt would be applied to the same device and even for typically encountered built-in voltages of 0.1 V the Casimir contribution would be unnoticeable. On the other hand, assume thin dielectrics of d = 2 nm and ε = 9.1 (Al_2O_3) and both electrodes grounded (V = 0). It would require a built-in voltage of only 0.017 V and a corresponding surface unit charge density of $n_e = V_{\mathrm{bi}}\varepsilon\varepsilon_0/2ed \sim 2100\ \mu\mathrm{m}^{-2}$ (= 0.34 mC/m²) to create a force of magnitude similar to the Casimir force, which is a relatively low, but not uncommon charge density in the practice of capacitive MEMS, which underscores the necessity of accurately deal with trapped charges (prevention, removal or force compensation) if Casimir force measurements are being planned.

If the MEMS device in the calculated example would have a spring constant of 100 N/m, the Casimir force would displace it over 13 pm, which results in an increase in capacitance (~ 0.1 fF) that is within the reach of accurate contemporary measurement systems. However, in the $C(V)$ measurements performed here, the Casimir force is overshadowed by much more important contributions to the total force.

2.5.3 Fringe forces, charge distribution

The electrostatic force on the top electrode of a parallel plate capacitor is usually calculated with Eq.(2.7), in which, depending on the context, V may be replaced by $z\sigma/\varepsilon_0$ (z = gap, σ = surface charge), or $zQ/A\varepsilon_0$ with Q the total charge and A the capacitive surface. The derivation of this expression assumes a homogeneous electric field between the electrodes. At the edges however, the field lines are found to bulge outward (Fig. 2.32)

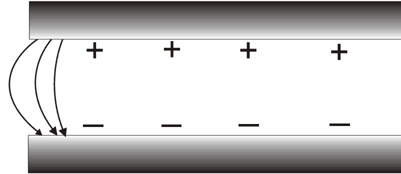


Fig. 2.32 Boundary field lines in a charged parallel plate capacitor

The force between capacitive plates has been calculated by 3D Finite Element Modeling in COMSOL. The geometry consisted of a square plate capacitor of side L and thickness h = 1 µm, at a voltage ½V floating a distance ½z above an infinite grounded plane. The capacitance and force is then equivalent to a parallel plate capacitor at voltage V and gap z, see Fig. 2.33:

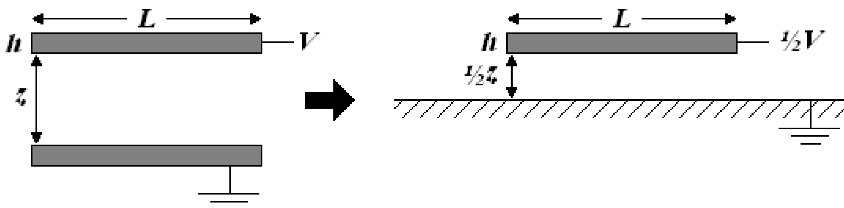


Fig. 3.33 Modeled geometry

As is known since the times of Maxwell, The equipotential surfaces are curling upwards close to the edges of the parallel plates, see Fig. 2.34

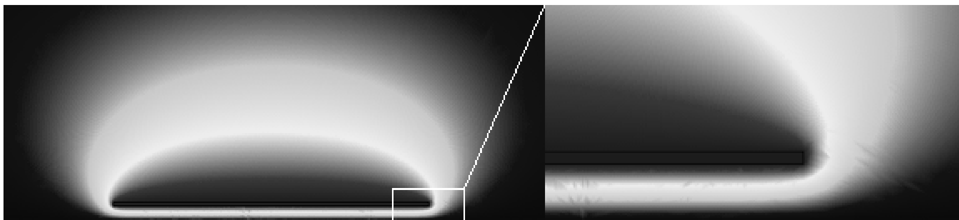


Fig. 2.34 Potential contourplot of a finite capacitive electrode above an infinite grounded plan, for L = 10 µm and z = 1 µm. Light regions correspond here to intermediate voltages. The dark southern region is close to 0 V, while the electrode is at V = 1 V (in the actual simulation: ½V = 0.5 Volt.).

Towards the perimeter of a parallel plate capacitor, the electric field midway between the electrodes becomes weaker, while at the electrode itself the field is very strong, leading to concentrations of surface charge there. Neglect of the fringe effect underestimates the

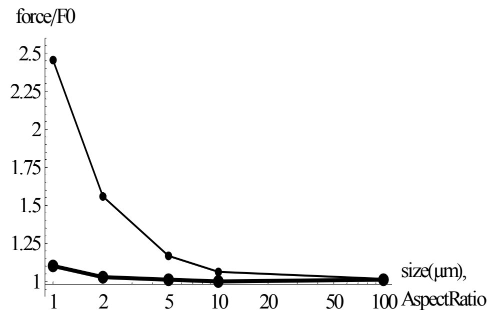


Fig. 2.35 Double logarithmic plot of Finite Element Modeling of the force between parallel plate electrodes. The vertical axis denotes the simulated force divided by the theoretical force, given by Eq. (2.7).The small dots represent the force on the underside only of the plate electrode, while the thick dots are the total force, which is generally lower: for an electrode of finite size, there is a small electric field present above it, tending to pull it upward.

actual capacitance. For various aspect ratios L/z (L varied, z kept constant), the FEM force is compared to the force given by the standard formula Eq. (2.7), see Fig. 2.35.
It is seen that noticeable deviations from the standard parallel plate formulary occur for aspect ratios below 10 or 5, a bit dependent on what precision is demanded[18]. Aspect ratios for MEMS capacitors encountered in this thesis are typically of the order of 100, in which case the thus introduced error is within the order of one percent.

2.5.4 Breakdown

As encountered in measured $C(V)$-curves in Chapters 3 and 4, typical voltages occurring in MEMS-devices can be of the order of at least a few Volt. Given that gap distances are in the order of a micrometer, this raises the question of whether an avalanche discharge may occur if the MEMS is operated in vacuum or in air, given that the breakdown field of air at atmospheric pressure is about $3\cdot 10^6$ V/m. In 1889, Paschen [24] found however that the breakdown voltage between two plate electrodes depends not only on the gas pressure, but also on the electrode separation (at close distances). For the critical voltage V_c at which breakdown occurs he established

$$V_c = \frac{a(Pz)}{\ln(Pz) + b}, \tag{2.130}$$

[18] Incidentally, the force between two point charges is a pure fringe effect.

where P is the pressure and z the electrode separation. a and b are constants that depend on the composition of the gas. For P in atmosphere and z in meters, a = 43.6 MV and b = 12.8. If the plates are brought closer, the breakdown voltage decreases, reaches a minimum and then rises steeply, as with decreasing distances it is harder to ionize enough gas atoms to create an avalanche effect. At low pressures or very close spacings, the gap becomes at the order of the mean free path of electrons. Paschen curves for various gases are plotted in Fig. 2.36.
According to Paschen's law Eq. (2.130), there exists for every pressure a *minimal* breakdown voltage (*Paschen minimum*): $\frac{dV_c}{d(Pz)} = 0$ gives $(Pz)_{\min} = e^{1-b}$, so that

$$V_{c,\min} = ae^{1-b}. \qquad (2.131)$$

In air, one cannot have a discharge below 327 V, *irrespective of pressure and electrode separation*. At atmospheric pressure, this minimum occurs at z = 7.5 µm. For MEMS devices, it is in principle inconvenient that the Paschen minimum occurs at this particular (close) distance, but the involved breakdown voltage is still comfortably large.
The case is however more complicated than this. Paschen's law predicts that for gaps decreasing to 2.76 µm, at atmospheric pressure the breakdown voltage tends to infinity. It has been experimentally established, that breakdown still occurs at this close distance. Experiments by Dhariwal et al. [25] indicate that for gaps below ~ 10 µm, differences with Paschen's law start to occur and that below 4 µm Eq. (2.128) ceases to satisfactorily predict the breakdown voltage. Their results for nickel electrodes at P = 1 atm. are presented in Fig. 2.37

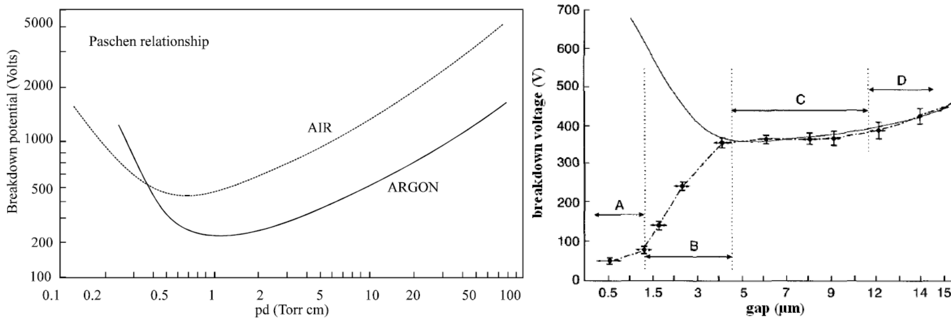


Fig. 2.36 (left) Paschen curves for air and argon
Fig. 2.37 (right) Breakdown voltage at P = 1 atm.versus gap distance for clean nickel electrodes. The solid concave curve represents data gathered by CIGRE (International Conference on Large High Voltage Electric Systems) in the 1970's [6], extrapolated to small gap distances. Reproduced from [25].

Paschen's law is still valid in region *D*. Below z ~ 10 µm there is a plateau *C* just above the theoretical Paschen minimum. For lower gaps (*A* and *B*), breakdown behavior is observed opposite to Paschen behavior. Breakdown at 12 V (air) has been observed for z = 0.25 µm.

The deviations are amongst others attributed to field emission (see also subsection 2.3.3), and therefore the electrode material starts playing a role (through the work function). Surface roughness starts playing a role as well, as small protrusions cause high local electric fields, facilitating emission. The exact reasons and underlying mechanisms for the observed behavior are still not very well understood.
Consequently, MEMS devices should not experience problems at ambient conditions and moderate voltages, but breakdown is certainly an effect that has to be taken into account. The flexibility of moving elements in actuators should be such that low voltages are sufficient to displace them. If a microswitch is designed too stiff or with a too small actuation surface, breakdown voltage might be below the pull-in voltage. This does not require extremely pathetic design parameters, but for typical values there is no danger of $V_{\text{pull-in}}$ approaching V_c. For example: for $z = 1$ μm, $k = 1000$ N/m and $A = 10^{-7}$ m^2 , (a relative stiff device) $V_{\text{pull-in}} = 2.3$ V, while for this gap breakdown occurs at around 50 V (material dependent). However, in designing and operating MEMS devices, it is still important to ascertain breakdown will not affect proper functioning.

REFERENCES

[1] T. Rijks et al., "MEMS tunable capacitors and switches for RF applications", *Proceedings of 24th Intern. Conf. on Microelectronics (MIEL)*. VOL **1** (2004)

[2] A. Oja, J. Kyynäräinen, H. Seppä, and T. Lampola, "A micromechanical dc-voltage reference," *Proc. Conf. Dig. CPEM*, 2000, pp. 701–702.

[3] E. F. Dierikx, "A MEMS-Stabilized AC Voltage Reference Source", *IEEE Transactions on Instrumentation and Measurement,* **56** 313 (2007)

[4] A. Kärkkäinen et al., "MEMS-based ac voltage reference", *IEEE Transactions on Instrumentation and Measurement*, **54** 595 (2005)

[5] L. M. Castaner, and S. D. Senturia, "Speed-Energy Optimization of Electrostatic Actuators Based on Pull-In", IEEE Journal of Microelectromechanical Systems **8** 290 (1999)

[6] J. I. Seeger., and B. E. Boser, "Charge Control of Parallel-Plate, Electrostatic Actuators and the Tip-In Instability", *Journal of Microelectromechanical Systems*, **12** 656 (2003)

[7] S. P. Timoshenko, and S. Woinowsky-Krieger, "Theory of plates and shells", McGraw-Hill, New York (1959)

[8] M. Elwenspoek, and R. Wiegerink, "Mechanical Microsensors", Springer, Berlin (2001)

[9] S. Pamidighantam, R. Puers, K.Baert, and H. A. C. Tilmans, "Pull-in voltage analysis of electrostatically actuated beam structures with fixed-fixec and fixed-free end conditions", *Journal of Micromechanics and Microengineering* **12** 458 (2002)

[10] R. J. Roark, and W. C. Young, "Formulas for stress and strain", McGraw-Hill, New York (1989)

[11] J. Wibbeler, G. Pfeifer, and M. Hietschold, "Parasitic charging of dielectric surfaces in capacitive microelectromechanical systems (MEMS)", *Sensors and Actuators A* **71**, 74 (1998)

[12] R. W. Herfst, P.G. Steeneken, and J. Schmitz, "Indentifying degradation mechanisms in RF MEMS capacitive switches", *Proceedings MEMS* (2008)

[13] X. Rottenberg, B. Nauwelaers, W. de Raedt, and H. A. C. Tilmans, "Distributed dielectric charging and its impact on RF MEMS devices", *Proceedings 34th Eurpoean Microwave Conference* - Amsterdam (2004)
[14] Sze, S. M., "Semiconductor devices: Physics and Technology", John Wiley & Sons, New York (1985)
[15] Sze, S. M. "Semiconductor devices: Physics and Technology 2nd edition", John Wiley & Sons, New York (2002)
[16] J. G. Simmons, "Poole-Frenkel effect and Schottky effect in metal-insulator-metal systems", *Physical Review* **155**, 657 (1967)
[17] J. Frenkel, "On pre-breakdown phenomena in insulators and electronic semi-conductors", *Physics Review Letters* **54** 647 (1938)
[18] J. Mizsei, "Fermi-level pinning and passivation on the oxide-covered and bare silicon surfaces and interfaces", *Vacuum* **67** 59 (2002)
[19] Y. Khlifi, K. Kassmi, L. Roubi, and R. Maimouni, "Modeling of Fowler-Nordheim current of metal/ ultra thin oxide/ semiconductor structures", *M.J. Condensed Matter* **3**, 53 (2000)
[20] M. Lenzlinger, and E. H. Snow, "Fowler-Nordheim tunneling into thermally grown SiO_2", *Journal of Applied Physics* **40**, 278 (1969)
[21] N. F. Mott, "III. Localized states in a pseudogap and near extremities of conduction and valence bands" *Philosophical Magazine* **19** 835 (1969)
[22] H. A. C. Tilmans, "Micro-mechanical sensors using encapsulated built-in resonant strain gauges", Ph.D. thesis, University of Twente (1993)
[23] G. Bressi, G. Carugno, R. Onofrio, and G. Ruoso, "Measurement of the Casimir Force between Parallel Metallic Surfaces", *Physical Review Letters* **88** 041804-1 (2002)
[24] Paschen, "Ueber die zum Funkenübergang in Luft, Wasserstoff und Kohlensäure bei verschiedenen Drucken erforderliche Potentialdifferenz", *Annalen der Physik* **273** 69 (1889)
[25] R. S. Dhariwal, J.-M. Torres, and M.P.Y. Desmulliez "Electric field breakdown at micrometre separations in air and nitrogen at atmospheric pressure", *IEE Proc.-Sci. Meas. Technol.,* **147** 261 (*2000*)

CHAPTER 3

MEASUREMENTS ON RF POWER SENSORS

The EMMA (ElectroMechanical Microcomponents for precision Applications) – project (IST-2000-28261) [1] was a three-year European project running from 2001-09-01 to 2004-08-31. It was put forward for the development, improvement, characterization and packaging of a variety of stable capacitive MEMS-based components, among which power sensors, accelerometers and voltage standards [2]. Some publications in the spirit of this project are [3, 4]. The participating consortium consisted of seven partners, of whom the University of Twente cooperated with VTT (Finland) for the development of a sensitive HF power sensor.

This activity was guided by challenging demands concerning sensitivity and stability of the High Frequency (HF) power sensor. Regarding the former, commendable progress was achieved. Regarding the latter, it had been recognized that charge trapping could pose a limitation. Presented in this chapter are measurements on charge trapping in the full-grown, real(istic) devices developed in the context of the above-mentioned EMMA-project. In Chapter 6, it is calculated when trapped charges become a limiting factor.

Capacitive MEMS structures sometimes have to operate at low temperatures. This could be imposed by the ambient conditions, most notably space missions. In other cases, they are part of cryogenic electronics, either because of superconducting elements or just to reduce thermal noise. This is the motivation to study the effects of charge trapping at various temperatures in the cryogenic regime (down to 4 K). The large variations in temperature may well have their repercussions on the thermo-mechanical behavior of the device under test. It is necessary to pay attention to this influence.

The structures covered in this chapter are to be divided into two classes, most readily by the criterion of the type of bridge (top electrode) used; silicon nitride or aluminum. More details of these samples are to be found in the corresponding sections 3.2 and 3.3. For further research on charge trapping, performed on newly designed samples, be the reader invited to Chapter 4.

3.1 MEASUREMENT PRINCIPLES AND SETUP

3.1.1 Impedance measurement

The probing of the samples in this chapter was invariably according to a four-contact set-up; High and Low Voltage and High and Low Current using four coaxial cables, see Fig. 3.1.

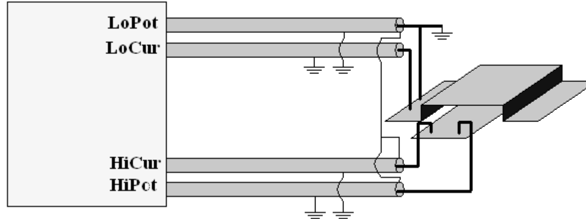


Fig. 3.1 Measurement connection scheme

An HP4194A applied a bias voltage between the voltage contacts, with an 800 kHz 0.3V amplitude AC-voltage on top. The impedance is determined by reading the current, and internally translated to capacitance. These latter values have been recorded in the data files. The grounds are connected as close as possible to the sample, in order to compensate for any parasitic capacitance of the cables. Obviously, the cables must be identical.

A typical example of a measured C(V)-curve is provided in Fig. 3.2. It is shown what information can typically be retrieved from such a curve and what orders of magnitude relevant quantities have. The measured device is an aluminum bridge as described in subsection 3.3.1., in particular Fig. 3.17. The capacitive area is 3600 µm × 100 µm. For this type of samples, parasitic capacitance is neglected.

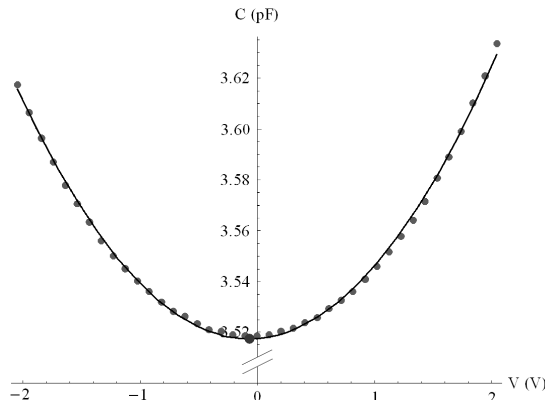


Fig. 3.2 Typical C(V) – curve, measured on an aluminum bridge RF power sensor, at 300K.

The dots represent actual data points (V_i, C_i), while the continuous line is a quadratic fit $C(V) = \alpha V^2 + \beta V + \gamma$. Obviously, we see that for $V = 0$ we must have $\gamma = C_0$. The big dot marks the minimum of the fitted curve. Assuming no significant parasitic capacitance, the

sample has a rest gap of $z_0 = \varepsilon_0 A/C_0 = 0.91$ µm. This is very close to the designed value of 1 µm, see subsection 3.3.1. At $V = 2$ V, the electric force F is 7.8 µN downward (Chapter 2, Eq. 2.7). On stiffer devices than this one, larger electric forces of up to several mN can be applied before the bridge electrode is pulled in. Assuming further that the plates remain parallel during actuation, we can write $dC/dz = -\varepsilon_0 A/z^2$ so that the displacement is $\Delta z \approx z_0^2 \Delta C/\varepsilon_0 A = 29$ nm downward, which is about 3% of the gap distance. The parabolic coefficient $\alpha = C_0^2/2kz_0^2$ is fit to 0.026 pF/V^2, and with a zero – capacitance $C_0 = 3.52$ pF and a gap $z_0 = 0.91$ µm, the spring constant is estimated to be 0.3 kN/m. From C_0 and α the pull-in voltage $V_{pull-in} = \sqrt{4C_0/27\alpha}$ (Eq. (2.15)) is estimated to be 4.5 V. This is just more than twice the maximum applied voltage, so it is to be expected that the $C(V)$-data are still in the "parabolic regime", i.e. they can be fit adequately by a quadratic function, as has been discussed in subsection 2.1.2. This is confirmed by inspection of Fig. 3.2.

The minimum of the fitted quadratic curve V_{offset} is retrieved by

$$V_{offset} = \frac{-\beta}{2\alpha} = \left. \frac{-\frac{d}{dV}C(V)}{\frac{d^2}{dV^2}C(V)} \right|_{V=0} . \qquad (3.1)$$

It has here the relatively small value of -0.065 V. A check will be provided whether the fit quality is such that this small offset is significant. For the second degree polynomial fit used here, let $\overleftrightarrow{V}$ be a matrix consisting of row vectors (V_i^2, V_i, 1) where V_i is the voltage of the ith data point. Further, let S be the sum of the squares of the *residuals*:
$S = \Sigma_i (C(V_i) - C_i)^2$. Then the standard errors σ in the coefficients α, β, γ are given by

$$\sigma_{\alpha,\beta,\gamma} = \frac{S}{n-3} (\overleftrightarrow{V}^T \overleftrightarrow{V})^{-1}{}_{\alpha\alpha,\beta\beta,\gamma\gamma} . \qquad (3.2)$$

Here n is the number of data points and must obviously be larger than the number of polynomial coefficients, 3.
The double indices indicate the diagonal elements of the 3 × 3 inverse matrix of the product of $\overleftrightarrow{V}^T$ (T is "transposed") and $\overleftrightarrow{V}$. The relative standard error in V_{offset} is then given by

$$\frac{\sigma_{V_{bi}}}{V_{offset}} = \sqrt{\left(\frac{\sigma_\alpha}{\alpha}\right)^2 + \left(\frac{\sigma_\beta}{\beta}\right)^2} . \qquad (3.3)$$

It is found that $\sigma_{\text{Voffset}} = 0.05\ V_{\text{offset}}$. The overall standard deviation of the fit comes out at $\sigma_C = 0.0013$ pF, about 1% of the total variation of the capacitance between 0 and 2 Volt. This confirms that the data points are well fittable and that the nonzero value of V_{offset} can be considered significant. It is known that electrodes of different materials can give rise such an offset, as mentioned in subsection 2.3.1. The electrodes of the device from which the curve in Fig. 3.2 was measured are made of identical material (aluminum). From a material point of view, an offset would not be expected.

3.1.2 insert, flow cryostat

The majority of the temperature-varied measurements have been carried out using a flow cryostat. The sample was mounted on an insert finger and connected to coaxial cables with a characteristic impedance Z_0 (for a lossless line $Z = \sqrt{(L/C)}$) of 50 Ω, the grounds being interconnected only here. The sample was surrounded by a mu-metal case and a brass case, shielding the sample from external magnetic and electric dc and low frequency fields respectively. On the opposite side of the sample, one leg of a chromel-alumel thermocouple was attached for temperature reading and control, governed by a Lakeshore 320 Autotuning temperature controller. This required a heating wire, which was wound around the outer (brass) case. A flow cryostat cools by pumping liquid helium through a tube spiraling around the chamber containing the insert. The helium is in thermal contact with the chamber wall, which in turn absorbs heat from the gas surrounding the sample case. Of course, all is isolated from the environment by internal radiation shields and a vacuum wall. Compared to a bath cryostat, a flow cryostat has the advantage of low helium consumption. Because the helium is pumped, it is possible to reach temperatures even a few degrees below helium liquefaction temperature (4.2 K). Because of degraded accuracy of the thermocouple in this low regime, the recorded temperatures are less reliable. Because the helium flow can be regulated, it is possible to conduct measurements at stabilized intermediate temperatures, e.g. 100 K, though at relatively ‘high’ temperatures the temperature control circuit allowed temperature fluctuations of about 1 K (at temperatures ≤ 30 K fluctuations are less than 0.1 K).

3.2 SILICON NITRIDE BRIDGES

3.2.1 Devices

The first class consists of capacitive MEMS of which the bridge was made of silicon nitride, bearing electrodes that formed a capacitive coupling to the underlying (conductive) substrate. These samples have been developed [5] in the context of the in EMMA project mentioned in the introductory text of this chapter. A cross section of the design (Fig. 3.3a) and a picture of the samples used are shown (Fig. 3.3b).

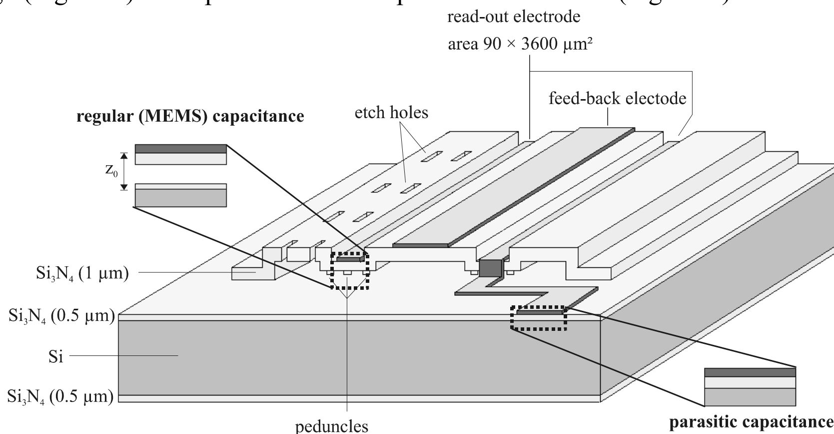


Fig. 3.3a: Schematic cross section Si_3N_4-bridge

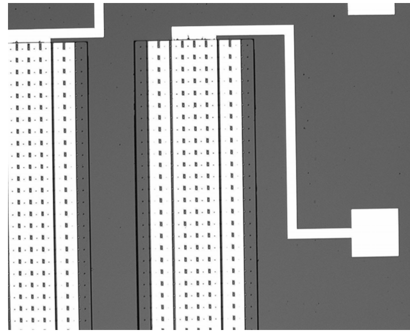


Fig. 3.3b: Top view Si_3N_4-bridge (optical microscope)

The device consists of three top electrodes, which are carried by a 1 µm thick silicon nitride (Si_3N_4) bridge[19]. The two flank electrodes act as read-out electrodes, while the central electrode is for feed-back purposes. The electrodes are capacitively coupled to the substrate. This substrate is a p-type, highly doped silicon wafer. The substrate is covered on both sides with 0.5 µm silicon nitride as isolation. On the left of Fig. 3.3 a magnified cross-section of the MEMS capacitor is shown. The top electrode is sputtered platinum (150 nm) on 10 nm sputtered chromium, which acts as an adhesion layer. The capacitive gap z_0 is partly filled with two layers of Si_3N_4, which has a relative dielectric constant ε of 7.5. The lower Si_3N_4 layer (0.5 µm) covers the conducting silicon substrate as an isolation layer. The upper layer (1 µm) forms the bridge that supports the top electrodes. The air gap was designed as 1.2 µm. Summing these three numbers gives a design value for z_0 of 2.7 µm.

Also drawn are peduncles, pending under the silicon nitride bridge. They very much reduce the chance of sticking of the bridge onto the substrate upon pull-in (or other circumstances that cause the bridge to collapse). This is because of the large reduction of the potential sticking area. Either they much reduce the distance over which the bridge collapses when pull-in is effectuated, or they prevent a proper pull-in induced movement altogether, that is, it is mechanically prevented that the bridge reaches its unstable point[20], irrespective of the magnitude of the applied voltage.

In a part of Fig. 3.3 and in Fig. 3.4 we see that the bridge is perforated with small etching holes. They facilitate wet-etching of the sacrificial polysilicon layer that sustains the bridge during the fabrication process. They reduce the effective capacitive area by about 5%. The measurements presented in this section have been done by using the two read-out electrodes, which have a combined capacitive area A of 2 × 90 µm × 3600 µm.

It should be noted that the bridge approximates in fact a *four*-sided clamped plate, rather than a clamped-clamped beam. Two opposite sides are clamped by the top silicon nitride layer, and two sides are partially clamped by the electrodes on top of the bridge. This is seen in the schematic in Fig. 3.3 from the right read-out electrode. The right side picture of Fig. 3.3 shows how the central electrode is connected to the contact pad.

The particular configuration of the electrodes entails a few notable points of attention. First and foremost, the equations of a pure parallel plate capacitor have to be extended so

---

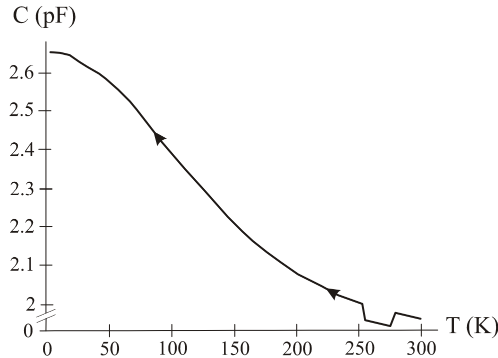
[19] Instead of the stoichiometric formula Si_3N_4, a different notation is sometimes used for silicon nitride: SiRN. Here "R" is written to emphasize the low-stress fabrication process of "Reaction bonding".

[20] For a pure parallel plate capacitor, with or without dielectric layers, the capacitance will never exceed 1.5 times its unactuated value before pull-in, see subsection 2.1.4.

as to encompass the presence of dielectrics by including a voltage-independent series capacitor. The capacitance C_d over the dielectric layer is 27.45 pF. This will have an immediate consequence for at which voltage pull-in is expected to occur, see subsection 2.1.4 for a more detail. Another important point to note is that the bonding pads and the interconnecting on-chip wiring make for a parasitic capacitance, as can be seen from the right inset in the schematic of Fig. 3.3. The metallic electrode is capacitively coupled to the substrate, separated only by the Si_3N_4 isolation layer. This constant parasitic capacitance is parallel to the variable MEMS capacitance. Its value is calculated to be 4.91 pF.

3.2.2 Thermal dependence

Before addressing the temperature dependence of charging-related effects it is necessary to survey the thermo-mechanical characteristics of the involved samples. During a cooling-down session in the flow-cryostat the capacitance (at constant zero bias voltage) has been followed. The sample is described in the previous sub-section. The advantage of measuring single capacitances over measuring a series of $C(V)$-curves is that the effect of a rapidly changing temperature is followed as closely as possible. On the other hand, this yields no information about what happens with other characteristics such as built-in voltage and curvature. This cooling down series, which took 45 minutes to reach 2K, is displayed in Fig. 3.4.



Fig, 3.4 Capacitance vs. temperature, arrows indicate direction of time. A calculated parasitic capacitance of 4.9 pF has been subtracted from the shown data.

After cooling down to 2K, the capacitance has increased by about 35% compared to the value at room temperature. The capacitance values as shown in Fig. 3.4 are assumed fully attributable to the MEMS structure itself; in other words: possible capacitive contributions from wiring, electronics, surroundings and readout signal (Eq. 2.79) are considered negligible. From a capacitive area of 2 × 90 µm × 3600 µm a capacitive gap z_0 of 4.4 µm (including the dielectric Si_3N_4 layers) at room temperature is derived, rather than the 2.7 µm as designed (see subsection 3.2.1). The observed amount of change in capacitance through temperature could be explained by a gap decreasing to 3.7 µm. The most natural explanation for the variation in capacitive gap is deformation of the bridge, imposed by differences in the expansion coefficients of the different materials used. It is not clear if the measured trend is conform expectation. The three materials that make up the device, silicon, silicon nitride and platinum, contract in unequal amounts when the

temperature is lowered. What can be said is that the metallic electrodes will try to enforce an upward curvature of parts of the silicon nitride bridge. However, the overall cross-sectional shape of the silicon nitride bridge is quite complicated. This makes predicting the exact deformation of the bridge difficult. Connected with this is the question whether the (effective) capacitive gap will increase or decrease.
For the temperature range 20-150 K available $C(V)$ - curves provide more insight in the temperature dependent behavior of various mechanical characteristics (Figs. 3.5 and 3.6).

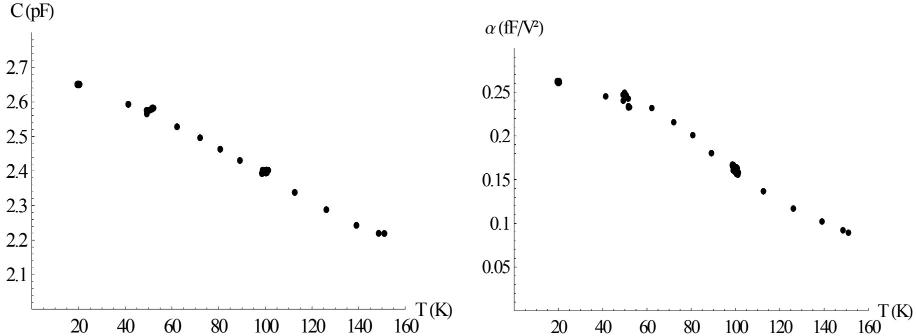


Fig. 3.5 (left) Capacitance variation with temperature
Fig. 3.6 (right) Curvature variation with temperature

The measurement series consists of phases of stabilized temperatures at 20 K, 50 K and 100 K, with intermediate curves taken at in-between cooling down stages. For the given temperatures, the capacitance values are in concordance with Fig. 3.4. As expected for a lowered bridge, the device becomes more sensitive, which is reflected in the increased parabolic curvature shown in Fig. 3.6. From this curvature, the stiffness k can be derived, as in Eq. (2.11). These stiffnesses are plotted against temperature in Fig. 3.7.

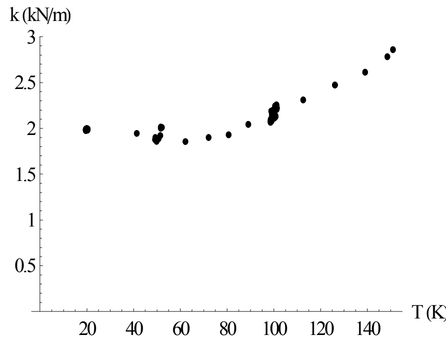


Fig.3.7 Estimated stiffness with temperature

The spring constant does not change too much, although the slight trend that is visible would be expected reversed. At lowered temperatures the silicon nitride, with a thermal expansion coefficient of 3 $\mu m\ m^{-1}\ K^{-1}$, is inclined to contract slightly more than the silicon substrate 2.6 $\mu m\ m^{-1}\ K^{-1}$. As the silicon nitride bridge is frustrated in its attempt to contract by the substrate, this would cause a tensile stress in the bridge which would increase the spring constant. As mentioned earlier, the platinum (9 $\mu m\ m^{-1}\ K^{-1}$) electrodes want to contract even more, also contributing to tensile stress.

3.2.3 Built-in voltage in silicon nitride bridges

For two stabilized temperatures (20 K and 294 K) the built-in voltage fitted from series of $C(V)$-curves are compared (Figs. 3.7, 3.8). $C(V)$ curves are voltage-swept in alternating directions, corresponding with thick and thin dots representing the positions of the minimal capacitances in successive curves.

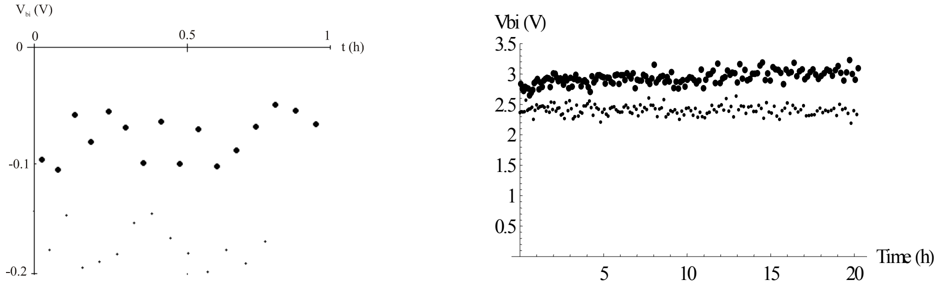


Fig.3.8 (left) Built-in voltage vs. time (T = 20 K)
Fig. 3.9 (right) Built-in voltage vs. time (T = 294 K) Thin and thick dots correspond to alternating directions of the voltage sweep.

In both cases, the most notable feature is the splitting of the two sweep-directions into two branches. At cryogenic temperatures this sample showed on average a difference of 0.1 V in the built-in voltage of upramping and downramping voltage swept curves. For a long series at room temperature, the average splitting turned out to be five times as high (0.54 V). The thick dielectric layers in this sample prove ample trapping states, which facilitates the occurrance of relatively high values for the built-in voltage (here 2.5 V at room temperature; the actual, or average value has not been found to display a correlation with temperature). For aluminum electrodes with their thin native oxide (section 3.3 and chapter 4), the built-in splitting has generally been found to disappear at temperatures below about 150 K, this in contrast with the case at hand where the phenomenon is still observed at $T = 20$ K. Observation of 3.9 also suggests a long term effect in the splitting, which widens at a rate of about 11 mV/hr.

3.2.4 C(V)-asymmetry

Initial cryogenic $C(V)$ - measurements were not performed in a flow cryostat, but by using a cylindrical aluminum package, half-submersed in a bath of liquid nitrogen. This easy but rather primitive solution for cryogenic measurements has the advantage of quick sample mounting and absence of pumping installations, which is a potential source of noise. The disadvantages are that between room temperature and the boiling temperature of liquid nitrogen no intermediate temperature can be maintained and that the wiring and the aluminum case may introduce uncertain sources of off-chip parasitic capacitance. Shown below (Fig. 3.10) are two consecutive curves from one of the ‘nitrogen-series’, a backward-swept voltage followed by a forward-swept voltage.

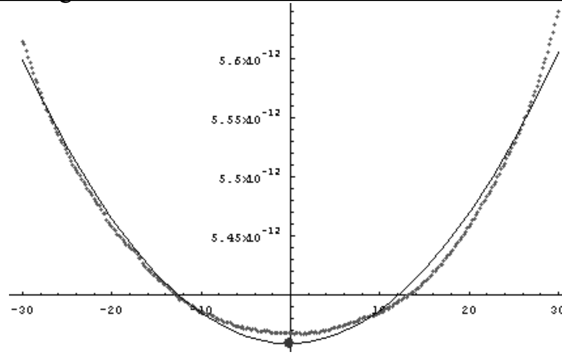


Fig. 3.10 Asymmetric C(V)-curve

The strings of small blue dots indicate measured capacitances for voltages varying between –40 and +40 Volt in this or in reverse order, in steps of 0.2 V. Both datasets are fitted by a quadratic function (thin black line). The big dot at the minimum in this fit marks the offset, or built-in voltage and the unactuated total capacitance, this is the device's zero capacitance plus parasitic capacitances. Small deviations in the parabolic fit are visible in the minimum and at the outer edges, as is to be expected. If improved fitting is wanted, it is possible to select from the $C(V)$ - dataset only those points closest to the minimum, because the quadratic fitting improves a lot far from pull-in. In some cases it it is possible to end up with too few points, however. A second possibility is the inclusion of the quadratic Taylor term Eq. (2.13). A third possibility involves the assumption that the data are symmetric around the offset voltage. The usual parabolic fit should find this offset, even if curvature and minimum capacitance are less accurately determined. Starting from

$$\frac{1}{C^2}\left(\frac{1}{C_0}-\frac{1}{C}\right)=u(V-V_{offset})^2\,, \tag{3.4}$$

its standard approximation

$$C=uC_0^4(V-V_{offset})+C_0^4 \tag{3.5}$$

yields a usable value for V_{offset}. Here $u \equiv (2k\varepsilon_0^2A^2)^{-1}$ and we identify the parabolic curvature α with uC_0^4. Now, an array of X-values can be constructed by $X_\mathrm{i} = C_\mathrm{i}^2(V_\mathrm{i} - V_\mathrm{offset})^2$ and an array of Y-values $Y_\mathrm{i} = C_i^{-1}$. If Eq. (3.4) is an accurate description of the data-points, the (X_i, Y_i) should lie on a straight line

$$Y=-uX+C_0^{-1} \tag{3.6}$$

The verification of this for the two mentioned $C(V)$-curves yields (Fig. 3.11ab).

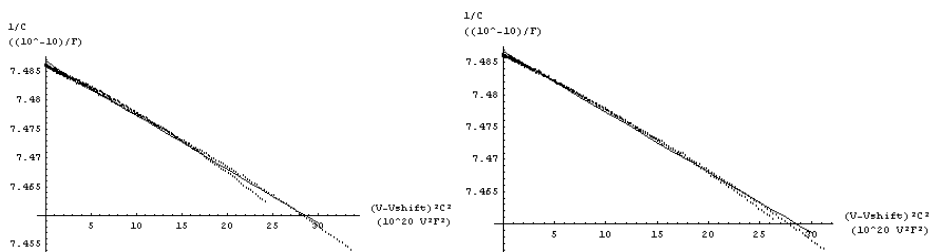


Fig. 3.11a (left). Compliance of C(V) data with straight line
Fig. 3.11b (right) Idem, for second series of C(V) data

Tiny but persistent aberrations from the ideal case (1)-(3) are to be remarked. In the first place, the data show convex curvature with respect to a fitted straight line. In the second place, for high voltages and capacitances (right side of the graph) the data sets split into two branches, pointing at a slight asymmetry in the capacitance values around the offset voltage: $C(+|V| - V_{offset}) - C(-|V| - V_{offset}) \sim 2$ fF for voltages higher than 30 V. These aberrations are more clearly demonstrated by subtracting the fitted line from the data (Fig. 3.12ab).

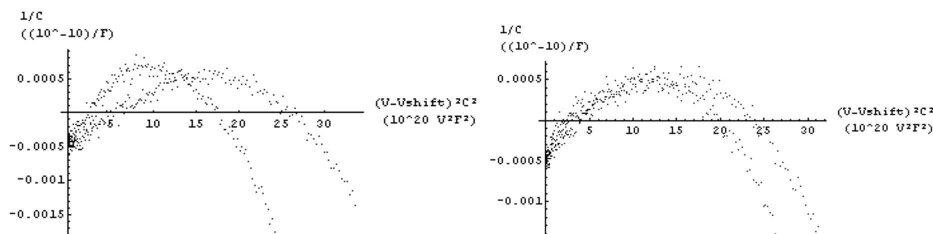


Fig. 3.12a (left) Data aberrations from implicit C(V)-curve
Fig. 3.12b (right) idem, for a different curve

Before deciding upon whether we are dealing with a genuine effect, the accuracy of the impedance analyzer should be considered. The applied DC bias had an error of ±(0.12% + 12 mV) which for the maximum applied voltage of 40 V comes down to ±0.05 V of systematic error. In the read-out capacitance a systematic error of ±0.25% is possible. For the capacitances under test this clearly exceeds the magnitude of the effect we are looking at. However, it should be noted that the indicated error is systematic, rather than random. Plausible candidates for explaining this effect, if any, are multiple; further measurements are needed to create more clarity. A few possibilities are mentioned: Deviations from a parallel plate geometry, on which Eq. (3.4) is based, are not expected to yield data sets asymmetric in the voltage. An additional force linear in the voltage, when parasitic charges are in the game, should only have an influence on the location of the offset voltage. However, as the measurement of a $C(V)$-curve is a time-extended process, time-dependent quantities in Eq. (3.4) could play a role. For the duration of a voltage sweep, the zero capacitance may drift; perhaps a relaxation-effect in the membrane, or perhaps the measurement system. If this drift can be considered linear during a sweep, at first glance it should have little influence: Let the unactuated capacitance drift as

$$C_0 = C_0(t) = C_{00} + ht \tag{3.7}$$

and the voltage be swept like

$$V = V(t) = V_{00} + ft\ . \tag{3.8}$$

Then in Eq. (3.4), the capacitance drifts linearly along with the voltage, with a rate (Farad/Volt) which is the ratio of the voltage increasing and the zero capacitance drift rate. In a parabolic approximation this would add a term linear in *V*, which has an effect on the offset voltage only. At a second glance, high drift rates make this argument problematic, so a further test has been performed, Fig. 3.13.

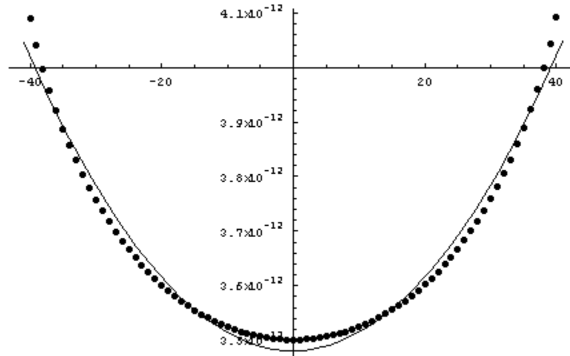


Fig. 3.13 Simulation of linear capacitance drift throughout C(V)-measurement

A 3.5 pF MEMS capacitor without parasitic charge is applied with a voltage sweeping from –40 V to +40 V in 60 seconds. Capacitive area and spring constant are 0.72 mm^2 and 8000 N/m respectively. The applied capacitance drift of 100 fF/hr is clearly higher than often seen in praxis (not more than in the order of 10 fF/h in stabilized conditions). The shortcomings of a quadratic fit are clear; though it might still have some use. The capacitance drift mimics a voltage offset of 46 mV.

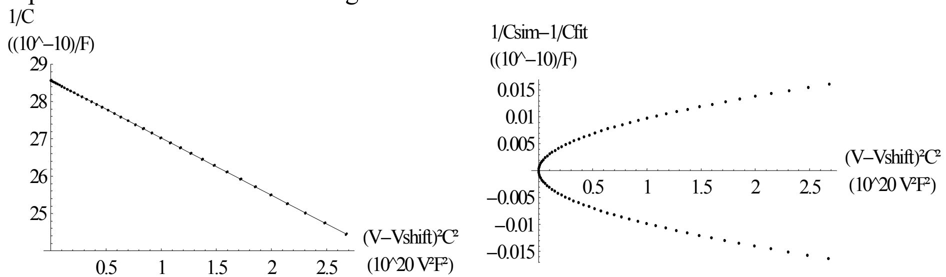


Fig. 3.14 (left) Compliance of simulated C(V) - data with straight line
Fig.3.15 (right) Simulated C(V)-data aberrations from implicit C(V) - curve

Although not immediately numerically comparable, the aberrations from (1) (fitted straight line in Fig. 3.14) are clearly different in form (Fig. 3.10) compared to the measured data in Fig. 3.12ab.

But an important candidate for explaining the effect is a time-dependent amount of parasitic charge, which changes significantly during the sweep. In order to increase the

capacitance by 2 fF, its gap has to be narrowed by only 0.5 nm. A charge density in the order of 10^{-6} C/m^2 is sufficient, which is not unrealistic and a density quite typical for devices like these.
Another important candidate is mechanical hysteresis, which should be looked at further in order to evaluate its possible influence on $C(V)$ - asymmetry.
The capacitance changes in the order of 10%, which is easily understandable when considering the composition of materials the structure is made of: silicon (substrate / 4.7), silicon nitride (bridge / 3,[21]) and chromium-platinum (top electrodes / 4.9-8.8). Numbers indicate thermal expansion coefficients in μm m^{-1} K^{-1}.

3.3 ALUMINUM BRIDGES

3.3.1 Devices

Characteristic for the second class of samples, UT/EMMA RF power sensors [6], is the aluminum bridge. It simultaneously functions as flexible bridge and as top electrode. For an example, see the SEM picture in Fig. 3.16. The exact dimensions of the samples measured in this section, see Fig. 3.17. The capacitive area is larger than that of the sample shown in Fig. 3.16, which is beneficial for the sensitivity of the device.

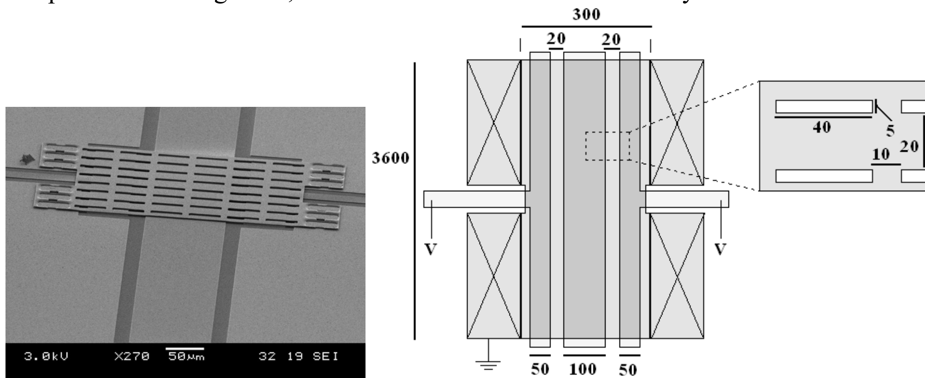


Fig 3.16 (left) Aluminum bridge MEMS
Fig. 3.17 (right) dimensions of the measured device in μm. Crosses indicate areas where the aluminum bridge (blue) is in contact with the (Pyrex) substrate. The inset shows the dimensions of the etch holes in the aluminum bridge. Dark areas represent where the electrodes overlap.

The bridge is suspended above aluminum electrodes, with boron-doped glass as substrate. The capacitive gap z_0 is designed to be 1 μm. This device can be considered as an alternative for / follow up of the silicon nitride bridges. In applications, this design has several advantages with respect to the silicon nitride bridges considered in the previous section. Firstly, the fabrication process is much facilitated compared to the SiRN samples. Secondly on-chip parasitic capacitance is practically eliminated. This is because the aluminum bridge samples have bottom electrodes that have well-defined and

[21] http://www.azom.com/details.asp?ArticleID=77

restricted shapes. The geometrical overlap of the top and bottom electrodes occurs only at the movable bridge. In case of the silicon nitride bridges, the complete substrate acted as the lower electrode. This introduces parasitic capacitances to the bonding pads, as explained in the previous section.
Thirdly, thick dielectric layers are absent in the aluminum bridges. Only 2.5 nm natural aluminum oxide is present. Based on the discussion in section 2.3, it is expected that the effects of charge trapping will be reduced for a couple of reasons. In the first place, with thin layers there are less (bulk) sites to trap charge in. In the second place, thick dielectric layers enhance the influence of the trapped charges on the total electrostatic force. At last, in thick dielectric layers, transport mechanisms other than metal-surface tunneling will start to become important, which affects the migration behavior of trapped charge.
As a last (fourth) advantage, improved thermo-mechanical properties might be mentioned. In [7], pp 65, this expectation is put forward. This is based on that the aluminum bridge consists of a single material. In contrast, the silicon nitride bridge has electrodes that have considerably different thermal expansion coefficients. The latter structure acts as a bilayer that potentially deforms when subjected to thermal variations. The aluminum bridge should not suffer from this effect. However, this does not take into account the potential influence of the substrate. This had already been signaled in [8, 9]. The thermo-mechanical properties of this device will be considered in the next subsections, 3.3.2 and 3.3.3. Subsection 3.3.2 discusses an experiment in which a series of $C(V)$-curves has been quickly recorded during a trajectory in which a sample was rapidly cooled down. It is used as a first test of how the characteristics of the device change with temperature. The experiment covered in subsection 3.3.3 consists of series of $C(V)$-curves at *stabilized* temperatures. Here, the stiffness is explicitly linked to the change in temperature and discussed. The relatively simple design of the aluminum bridge encourages attempting such an analysis. On the other hand, the complicated silicon nitride bridges provide less tangible handles to proceed thus.
At last, in section 3.3.4 the aluminum bridge will be tested for charge trapping effects. The role of temperature in these experiments is prominent.
All aluminum bridge measurements have been performed in the flow cryostat, as described in subsection 3.1.2.

3.3.2 Influence of rapidly changing temperature on aluminum bridge characteristics

The measurement series presented in Fig. 3.18 and 3.19 was taken during a cooling stage from room temperature down to approximately 2 K. The total time of the series was 3104 sec. In this period 304 $C(V)$-curves were recorded, taking 10.4 seconds per curve, including data saving for each curve. Each curve ranged from –3 to + 3 V (or vice versa). A curve consists of 13 sample points spaced 0.5 V. The temperature decreased linearly with time during the whole series except for the first and last few minutes. This implies that within the time-span of a complete $C(V)$-curve, the temperature dropped with about 1K. Charge-related and other effects that produce small shifts in these curves are therefore not to be distilled from this kind of series (which was never intended), but parabolic fits to the data are satisfactory and provide reasonable values for capacitance and curvature.

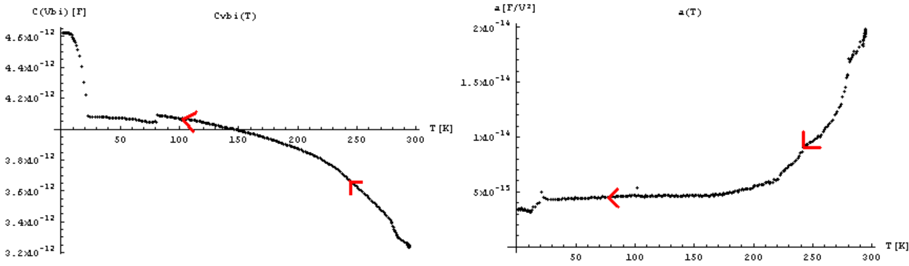


Fig. 3.18 (left) Capacitance vs. temperature; arrows
Fig. 3.19 (right) Curvature vs. temperature indicate direction of time.

In Fig. 3.18 we observe that the capacitance increases when the temperature goes down. At some point, the rate of increase slows down. Suddenly, below 20 K there is a sudden increase. In Fig. 3.19 a reverse trend is noted. The parabolic curvature, a measure for the sensitivity of the device, decreases rapidly with decreasing temperature. Even more pronounced than in Fig. 3.18 is the fact that the biggest change in the parameter occurs at still relatively high temperatures. Also here "something happens" at around $T = 20$K, but it is less noticeable.
By far the most plausible explanation for the gradual but considerable change in capacitance between 300 K and 30 K is a decrease of the capacitive gap. Autonomous capacitance drifts of this order are not expected.
But this behavior of the characteristics offers a few more points of concern. Namely, at first thought more plausible seems the reverse trend: *decreasing* capacitance with decreasing temperature. Suppose one starts with a bridge with a slightly concave (hollow, curved upwards) shape. The aluminum bridge has a thermal expansion coefficient of 23 µm m^{-1} K^{-1}. The glass substrate comes out at only 3 µm m^{-1} K^{-1}. Upon cooling down, the aluminum bridge wants to contract considerably more than the substrate. This would stretch the bridge straight like a taut sheet in case of a pure and strict clamped-clamped situation. This would increase the gap and decrease the capacitance. However, there are a few possibilities. The first is that the bridge starts at room temperature in a *convex* (curved downwards) and (relatively) stress-free state. Cooling down and stretching would now decrease the gap and increase the capacitance. A second possibility is hypothesized in Fig. 3.20 [7].

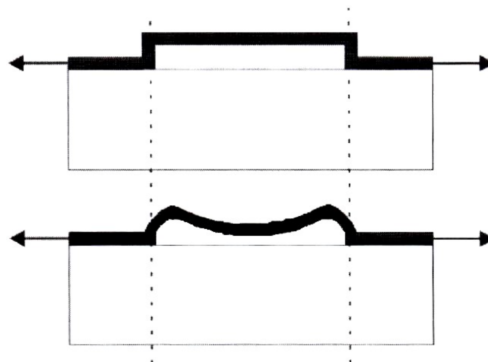


Fig. 3.20 Example of tensile stress causing sagging of the bridge [7]

The contracting bridge pulls the suspension corners out of the plump line towards each other. This forces the bridge in a concave shape, which decreases the gap.
A second concern is the sudden 'jump' in the capacitance below ~20 K, as was already mentioned. No convincing explanation has been found for this, but a hypothesis to explain this anomaly is shortly discussed at the end of subsection 3.3.3.
The third concern is that the curvature shows a trend in the temperature that is reversed with respect to the capacitance. This is counter-intuitive when recapitulating Eq. (2.11), relating capacitance C_0 to curvature α. In this relation, k is the spring constant and A is the capacitive area. Irrespective of a change in the unactuated capacitance is due to a change in capacitive surface A (quadratic dependence of curvature), or due to an altered gap (quartic dependence), an increased capacitance is expected to increase the curvature[22]; less electric force (voltage) is needed to effectuate the same amount of absolute capacitance increase.
Further considerations touching upon the effects observed here will be discussed on the basis of temperature-stabilized measurement series presented below.

3.3.3 Influence of varied stabilized temperature on aluminum bridge characteristics

Even if the temperature is stabilized to a fixed value, the mechanical characteristics of the aluminum samples show practically the same dependence on the temperature (Fig. 3.21, 3.22) as in the case of a rapidly declining temperature, as presented in the previous subsection.

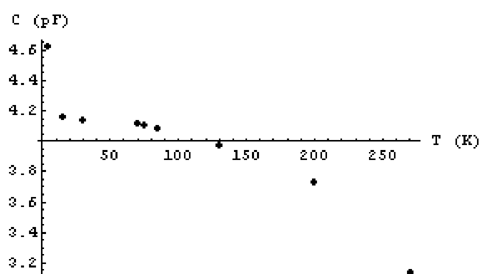


Fig. 3.21 Thermal dependence of C

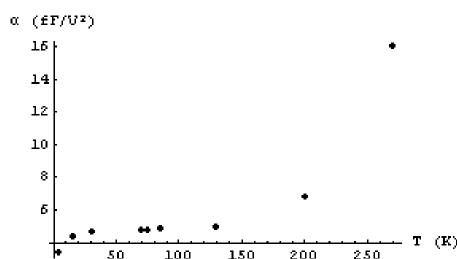


Fig. 3.22 Thermal dependence of curvature

Measurements were carried out at temperatures in a range of 4-270 K on the exact sample in subsection 3.3.1. Temperatures were stabilized and curve series lasted about 25 minutes. A single curve took 40 seconds to record 31 sample points, between –3 V and 0 V (not +3 V, as in subsection 3.3.1), so that the voltage stays unipolar during a curve. The thermo-mechanical results for 0 V to +3 V - curves are similar to those presented in the current subsection. Assuming at 270 K a gap of 1 µm, in accordance to the sample design, a spring constant can be estimated from the capacitance and the curvature, by making use of Eq. (2.11), See Fig. 3.23a.

[22] This 'expected' correlation has indeed been observed for different samples (Capacitance still increased for decreasing temperatures). However, also here the change in capacitance and curvature do not match unless the stiffness is assumed to vary considerably.

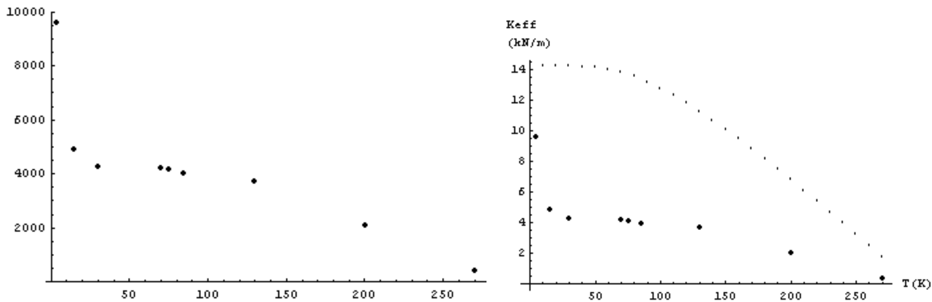


Fig. 3.23a Spring constant vs. temperature
Fig. 3.23b Experimentally derived spring constant derived from the parabolic curvature together with an estimate by calculation assuming a capacitive gap of 1 μm

Fig. 3.23b combines these data with a calculated value for the spring constant, which has sample geometry, material properties (linear thermal expansion coefficients of bridge and substrate, Young's modulus, Poisson ratio), stress due to thermally-induced strain and an offset residual film stress as input. The material properties are in decreasing order of temperature dependence, which has been taken into account for all of them. For the residual film stress a value of 35 MPa [6,7] was taken.
In order to accommodate the observed trend in curvature with the trend in capacitance, the spring constant needs to vary for more than an order of magnitude. As calculation shows in Fig. 3.21b, a build-up in tensile stress can very easily accomplish this. A further argument for the spring constant indeed being influenced by thermo-mechanical stress is the flattened trend between 15 K and 130 K with respect to 130 K-270 K (Fig. 3.23a). At lower temperatures, the thermal expansion coefficient of aluminum tapers off to zero, so that further cooling down adds relatively little additional stress; be it that calculation predicts this flattening to happen at lower temperatures, say 70 K. At these temperatures, the plate modulus of aluminum has increased by about 13% compared to room temperature, which predicts the flattening to happen at somewhat lower temperatures; this is a mild effect however.
Obviously, far stronger hitting the eye is the difference in magnitude between experimentally-derived and calculated spring constant. As displayed in Fig. 3.23b, two considerations have already been taken into account. The first one is that the aluminum bridge, like the silicon nitride bridge, is perforated with long rectangular etch holes. They make the bridge effectively narrower by about (only) 20%[23], a correction which has been put through in the calculated spring constant. At the same time, the capacitive area is reduced by 16% because of these holes. The measured room temperature value of the capacitance of 3.2 pF is in very good agreement for the value of a capacitor with an unperforated bridge. As parasitic capacitance is mainly stray capacitance (about 0.2 pF), the 'correct' value for C_0 is restored by decreasing the gap. This increases the from measurements interpreted spring constant by about 40% ($1/gap^2$).

There are some important remarks to be made, however. The parabolic curvature α Eq. (2.11), from which the spring constant is derived, contains the capacitance to the fourth

[23] The picture in Fig. 3.16 gives a distorted impression in this respect, where the etch holes seem considerably larger. The bridge for that particular sample spans 300 μm and is 200 μm wide, though the picture suggests a much larger aspect ratio. The sample actually used for the measurements discussed here measured 3600 μm width instead of 200 μm.

power. If the bridge capacitance is not determined very accurately, for example when misestimating the parasitic component in the measured total capacitance, errors in the estimated spring constant blow up very quickly (although a parasitic capacitance will affect the spring constant only *quadratically*). The larger the portion of parasitic capacitance in the measured capacitance (which is expected small, however), the lower the estimated spring constant will be. After all, the fitted curvature is unaffected by the assumption of a parasitic capacitance, so that the same change in capacitance (upon DC bias actuation) has to be caused by a smaller capacitor, which can only be done by having weaker springs.
More important is the distribution of the electrodes. The spring constant is calculated by expression Eq. (2.66), which is however valid for a centered electrode. The actual electrodes are located more off-center, which will increase the effective stiffness and wider the gap between calculation and experiment.
Possibly partly countering the effect of the electrodes is if the stretching behavior of Fig. 3.20 is assumed: The points which the bridge is clamped to are not rigid but are flexible.
By the way, in-plane stress perpendicular to the beam axis will be present at the areas of contact between the bridge and the substrate, but rapidly subside further away from these areas. It is expected that this will have a comparatively low influence on the elastic behavior of the beam. Nevertheless, it is advised that this is investigated further by use of for example finite element simulations.
An extra complicating factor is the large spatial extension of the bridge, which can show non-uniformities over its area, that can especially come to light when the bridge is exposed to stress.

At last, concerning the sudden increase in capacitance, the speculative idea of the bridge touching the central floating electrode is briefly considered. A large jump in stiffness is easily understood as the effective beam length is suddenly cut in half. However, this increase is rather associated with a considerable discontinuity in the curvature rather than the capacitance. The opposite is observed though: the 20K-jump in the capacitance (Fig. 3.18) is much larger than the corresponding discontinuity in the curvature (Fig. 3.19).
Let us first consider the stiffness. It is assumed that the bridge undergoes a gradual downward deformation when the temperature is decreasing. Below a certain temperature it will make a physical contact with the center electrode. In practice, it is expected that his contact will only be at one or a few small areas, “contact points”. In this case, the stiffness should be increase less than would be the case when the bridge makes physical contact over its full width (3600 µm, the vertical direction in Fig. 3.17).
Now consider the capacitance. When the bridge makes *conducting* contact with the central floating electrode, the latter becomes a grounded electrode like the bridge. Using a 2D finite element solver, an estimate has been made of the extra stray capacitance trough the substrate (Fig. 3.24). As soon as the central electrode becomes grounded, an extra stray capacitance of 0.2 pF adds to the measured capacitance. Not quite sufficient to explain the large step (0.5 pF), but not orders of magnitude off. Also, close inspection of Fig. 3.18 suggests that the increase of 0.5 pF does not happen in a single ‘leap’, but rather in a quick but smooth variation. Further deformation of the bridge also keeps on increasing the variable MEMS capacitance. In any case, the stray capacitance is a *parallel* capacitance. It does not add to the sensitivity of the MEMS device. It should not have an effect on α.

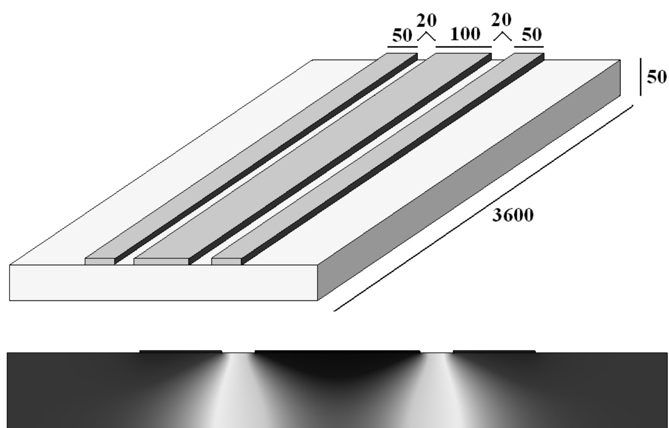


Fig. 3.24: Cross section of a piece of substrate with three parallel electrodes; Numbers are dimensions in µm. and indicate the values that have been used. Potential ranging from 0 V (blue, center) to 1 V (red, flanks)

The described scenario is consistent for explaining a sudden capacitance increase, while at the same time the curvature shows a rather small *decrease*. Without anything special happening, such as for example physical contact as described above, Eq. (2.11) shows that only a *large increase* of α is compatible with a capacitance increase.

3.3.4 Built-in voltage

Fig. 3.25a shows the built-in voltage (or offset-voltage) plotted against time for a small MEMS capacitor, in a series lasting for over four days at room temperature.

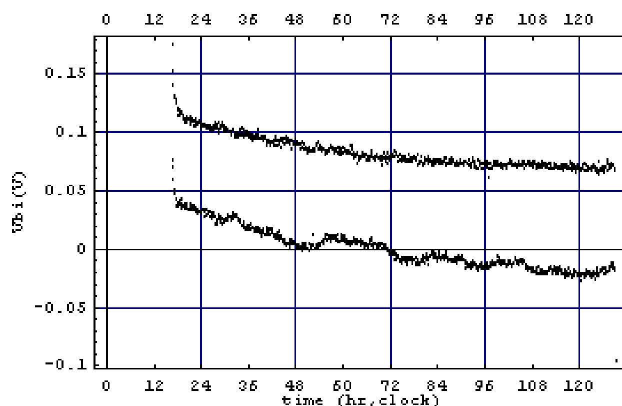


Fig. 3.25a Built-in voltage vs. clock time

Fig. 3.25b shows the capacitance in this minimum.

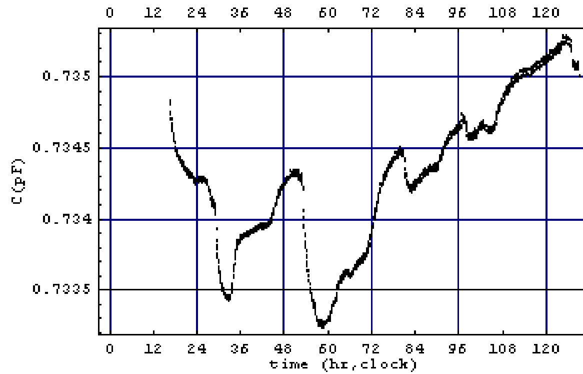


Fig. 3.25b Capacitance vs. clock time

Again, the two branches in the built-in voltage arise from the series having the curves in alternating sweep-directions. The minimum of the curve swings back and forth for every sweep. In the meantime a slow, exponential overall drift is visible.
As remarked in section 2.3 the potential causes of built-in voltage are manifold. As the grounded and actuated electrode are of identical material, a difference in work function is not expected and henceforth a (constant) contribution. The capacitance trend as displayed in Fig. 3.25b shows drift speeds up to 1 fF/hr, which is regarded as too low for being able to mimic built-in voltages of 0.1 V. The capacitance does not split into two branches, denying rapid capacitance changes being a cause for the built-in splitting.
Meanwhile, though no temperature data are available from this measurement, the 24hr period of smooth changes in the capacitance, which is especially visible in the first half of the graph, tempts to link the mechanical state of the tested device to be in direct correspondence to the surrounding (laboratory) temperature. From subsection 3.3.2, Fig. 3.18 it is clear that temperature fluctuations of a few Kelvin are more than enough capable of swinging the capacitance for the 1.5 fF that is observed in Fig. 3.25b.
Linking the actual value of a built-in voltage to a temperature dependency is problematic, as the exact history of the sample is important, for example. What is expected to be temperature dependent are dynamic aspects of built-in voltage; long term drift and splitting (back sweep and forward sweep yield different values, cf. Fig. 3.25). A temperature dependence of this splitting has been repeatedly observed. The built-in voltage of a $C(V)$ - curve series during gradual warm-up, is displayed in Fig. 3.26.

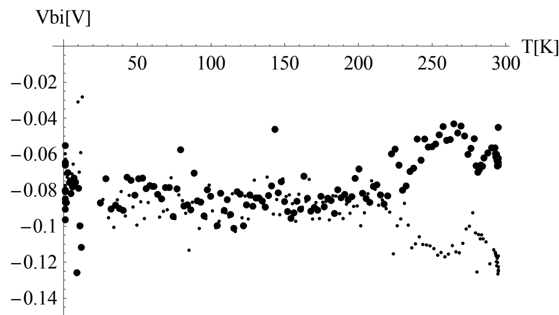


Fig. 3.26 Built-in voltage vs. temperature

Remarkably, in a big range of temperature there is no significant splitting, which only appears quite suddenly above approximately 200 K.
For the observation of the dynamic built-in voltage at stabilized temperatures, only unipolar *C*(*V*)-curves (cf. subsection 3.3.3) were measured, to attempt to enhance long-term drift. For a single curve, the voltage is swept between either –3 V and 0 V or 0 V and +3 V in steps of 0.1 V in 40 seconds. At each temperature (15, 30, 70, 75, 85, 130, 200 and 270 K) a series of negative and a series of positive 'halfcurves' was recorded, each taking 20-30 minutes. For both polarities, built-in splitting spread between –0.1 and +0.1 V, and drifts were in the order of –0.5 to +0.2 V/hour. In no case was a clear trend with the temperature discernible. It is immediately to be added that the error bars belonging to temperatures of 85 K and higher are itself several tenths of volts. It is insightful to single out a particular series, for example the positive half-curves at 130 K, for which built-in voltage and minimum capacitance versus time are given in Fig. 3.27ab:

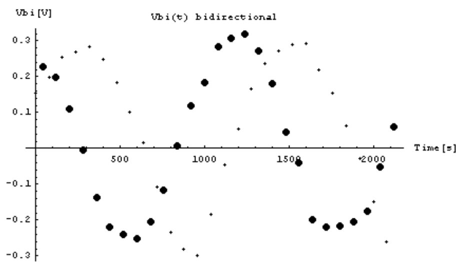
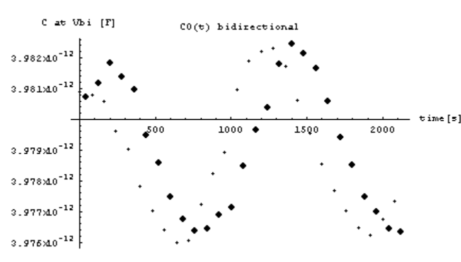


Fig. 3.27a Built-in voltage vs. time



Fig, 3.27b Capacitance vs. time

Fig. 3.28 combines them in a single graph, displaying the location and the value of the minima in the repeatedly parabolic fitted data in this *C*(*V*) – series.

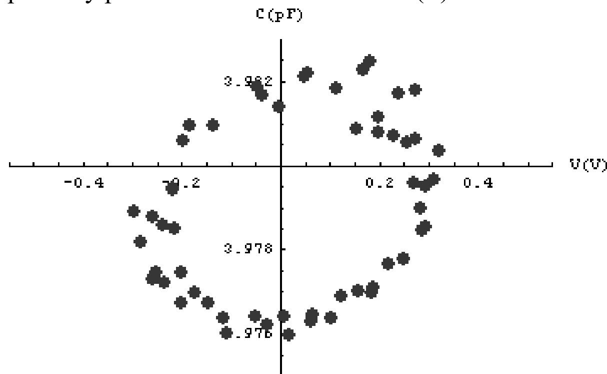


Fig. 3.28 Locus plot of C(V) minima, running just over two cycles in time.

The oscillatory trend of both built-in voltage and capacitance has a period of about 20 minutes. For these series no curve-to-curve temperature data are available. However, a *C*(*V*) - series in the exact same measurement setup for a silicon nitride bridge at 100 K did contain detailed temperature data, which showed a fluctuation with an amplitude of about 1.5 K in a comparable (slightly shorter) period. From Fig. 3.18 it is deduced that capacitance changes with temperature with almost –4fF/K. Consequently, thermal fluctuations resulting from a non-perfect temperature control system are capable of causing the observed oscillations of the capacitance. Assuming a sinusoidal temperature trend that changes 128.5 K to 131.5 K peak-to-peak in half a period (10 min.),

the maximal change rate (where the sinusoidal slope is steepest) is ½π × 3 K / 10 min. = 28 K/hr, which in turn would cause a maximal capacitance drift of about 100 fF/h. These high drift rates could cause a small but substantial portion of the built-in voltage, as estimated by the simulation shown in Fig. 3.13. This does not explain many phenomena like the phase shift in backward and forward curve built-in voltage, nor the full value of the built-in voltage, for which trapped charges and drifts therein still can be held responsible. However, it does show that temperature fluctuations claim a significant co-authorship concerning some of the observable phenomena. This illustrates that research of charge-trapping related phenomena meets with problems if performed on the aluminum bridge structures described in this chapter.

3.4 CONCLUSIONS AND DISCUSSION

The capacitance of the silicon nitride bridge HF power sensors shows clear dependence on the temperature. A plausible explanation for an observed small asymmetry in the $C(V)$ curves (measured in stabilized conditions) is charge movement during the course of the measurement of a $C(V)$-curve. This effect is however smaller than the possible systematic error of the impedance analyzer. Random noise in these curves is lower than the magnitude of the effect though.

Thermo-mechanical properties of the aluminum bridge RF power sensors are even worse; as smooth and repeatable capacitance changes up to 30% have been observed, against 10% for silicon nitride bridges. The possibly at first sight counterintuitively reversed trend of the curvature with respect to the capacitance change can be explained by a dramatic increasing stiffness of the beam. Calculations for the spring constant predict an influence of thermo-mechanical tensile stress that is even considerably higher than is interpreted from $C(V)$-curves. Too many complicating unknown factors can play a role for further resolution. A flattened trend for the stiffness at low temperatures is both seen in experiment and predicted by calculation, as a result of the decrease in the thermal expansion coefficient of the bridge material aluminum.

For both the silicon nitride bridge samples and the aluminum bridge samples dynamic aspects of built-in voltage (long term drift and splitting) have been observed, but especially in the latter case the effects have been demonstrated to be obscured by heavy dependence of the $C(V)$ - characteristics on the temperature. Therefore, in order to proceed with device-level study of temperature-dependent built-in effects, it is very desirable to employ structures which mechanical characteristics possess highly improved immunity against variations of ambient temperature.

REFERENCES

[1] The original website http://www.vtt.fi/tte/research/tte7/research/emma/index.html is no longer accessible at the time of writing (October 2008). A cached version is still available though. The announcement of the EMMA project can be viewed via the search function of the website http://cordis.europa.eu .
[2] A. Kärkkäinen, N. Tisnek 1, A. Manninen 1, N. Pesonen, A. Oja, and H. Seppä, "Electrical stability of a MEMS-based AC voltage reference", *Sensors and Actuators A* **137** 169 (2007)
[3] H. Seppä, J. Kyynäräinen, and A. Oja, "Microelectromechanical systems in electric metrology", IEEE Transactions on Instrumentation and Measurement **50** 440 (2001)
[4] J. Kyynäräinen, A. S. Oja, and H. Seppä, "Stability of microelectromechanical devices for electrical metrology", IEEE Transactions on Instrumentation and Measurement **50** 1499 (2001)
[5] B. de Jong, "Fabrication, modeling and characterization of test-devices for a high-frequency power sensor", Master thesis, University of Twente, September 2002
[6] L. J. Fernández, R. J. Wiegerink, J. Flokstra, J. Sesé, H. V. Jansen, and M. Elwenspoek, "A capacitive RF power sensor based on MEMS technology", *Journal of MicroMechanics and Microengineering* **16** 1099 (2006)
[7] L. J. Fernández, "A capacitive RF power RF power sensor based on MEMS technology", Ph.D. Thesis, University of Twente, December 2005
[8] M. C. van Essen, J. Sesé, L. J. Fernández, J. M. Reitsema, R. J. Wiegerink, and J. Flokstra, "Characterization of Capacitive MEMS at Cryogenic Temperatures", *proceedings of 6th Workshop on Low Temperature Electronics (WOLTE), Noordwijk, The Netherlands, 2004 pp. 137*
[9] M. C. van Essen, J. Sesé, L. J. Fernández, J. M. Reitsema, R. J. Wiegerink, and J. Flokstra, "Characterization of Capacitive MEMS at Cryogenic Temperatures", *proceedings of 15th MicroMechanics Europe (MME), Leuven, Belgium 2004 pp. 92*

CHAPTER 4

THERMO-MECHANICALLY STABLE MEMS

The primary goal of the research presented in Chapters 3 and 4 is to demonstrate temperature-dependent charge trapping effects. In the previous chapter it has been demonstrated that the characteristics (capacitance, sensitivity) of the EMMA/UT HF power sensors were amenable to fluctuations in temperature, even to such an extent that possible temperature dependence of charge trapping effects could not be decoupled. The secondary, serving goal to be addressed in this chapter is the realization of new test structures that are sufficiently thermo-mechanically immune. Section 4.1 describes the design and fabrication of these structures, based on two wafers: The top wafer provides a flexible element carrying a floating electrode; the bottom wafer supports two fixed electrodes. As fabrication, the top wafer undergoes wet bulk micromachining. A thorough account of this technique is given in [1, chapter 4] and incorporates the novel concept of double wet-etched beam springs, demonstrated for the first time in this thesis. To the bottom wafer, only surface micromachining is applied to create a capacitive gap embedding the electrodes. Section 4.2 describes a preliminary electromechanical test of one of the bulk micromachined top structures by combining a C(V) - measurement with Phase Shift Interferometry. Section 4.3 contains electromechanical characterization of complete devices in order to accurately determine the stiffness of the devices. In section 4.4, firstly the thermo-mechanical stability is verified. As stated above, this is a key property. Then, the temperature-dependent behavior of charge trapping in these devices is investigated. Section 4.5 contains a high-vacuum experiment at room temperature. Finally, conclusions are summarized in section 4.6.

4.1 DESIGN AND FABRICATION

4.1.1 General design description and considerations

The demand of thermal stability imposes minimal differences in the respective thermal expansion coefficients of the bottom and top parts of a parallel plate configured capacitive MEMS, for this, a two-wafer approach (see Fig. 4.1) is a good option, which provides flexibility in the choice of electrode material and offers a monocrystalline (silicon) suspension, which minimizes mechanical creep and hysteresis.

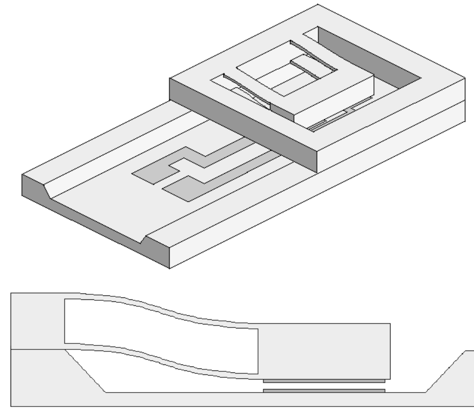


Fig. 4.1 Two-wafer capacitive MEMS

The top structures have been realized in a variety of forms, but the double springs are expected to keep the electrodes parallel during actuation even for one-sided suspended structures like in the illustration. These springs can be realized thin enough for sufficient sensitivity. The electrode material can be deposited on both sides to eliminate thermo-mechanical curvature of the structure.

For electric probing of the top electrode several solutions come into consideration, but this is not really necessary, if two bottom electrodes are used instead of one. The result is two capacitors in series, across which the applied voltage will be distributed.

Trapped charges will manifest itself via a constant offset voltage. This design has the obvious advantage of easy fabrication. There will be a loss of sensitivity compared to a non-floating top electrode. Still, the device can be made easily sufficiently sensitive for the purpose of charge trapping research.

The configuration of the electrodes is schematically drawn in Fig. 4.2.

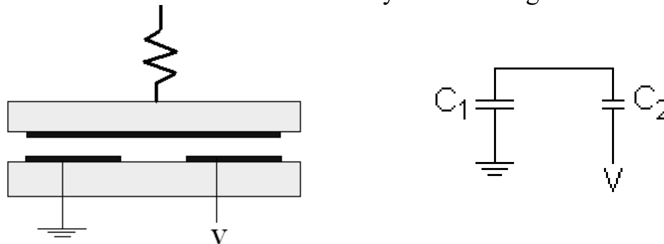


Fig. 4.2 Schematic electrode configuration

The sample capacitance consists of two, possibly unequal, capacitances in series. This effective series capacitance is valued

$$C_s = \frac{C_1 C_2}{C_1 + C_2} = \frac{\varepsilon_0 A_s}{z}\ , \qquad A_s \equiv \frac{A_1 A_2}{A_1 + A_2}\ , \tag{4.1}$$

and is smaller than each of the partial capacitances C_1, C_2.

The partial capacitances are intended to have equal gaps, wich justifies the second equality sign in Eq. (4.1). The 'series' area A_s is defined in terms of the capacitive areas A_1, A_2 of the partial capacitances and is the equivalent area of a single capacitor with gap z replacing the series capacitors. Because the total voltage V distributes over the two series capacitors, the voltages V_1, V_2 over the gaps are lower than would be in case of a connected top electrode. Consequently the magnitude of the electric force over the bridge is lower and comes out at:

$$F_{el} = -\frac{C_1}{z}V_1^2 - \frac{C_2}{z}V_2^2 = -\frac{\varepsilon_0 A_s}{z^2}V^2 , \qquad (4.2)$$

which demonstrates that this two-capacitor configuration can be fully described in terms of an equivalent single parallel plate capacitor with electrode area A_s. Now given a top electrode has a total capacitive area A and partial areas $A_1 = pA$, $A_2 = (1 - p)A$, $0 < p < 1$, the series capacitance C_s is (naturally) maximal for $p = ½$. In this case, $A_s = ¼A$. Consequently, for a fixed total voltage V over the device, the force on the bridge of the current three-electrode configuration is reduced with at least a factor 4 compared to a capacitor with two parallel plate electrodes of area A.

4.1.2 Top structure design

The top structures appear in frames each containing six members, see Figure 4.3. The total wafer mask is mirror symmetric so that it can be used on both sides of the wafer. Structures appear in two series of frames of each containing six structures, in which various forms are realized and tested.

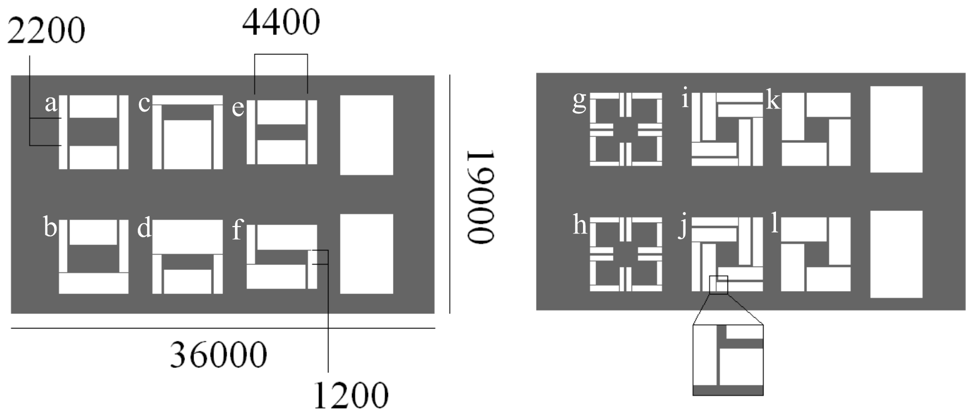


Fig. 4.3 Designed structures and dimensions in µm; the inset magnifies a corner etch prevention strip, measuring 1000 × 20 µm², see sub-subsection 4.1.5.3.

The frames measure 19 × 36 mm². The left frame contains one- and two-sided suspended structures; the right contains various four-sided suspended structures. All structures, except the top left and the top right in the left frame, and the four-leaf clover in the right frame are expected to release any remaining thermo-mechanical stress partly by freedom in geometry. But neither in the 'rigid' structures is significant deformation upon cooling expected. The top carrier and its suspension are of identical material (silicon) as the

surrounding frame. The empty spaces in the frames on the right sides are for contact to the bottom wafer electrodes. The structures will be formed by wet bulk micromachining. More on this process can be found in subsections 4.1.3-4.1.5.

The choice for the structures (labeled **a**, **b**, … **k**, **l**) is motivated as follows: Structure **a** has a relatively large capacitive area and is always parallel. **b** has the same area but the suspension is weaker. This makes for a more sensitive structure. Even with the springs only on one side, it is expected that the structure always remains parallel, in accordance to what has been stated in 4.1.1. If because of a remaining bilayer working there is some deformation effect or stress build-up, the structure has more freedom than **a** to alleviate this stress. **c** has springs of double length compared to most other structures. This greatly enhances sensitivity. **d** is a variation of **b** but with half the capacitive area. **e** is a variation of **a** but with half the capacitive area likewise. **f** can be considered a variation of **d**. Any thermo-mechanical stress can be alleviated by rotation of the suspended mass. The four-sided clamped structures (right frame) are pairwise identical and mirrored. The clover-shaped structures **g** and **h** are identical. Given the available space, they maximize the capacitive area while the springs are as simple as possible and not too short. Structures **i** and **j** have very long springs, which enhances sensitivity more than for example **c** while having the same capacitive area (4 mm^2). **k** and **l** are relatively rigid structures which can act as comparison structures for **i** and **j**. The latter four can release any remaining stress by rotation.

Bare corners of structures are especially vulnerable and should be protected against excessive etching. A common measure is to adopt sacrificial tree-like structures in the corners. In this case it was chosen to artificially 'conceil' the corner by elongating one mask edge with a narrow strip that can easily be removed with tweezers afterwards if desired, see the inset of Fig. 4.3. Some more spring sizes are given in the 3D–drawing in Fig. 4.4.

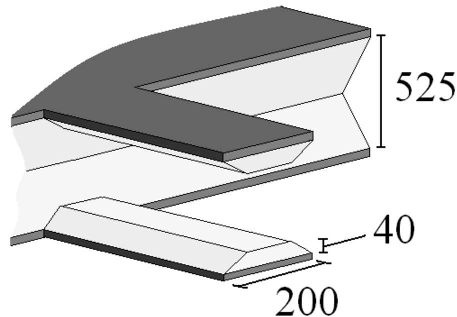


Fig. 4.4 Spring width and thickness and wafer thickness

It is assumed that the deflection of the rigid top plate under a distributed load (spread force) is equal to the deflection of a clamped-clamped beam under a centered point load (concentrated force, equal in magnitude to the spread force). For the calculation of the stiffness (spring constant), the first two cases sketched in Fig. 4.5 are both treated on equal footing as the third case. A structure containing N *pairs* of beam springs each of length l, (so a total of $2N$ springs), is thought to deflect equivalently to N parallel clamped-clamped beams of length $L = 2l$, see Fig. 4.5.

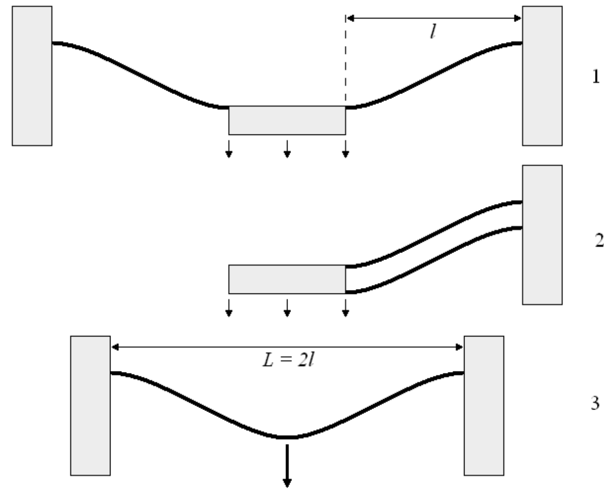


Fig. 4.5 The spring configurations 1 and 2 with a spread load on the rigid plate are assumed to show a deflection equal to a clamped-clamped beam subjected to a centered point load if the total force is the same.

For small deflections, a clamped-clamped beam subjected to a centered point load deflects according to

$$\Delta z = F / k \,, \tag{4.3}$$

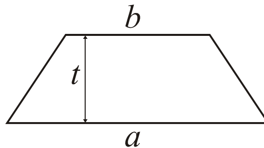
in which the stiffness k for prismatic (shape and area of cross section is uniform along the length) beams is given by

$$k = \frac{192EI}{L^3} \,, \tag{4.4}$$

where L is the total length of the beam, which is twice the length of the fabricated beam length l. The product EI, called *flexural rigidity*, contains the plate modulus E, which is the Young's modulus Y modified by Poisson's ratio ν according to

$$E \equiv \frac{Y}{1-\nu^2} \,, \tag{4.5}$$

which takes into account the slendering of a beam upon stretching.
For a trapezoidal cross section (see below left), the second moment of inertia (also called "area moment of inertia") I is given by [2]



$$I_{trap} = \frac{h^3(a^2 + 4ab + b^2)}{36(a+b)} \,, \tag{4.6}$$

which is cubically dependent on the thickness h and features symmetrically the lower width a and the upper width b. For a rectangular cross-section ($a = b$) and a triangular cross section ($b = 0$) the stiffness (4.4) reduces to the familiar forms

$$k_{rect} = \frac{16Eah^3}{L^3}, \qquad k_{tri} = \frac{16Eah^3}{3L^3}. \tag{4.7}$$

The former (k_{rect}) was already encountered in Eq. (2.73).

4.1.3 Fabrication process top structures

The main process steps of the fabrication of the top structure are summarized in Fig. 4.6.

Firstly, a monocrystalline (100), double sided polished, p-type silicon substrate receives a double sided coating of reaction bonded silicon nitride (SiRN, in stoichiometry: Si_3N_4) of 800 µm via Low Pressure Chemical Vapor Deposition (LPCVD). This layer will become a mask for the wet-etching process later on. The patterning is done by spinning primer (HexaMethylDiSilazane, or HMDS) and 1.7 µm of positive photo resist (Olin 907-17) on both sides. On one side standard photolithography is applied. After the photolithography, the bare parts of the silicon nitride layer are removed by plasma etching. As the plasma also attacks the photo resist layer, care must be taken that it is sufficiently thick and firmly baked out (30 min on 120°C). Now one side of the SiRN-layer is patterned. The photo resist is removed and reapplied on both sides, after which the other side of the wafer is patterned. Obviously, this requires a mask aligner that is able to align a mask to the back side of the wafer. This step is sketched (1) in Fig. 4.6.

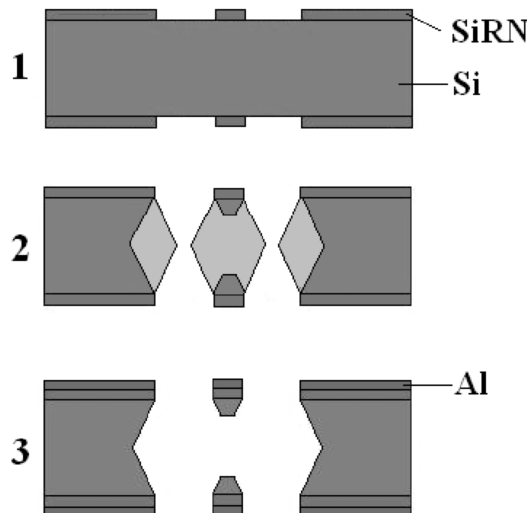


Fig. 4.6 Process steps top structure

Secondly, the springs and structures are formed out of the silicon substrate by wet-etching in an aqueous solution potassium hydroxide. At first, the wafer is etched through vertically (the white spaces in (2) in Fig. 4.6). Prolonged submersion in the etch fluid will then lead to the removal of the light grey areas in (2), after which of the centre part of (2) only the dark colored areas remain, which represent cross sections of the ultimate beam springs. The exact moment of ceasing the etching process determines the thickness of

these springs and thus their spring constant. The etching process will be considered in somewhat more detail in the next subsection. After the etching process, the structures will be provided with a conductor, in our case 100 nm thick aluminum. The conductor is sputtered on both sides, which results in the springs being bilayers (actually tri-layers) that are vertically mirrored (3). This minimizes mechanical deformations due to changes in temperature.
Some top structures from just after the etching are displayed in Figs. 4.7-4.9. No aluminum electrode layers had been applied yet at this stage. Fig. 4.9 already discloses the feasibility of the fabrication concept of double springs [6].

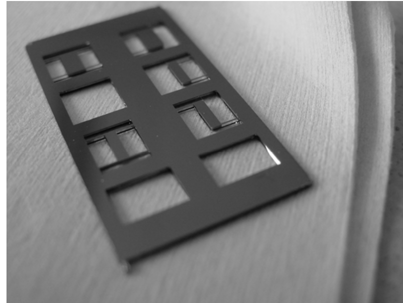


Fig. 4.7 One and two-sided suspended structures

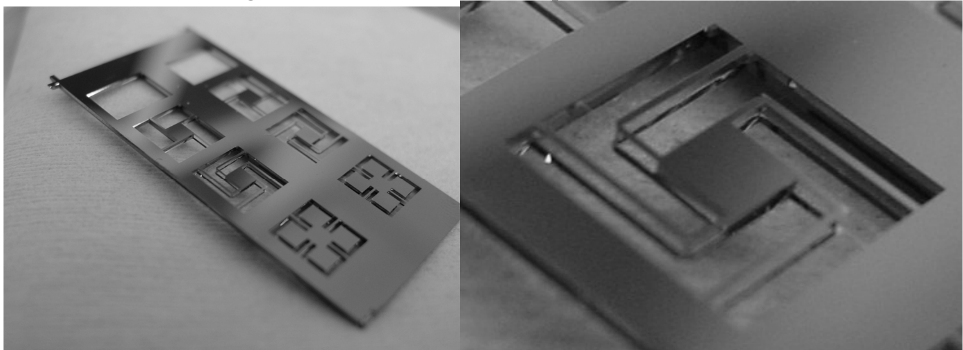
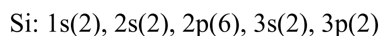


Fig. 4.8 Four-sided suspended structures　　*Fig. 4.9 Detail of Fig. 4.8*

4.1.4 Wet etching of silicon [1,3,4]

The wet-etching of the top-structures, as shortly described in the previous subsection on the basis of Fig. 4.6 (2), is done by submersing the wafer in an aqueous solution of 25 weight percent of potassium hydroxide (KOH), kept at a temperature of 75°C. To consider the process of wet-etching in some more detail, let us consider the electron configuration of the silicon atom:



Si: 1s(2), 2s(2), 2p(6), 3s(2), 3p(2)

Silicon needs four additional electrons to saturate its outermost shell and is therefore disposed to commit to four chemical bonds. In a bulk crystal, the silicon atoms form tetrahedra in a face centered cubic (fcc) configuration, similar to diamond. The outermost silicon atoms lack neighbors and are on the first place less firmly bound to the crystal

lattice and secondly exposed to the etchant. The orientation of this surface with respect to the bulk lattice determines the efficiency with which the etching takes place. A surface cut with the normal along the (100) direction has front atoms that has only two bonds with their backing neighbors and are relatively protruded out of the plane. This ensures ample contact opportunity with the KOH and H_2O atoms that constitute the etchant. Intermediate $Si^{+}_{4}(OH^{-})_4$ – combinations are easily formed, which results in a relatively high etch rate (defined as the distance the surface recedes per unit time; also called etch(ing) speed). Surfaces with a normal in the (111) direction have protruding atoms that have three bonds with atoms in that surface. These front atoms are little exposed to the etchant, which explains the (much) lower etch rate in this direction compared to the (100)-direction (more than a factor of 100).
The indicated etchant parameters (temperature and KOH concentration), dictated by the standardized cleanroom process, are close to optimal for a speedy, yet controlled etching process. In general, a higher temperature speeds up the etch rate as it enhances the chance a chemical reaction takes place. Concerning the concentration, for the etch rate there is an optimum at around 20% (weight) KOH. For the reaction mechanism both OH^{-}-ions and H_2O-molecules are important. A higher concentration of OH^{-} implies a lower H_2O-concentration, which above 20% reduces the efficiency of the etching process.

4.1.5. Wet etching of top structures

4.1.5.1 Etch path

The consecutive phases of the wet-etching process applied to the top structure are shown in Fig. 4.10, where the formation of the springs is concentrated on.

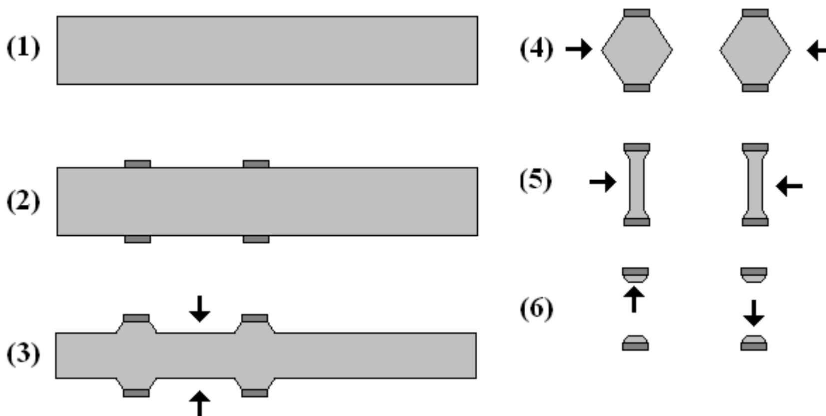


Fig. 4.10 Cross sections of various phases in the etching of the beam springs

Firstly, a silicon wafer (1) is provided with a double-sided silicon nitride mask (2). In (3), the exposed silicon is removed layer by layer by the etchant. Depending on the orientation of the mask on the wafer with respect to the wafer flat, the walls of the etched-out “lowlands” will make a certain angle. Masks edges parallel and perpendicular to the wafer flat (the straight edge of a wafer) will result in an etch angle ζ of 54.74° to

the horizontal plane in case of (100)-oriented wafers[24]. The slopes of these walls are hard-to-etch (111)-surfaces of the crystal. This implies that etch areas that are too narrow will practically stop being etched; in other words, the wafer will not completely etch through in vertical direction. This dictates for example a minimum distance at which neighboring parallel springs should be designed, which depends on the wafer thickness. For a standard 100 mm diameter wafer of 500 µm thick, areas to be etched through should be at least approximately 350 µm wide. In (4), the wafer has just been etched through, from which solid bars result. The corners are exposed most to the etchant and are especially a target for the etching. Continued etching will slimmer the bars (5), which will be etched through completely if the original mask strips are sufficiently narrow. As again, the etching will meet (111)-surfaces, which will almost stop large masked areas from being under-etched. When the bars are etched through, the top and bottom springs are finally separated, after which protracted etching will thinner the springs as indicated in (6). This etching process is illustrated in 3D once more in Fig. 4.11.

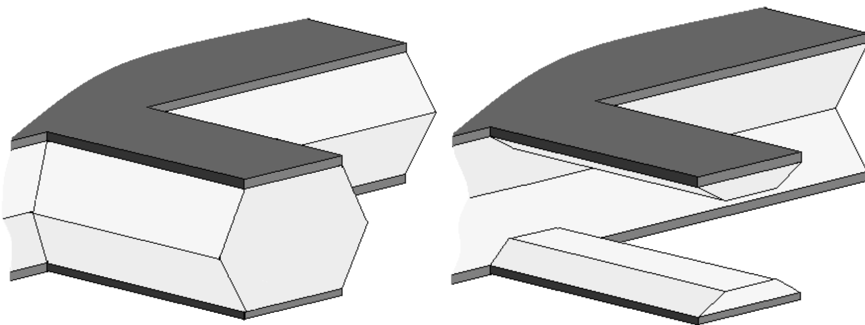


Fig. 4.11 Two phases in the etching process. Dark represents the silicon nitride mask, light represents the silicon substrate. The drawn thickness of the silicon nitride layer has been exaggerated with respect to the silicon.

The solid bar, which in Fig. 4.11 protrudes from the frame, is etched through, forming two springs. The continued etching will dig in the frame silicon as well, but stop at some extent (north east of Figs. 4.11). It is reminded that beams of different widths will consequently also have different thicknesses. As the etching time will be the same, wider beams will lag behind in thinnering and narrower will be ahead.

A detailed cross section of the originally intended spring etching process is provided by Fig. 4.12:

[24] If straight, vertical walls are desired, the mask edges should be designed with an angle of 45° with respect to the flat of (100)-wafers. This direction will etch quickly however, and it should be borne in mind that under-etching underneath the SiRN mask starts immediately after submersion, instead of only after vertical etch-through.

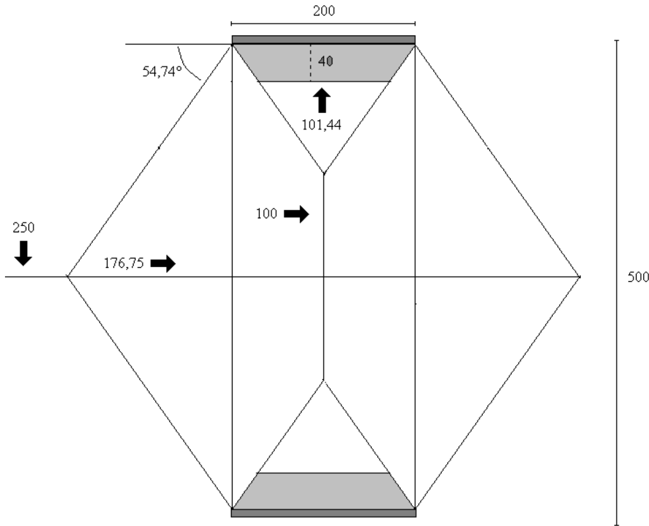


Fig. 4.12 Cross section of spring etching process

Black arrows indicate the "path length" of the etching, which starts with a vertical etching of the wafer, is continued by a slimmering of the resulting bar, after which the top and bottom beams become separated and ultimately thinnered. Typical etch speeds are known from literature [1], which provides an estimate of the total etching time, but some checking structures on the wafer are desirable. Something of this kind had not yet been included in the original process, but a suggestion for this might look like the following (Fig. 4.13):

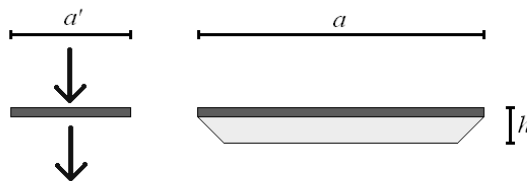


Fig. 4.13 Transparency check for wet-etching

Given that a beam of width *a* is to have a thickness *h*. On some part of the wafer, preferably not too far from the beam, is designed an etch-gauge beam of a narrower width *a*'. This *a*' should be chosen such that as soon as all silicon of this gauge beam is removed and thus becomes transparent, the proper beam having width *a* has attained the thickness *h*. This would provide an easy, initial optical indication of the springs being etched just enough.

4.1.5.2 Etch rate monitoring

The etch process, stopped after 420 minutes in total, had been interrupted a few times (after 150, 240 and 286.5 minutes) in order to determine the progress and the speed of the etch process by a standard microscope. The pictures of Fig. 4.14 are taken after exactly 2½ hours of submersion.

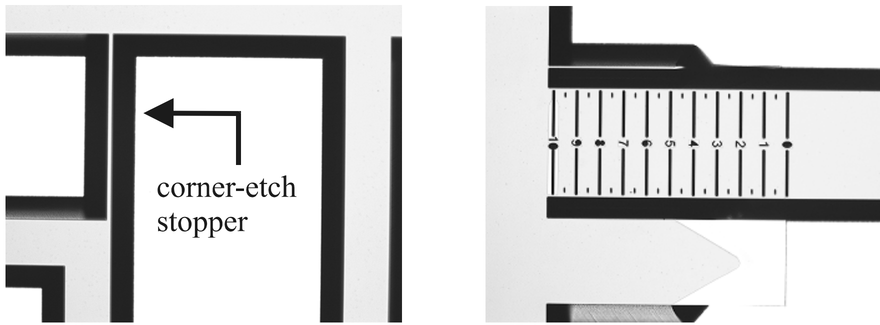


Fig. 4.14 (Left) Top-view after 150 minutes etching of the corner of an L-shaped spring. Grey areas are the silicon nitride mask. White areas are depressed, silicon areas, while the black edges are the slant walls of these areas, that reflect no light back into the camera. The narrow line in the "north west" is a corner-etch stopper.
(Right) Top-view after 150 minutes etching of a sacrificial cantilever structure next to a "ruler" in order to determine the speed of under-etching. The grey areas indicate where there is still silicon immediately underneath the silicon nitride.

Given the angle of the slant walls of the depressed areas of the wafer, the depth etched and with it the vertical etch rate can be determined from the observed width of the dark edges. Fig. 4.14 displays a sacrificial cantilever next to a ruler. It enables to monitor the amount of under-etching in time. The results of the derived etch speeds are summarized in Fig. 4.15:

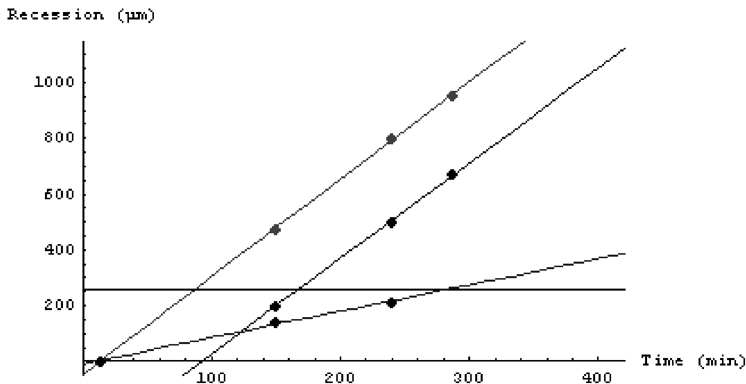


Fig. 4.15 Etching progress for the total etching time of 420 minutes. The two parallel steep lines indicate the recession of the arrow-shaped structure in Fig. 4.14 (right) with respect to the next-lying ruler; the lower one is for the head of the arrow, while the upper one represents its 'shoulder'. The slightly slant line corresponds to vertical etch depth. The horizontal line indicates half of the measured wafer thickness ($½h_w$ = 260 µm) and determines when vertical etch-through is supposed to occur.

From Fig. 4.15 it is seen that arrow head and shoulder recede with similar speeds (fit to 3.43 and 3.50 µm / min. respectively, in the direction parallel to the ruler and the wafer-flat). After the total etch time of 420 minutes it is expected that the arrow (originally

1000 μm long) is completely removed, which is confirmed by inspection of the microscope pictures. For the vertical etch speed there are only two data points (at 150 and 240 minutes) available, as at the third instant of inspection etch-through had (just) already occurred. The line determined by these two points alone would yield the undesirable results of a negative time-offset and a just still solid wafer at 286.5 min. By imposing the physically plausible condition that the time-offset of the vertical etch is identical to the arrow-shoulder (13.5 minutes), etch-through happens just before the inspection moment at 286.5 minutes. This time-offset can be explained by the fact that silicon oxide is formed during the time that the wafers are out of the solution and exposed to air at the moments microscope pictures are taken. This oxide layer had not been removed by a treatment with hydrogen fluoride solution (HF) before re-submersion, as a result of which the KOH-solution first has to remove the silicon oxide before the proper silicon etching can be resumed. The time offset is quite typical for KOH to strip the layer of native silicon oxide. Finally, the vertical etch rate is thus established on 0.94 μm / min. This is slightly slower than 1.1 μm / min, as in [1].

4.1.5.3 Prevention of corner under-etching

It is found desirable to place a few comments on the effectivity of the protection of corners. It was found that the narrow silicon nitride strips could delay but ultimately not prevent the etching of corners, as is seen in Fig. 4.16

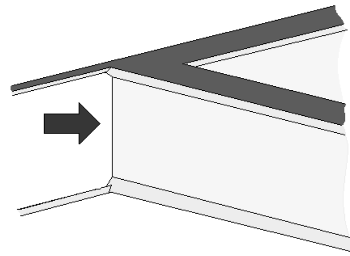
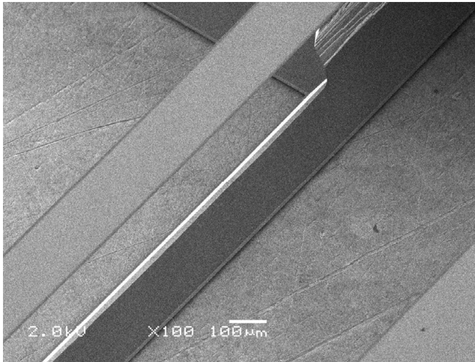


Fig. 4.16 (left) Under-etching at the corner of an L-spring wth an etch protecting strip. Dark areas are open parts of the wafer. Dark grey is silicon nitride with silicon immediately underneath; light grey is floating,transparent silicon nitride.
Fig. 4.17 (right) Under-etching of corners

For springs this is a point of considerable concern, as this can deteriorate or at least influence the desired or expected spring specifications, let alone that extended etching could remove the complete corner. The strips are effective during the phase of vertical etching and during horizontal etching for as far the silicon walls are still convex (are still bulging from underneath the mask; cf. Fig. 4.11). However, at this point the narrow strips will be under-etched quite quickly and as soon as they become separated, the corner of the still solid proper spring becomes exposed (Fig. 4.17). Properly designed L-springs and carefully weighted spring widths and thicknesses and tuned widths of the corner protection strips are still viable, but some under-etching of the corner always has to be taken into account. Sacrificial structures, that are designed to be completely etched away just when the desired spring thickness is attained, are certainly safer for L-spring corners, although their tree-like design is all but trivial and should be designed with great care as well in order to arrive at the desired specifications.

4.1.5.4 Results of the etching process

Let us finally take a closer look at the results of the etching process:



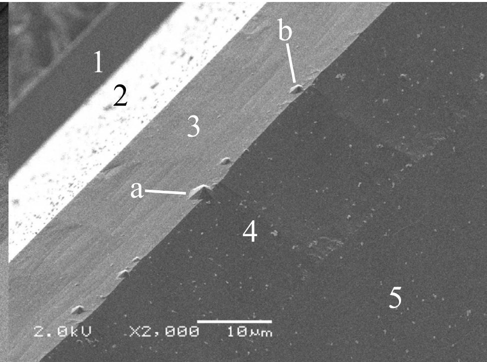


Fig. 4.18 Two springs *Fig. 4.19 Zoom on the lower spring of Fig. 4.18*

Fig. 4.18 is a Scanning Electron Microscope (SEM) picture of a pair of springs, holding a rectangular silicon plate (just inside the scope of this picture, in the north-east). The great majority of the springs, like the ones on this picture, came out of the process unscathed and well separated. The etching process was terminated after 420 minutes, as by inspection all springs were visibly separated, but little could be said about the thinnering of the springs. It turned out that, as can be deduced from Fig. 4.19, the thinnering of the springs had only just commenced. Let us look to the cross section of the spring in more detail, as illustrated in Fig. 4.20:

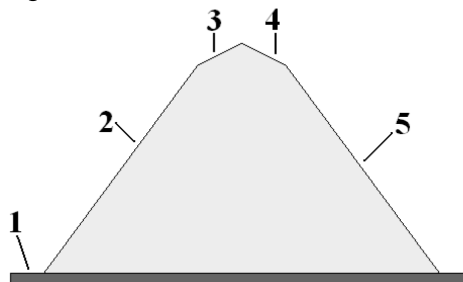
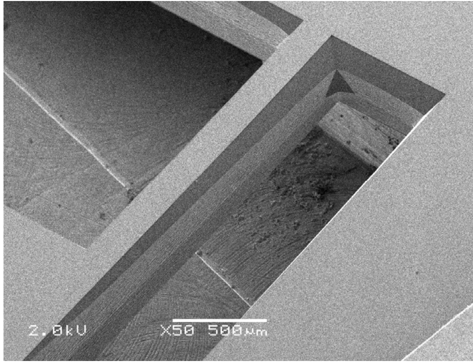


Fig. 4.20 Cross section of a spring; numbers correspond with areas in Fig. 4.19. Light grey is silicon, dark is silicon nitride.

In the order of ascending numbers we first encounter that there is a small recession / under-etching of the silicon with respect to the silicon nitride mask, creating a terrace of about 10 µm width (1). This is consistent with the etch rate of a (111)-plane compared to observed with (100)-plane etching. The western slope of the spring (2), looks narrow in Fig. 4.19 because the 'camera' is almost aligned with this slope. (3) and (4) are the top surface of the spring, which has just begun flattening. In Fig. 4.19 the eastern slope (5) is also still visible. Characteristic of this etching process is the abundance of tiny, rectangular and octagonal cones or pyramids, (a) and (b) in Fig. 4.19. In [5], a few, not mutually excluding, hypotheses are put forward for this. One is that tiny H_2-bubbles at the surface act like micro-masks, around which the etching continues that leave (111)-

planes behind. Another one is that they are in fact SiO_2-precipitates. The number density and the size of them increase with decreasing etch temperature and KOH-concentration.

A small area on one of the wafers lagged behind in etching, for unclear reasons. This provides however a glance at various etching stages in the process. Shown here is a case where the beam springs are still connected together as a solid bar (Fig. 4.21) and a case where they are just about to be separated (Fig. 4.22).



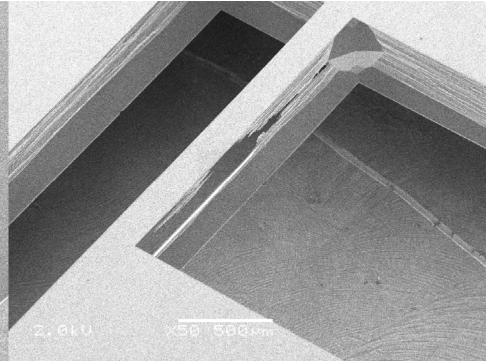


Fig. 4.21 Unseparated springs

Fig. 4.22 Almost separated springs

4.1.6 Bottom mask design

The bottom wafer embeds the electrodes that actuate the floating top-electrodes. The electrodes are matched to the shape of the top structures, see Fig. 4.23.

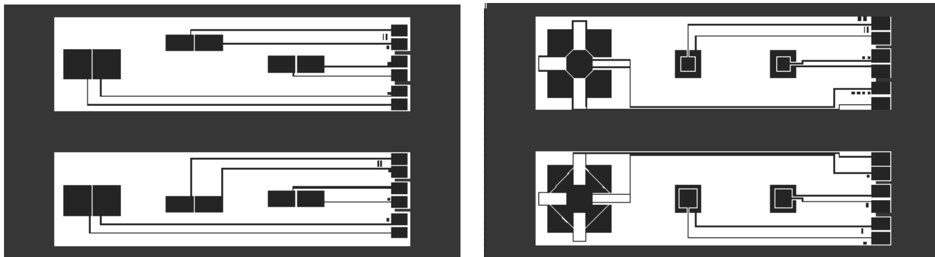


Fig. 4.23 Mask design for bottom wafer. The shown electrode chips match the top structures as in Fig. 4.3. A variation to the left design that has been realized is with a ring electrode surrounding a center electrode. Both this latter configuration and the depicted one were fabricated for p = 0.2 and p = 0.5 (A_1:A_4 = 1:4 and 1:1), see also subsection 4.1.1.

White areas are regions that are lowered with respect to the surroundings that support the top frame. These depressions define the capacitive gap. In the lower regions the electrodes (dark) and the on-chip paths are located. On the right are the contact (bonding) pads that are reachable through the contact holes as remarked on the basis of Fig. 4.3. The on-chip paths are routed such that parallel capacitance to the top structure is minimal and negligible.

4.1.7 Fabrication process bottom structures

The first batch of bottom electrodes was fabricated along the steps illustrated in Fig. 4.24.

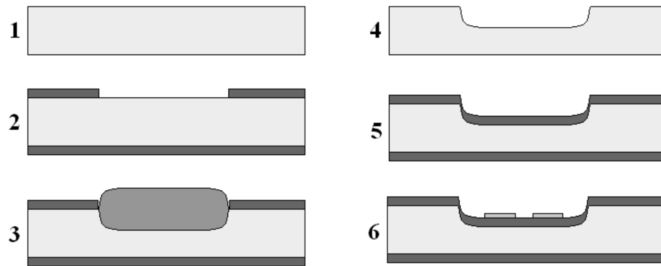


Fig. 4.24 Process steps silicon bottom wafer

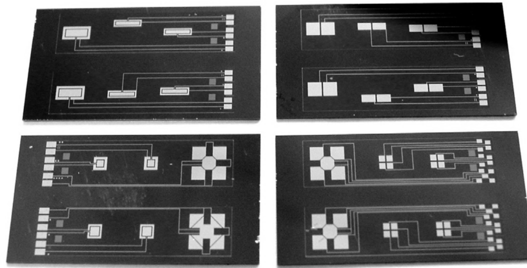


Fig. 4.25 Silicon bottom wafer electrodes

A double-sided polished (100) silicon p-type wafer (1) is provided with 800 nm silicon nitride on both sides by LPCVD. On one side, the parts of the silicon nitride are removed (photolithography and plasma etching) where the depressions are going to be made (2). The bare silicon is oxidized (dry) at 800°C, which will form a layer of SiO_2 that partly digs in the silicon (3). Treatments with buffered HF and HF 50% will consecutively remove the silicon oxide and the silicon nitride, after which just silicon remains with lowered areas (4). In (5) silicon oxide was sputtered on the top side. Finally, the electrodes (aluminum, 100 nm) are sputtered and patterned (6) by photo lithography and wet-etching. The resulting samples are shown in Fig. 4.25.
The disadvantage of this design however is that the usual AC-readout by the impedance analyzer runs into difficulties as the silicon oxide isolation layer permits a capacitive coupling from the electrodes to the silicon substrate, thus shortcutting the electrodes in AC-mode. However, even proper DC-experiments turned out to be not feasible, as the electrodes proved to be genuinely cut short. This is probably due to the aluminum sputtering damaging the isolating silicon oxide.
For the second batch of bottom electrodes, borosilicate glass (Pyrex) substrates were chosen. Admittedly the two constituting parts of the ultimate device are then not of identical material, but Pyrex is developed to be compatible for wafer bonding to silicon, for which similar thermal expansion coefficients is a demand. Even then, as the bottom and top parts are clamped, not bonded, there is still some freedom left for differences in expansion to occur. The idea is that because of the insulating substrate, no coupling of significance can take place as in the previously described case of the silicon bottom wafer. The process is summarized in illustration in Fig. 4.26.

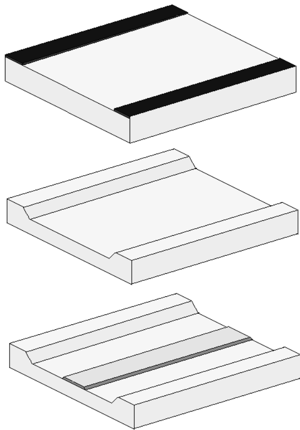


Fig. 4.26: Process steps pyrex wafer

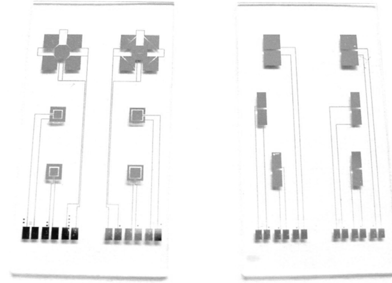


Fig. 4.27: Pyrex bottom wafer electrodes (picture)

At first, photo resist is patterned on Pyrex (top). The bare Pyrex is etched by HF, after which the photo resist is stripped. This defines the capacitive gap (middle, about 2 µm). It should be noted that considerable under-etching takes place, which is however not harmful in our case. At last, the electrodes (aluminum, 100 nm) are applied by sputtering and patterning (bottom). Some resulting makes are shown in Fig. 4.27.

4.2 TEST TOP STRUCTURE

This section describes a preliminary test of a top structure in order to verify the capacitance-voltage behavior. By simultaneously measuring displacement and capacitance variation as a function of applied dc voltage, parameters like stiffness, parasitic capacitance and capacitive gap can be retrieved.

4.2.1 Experiment setup, phase shift interferometry

For the detection of the bridge displacement Phase Shift Interferometry (PSI) had been invoked. A monochromatic light source (green, λ = 455 nm) is directed vertically downward on a slightly tilted surface (the sample). The reflected light will interfere with the incident light. This will be alternatingly constructive and destructive in the direction of the slope, thus causing a pattern of light and dark stripes, or *fringes*. Vertical movement of this surface will shift this pattern horizontally. This shift is recorded with a CCD – camera. An advantage of PSI is that no reference surface is required. By applying it at several points, the uniformity of the movement can be determined. More details on PSI can be found in [7]. The structure had been provided with 100 nm aluminum on the bottom side only. On the top side there were copper elements fixed which were brought in contact with the aluminum layer with silver paint. This enabled contacting the bridge with an electrical probe from above. A copper strip on a home-made PCB formed the capacitive counter-electrode, which matched the top electrodes completely. A firmly baked photo-resist layer was used as spacer, cf. Fig. 4.28. The actuation and read-out was

by the HP4194A impedance analyzer mentioned in Chapter 3. With this set-up, capacitance change and displacement could be measured simultaneously.

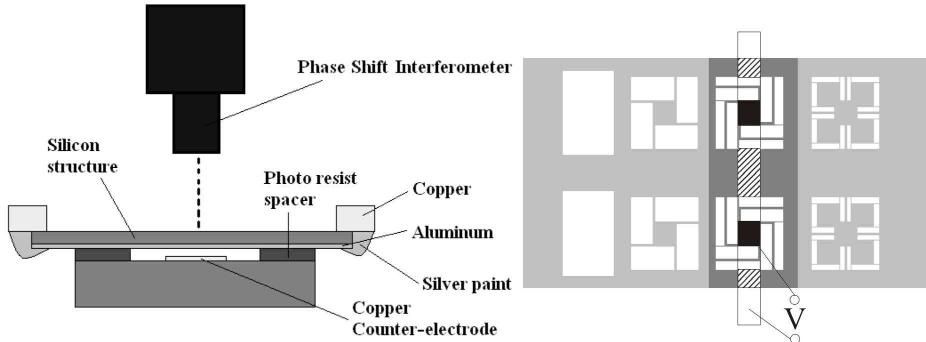


Fig. 4.28 (Left) Phase shift interferometry set-up; (Right) Actuation pattern. The vertical bar is the bottom electrode running underneath two structures simultaneously. The MEMS capacitive areas are marked black. The cross-hatched areas indicate parallel capacitance. A voltage V is applied between the top and bottom electrodes.

4.2.2 Results

Because the springs were fabricated thicker and thus much stiffer than originally intended, only the structure with the long (5 mm) L-shaped springs (see Fig. 4.9) was found sufficiently sensitive for the current test and even then only high voltages sufficed to detect displacement.
The preliminary measurements (at room temperature) of the top structures consisted of capacitance measurement as a function of voltage simultaneous with displacement measurement. The chosen measurement set up allows the closure of all mathematical relations and provides figures for the spring constant, the parasitic capacitance and the gap distance. This will be discussed shortly. The devices were actuated and read-out together: a combined capacitance was measured. Compared to a single device, the combined electrostatic force is twice as large. However, also the stiffnesses of the individual devices add up. Hence the displacement of either structure is the same as in the case that only a single structure would be actuated. The voltage is swept from –40 V to +40 V. Figure 4.29 shows a curve from the $C(V)$ analysis of two sensors of the type in Figure 6c simultaneously. A parabola has been fit through the measurement points.

The displacement has been determined by green light (455 nm) phase interferometry. The fringe shift provides an estimation of the vertical movement of the test structure upon actuation. The unactuated measured capacitance C_{m0} is the sum of the test structure capacitance C_t and the parasitic capacitance C_p (assumed parallel to C_t). These are related to the unactuated gap z_0:

$$C_{m0} - C_p = \varepsilon_0 A / z_0 , \qquad (4.8)$$

with A the capacitive area of the test structure and $\varepsilon_0 = 8.85 \cdot 10^{-12}$ F/m the permittivity of the vacuum. The second relation is the force balance (see also Eq. (2.7)):

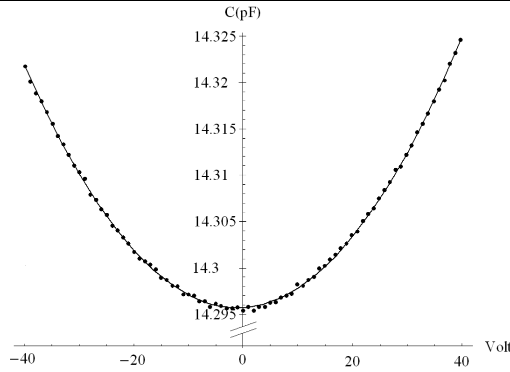


Figure 4.29. Capacitance measured versus voltage together with a quadratic fit
A built-in voltage of a magnitude of about 1 V has moved the curve a little to the left. This we expect to be due for the greatest part to the difference in work function of the electrode materials (4.20 eV for Al-(100) and 5.10 eV for Cu-(100) [8]).

$$\frac{\varepsilon_0 A V^2}{2(z_0 - \Delta z)^2} = k\Delta z \,. \tag{4.9}$$

Δz (>0) is the change in gap height when a voltage V is applied. For this, the interferometry was invoked. For 40 volt this gave 57 nm. Thirdly, for small displacements the force balance (4.9) can be simplified to yield a parabolic relation between capacitance and voltage. The quadratic coefficient α relates capacitance, gap height and spring constant by

$$\alpha = \frac{(C_{m0} - C_p)^2}{2kz_0^2}\,, \tag{4.10}$$

and is taken from the fit. Combining Eqs. (4.8), (4.9) and (4.10) determines the spring constant on 6.9 kN/m. This is considerably larger than the expected value of 2.7 kN/m. This is discussed in the next subsection 4.2.3.
The effective gap height comes out at 12 µm, which is also a lot higher than expected (5 µm, the thickness of the photo resist spacers). For this discrepancy no conclusive explanation has been found. Finally, the parasitic capacitance C_p came out on 8.4 pF. From the design (see also Fig. 28 (right)) it was expected that the parallel capacitance C_p would be 64% of the total capacitance. The copper electrode also runs underneath parts of the rigid frame, which also has aluminum. From a total capacitance of 14.3 pF, C_p would then amount to 9.1 pF. This shows that the experimentally derived value is conform expectation. Using the derived gap and the empiric value for the spring constant, the pull-in voltage (Eq. 2.14) can be calculated. Recapitulating:

$$V_{pull-in} = \sqrt{\frac{8kz_0^3}{27\varepsilon_0 A}}\,, \tag{4.11}$$

Inserting the quantities yields $V_{\text{pull-in}}$ = 225 V. Connecting the device to a high-voltage dc source enabled optical observation of pull-in, which occurred close to 300 V.

4.2.3 Spring constant anomaly

The spring constant k of the two test structures together is calculated by (see subsection 4.1.2)

$$k = \frac{16NEah^3}{3(2l)^3}\ . \qquad (4.12)$$

Inspection (Figs. 4.19 and 4.20) motivated to assume the cross-section of the beams to be triangular. N is the number of beam spring *pairs* ($N = 8$ for the two structures together).
E is calculated with Eq. 4.5. In the relevant direction, Young's modulus Y is 168 GPa and Poisson's ratio $\nu = 0.065$ [9]. In this case, Poisson's ratio has little influence. For the length of the springs, we have $L = 2l = 2 \times 5$ *mm*. The width of the springs was designed to be 200 μm. In subsection 4.1.5.4 however, it is shown that underetching of the silicon in the (111) direction reduces this by about 10 μm from each side. A width $a = 180$ μm is therefore taken. Finally, assuming a slope angle of 54.74° implies a thickness $h = 122$ μm for a triangular cross section. Inserting these parameters gives $k = 2.7 \cdot 10^3$ N/m. This is a factor of two-and-a-half too low compared to empiricism. However, after taking a close look at the L-spring samples (Fig. 4.30), it was noticed that the springs still had vertical interconnections in the corners of silicon columns (Fig. 4.31). This entails that the effective length of the spring is decreased by 20%, which 'upgrades' the spring constant by a factor of $(0.8)^{-3} \sim 2$ to $5.4 \cdot 10^3$ N/m, which matches the experimentally determined spring constant to roughly 20%.

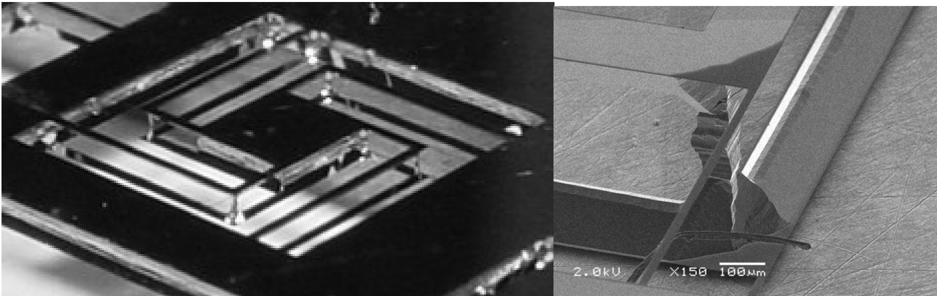


Fig. 4.30 L-spring structure, here with aluminum. *Fig 4.31 Interconnection between springs in L-shaped structure (SEM).*

4.3 MECHANICAL SAMPLE CHARACTERIZATION

In the previous section the stiffness of the springs has been derived with an indirect, quasistatic experiment. In this section, the stiffness will be obtained by the more direct and well-known method of resonance measurements. The experiments presented in this section arguably extend somewhat beyond this primary goal by exploring the dependence of resonance parameters on bias voltage and ambient pressure, though the former dependence is still valuable for more accurately determining the intrinsic spring constant.

4.3.1 Sample description

In this sub-section, the samples used are described. This is done separately for *A) Top structure* and *B) Bottom structure.*

A) Top structure:

The resonance measurements are preferably done on samples with a simple configuration and simple springs. More complicated structures, like the L-shaped structure in Fig. 4.30, may introduce hard-to-evaluate contributions (vibration modes) to the overall signal. The sample shown in Figure 4.32 meats these requirements.

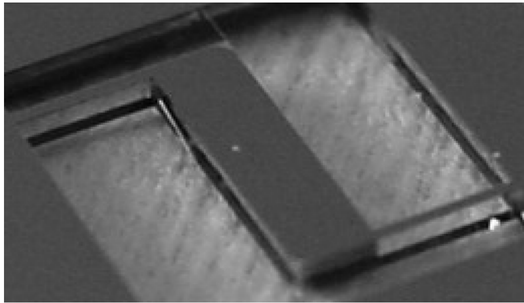


Fig. 4.32 (Left) Basic top structure, here still without aluminum electrode, which had been applied on one side only in the resonance measurements presented in this section. The block area measures 4.4 × 1.2 mm². The two pairs of beam springs are clearly visible. In the top, a very narrow nitride beam, used for corner etch delay, can be discerned with some difficulty.

For the inertial (dynamic) resonating mass, the gravitational mass of the suspended structure is calculated by its geometric volume $\Omega = 4.4 \times 1.2 \times 0.52$ mm³ and the mass density of silicon ρ_{Si} =2330 kg / m^3. Hence $m = \Omega \times \rho_{\mathrm{Si}} = 6.40$ mg. The mass of the springs, the thin silicon nitride layers and the aluminum layer have been ignored, as well as the fact that because of the etching process the vertical walls of the massive block are not straight but rather concave. This is drawn in Fig. 4.4 and can be observed in Figs. 4.22., 4.30 and 4.39. The two neglected contributions may partly cancel: the former is additive, the latter is subtractive. The stiffness of the suspension is again calculated by Eq. (4.12), see also subsections 4.1.2 and 4.2.3. As before, triangular cross-sectional springs are assumed. The number of pairs $N = 2$, while the lower width a is taken to be 180 µm, as under-etching is estimated to have eroded approximately 10 µm from the designed 200 µm on both sides, see figures (4.19) and (4.20). The thickness $h = 127.3$ µm is derived from this width and the etching angle ζ of 54.74° ($h = ½a \tan \zeta$, see sub-subsection 4.1.5.1, Fig. 4.12). This results in a spring constant $k = 10.46$ kN/m. The contributions of the silicon nitride layer and the aluminum layer are ignored. The former layer thickens the springs with ~ 0.8 µm and would therefore increase the stiffness by a percent or two. The four nitride strips for corner under-etch delay combine to weigh in about 1 mN/m. Being a factor of 10^7 lower compared to the (main) springs, they play absolutely no role.

B) Bottom structure:

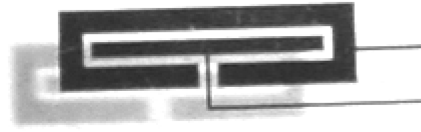


Fig. 4.33 Bottom electrodes structure used in this experiment series (picture). Lying on top of a Pyrex substrate (not visible), their shadows are cast on the underlying white surface.

For the given configuration, the effective (series) capacitive area (see subsection 4.1.1) is calculated $A_s = 0.71$ mm², and has a partial area coefficient of $p = 0.2$. The electrodes are in an area that is lowered by 1 µm with respect to the bottom sample edges supporting the frame edges of the top structure. However, over the length scales of the frames (3.6 cm), wafers can vary in thickness and be curved. These factors introduce an uncertainty in the gap of a maximum of a few µm, which makes the gap and therefore the capacitance an a priori unknown parameter.

4.3.2 Single peak characteristics

The locus plot of a narrow-band frequency sweep of the admittance of the sample configuration described above is presented in Fig. 4.34. The original data consisted of absolute impedance $|Z|$ and phase θ, from which the conductance G and the susceptance B have been derived by $G = |Z|^{-1} \cos\theta$ and $B = -|Z|^{-1} \sin\theta$.

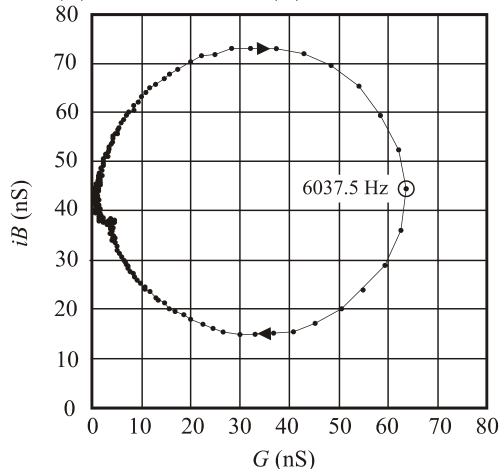


Fig. 4.34 Locus plot of admittance at 40 V bias voltage (5800 Hz ≤ f ≤ 6500 Hz). Conductance G and susceptance B are in units of nanoSiemens (nS). Arrows indicate ascending frequency. The encircled point is at f = 6037.5Hz and is very close to series resonance frequency. The quality factor is not sufficient for intersection with the B = 0 axis. We also observe that the locus plot is not exactly circular but is flattened by about 10% with respect to the horizontal direction. This is indicative for a damping that is not constant with frequency. The small bump in the 'west' part of the circle is probably due to a co-resonating structure, see below.

In Fig. 4.35 the conductance versus the frequency is shown. The black dots are fitted according to the theory described in section 2.4, where a small vertical offset in the conductance function has been allowed in order to reach optimal fitting of the peak. This offset could be caused by a small (additional) phase shift in the primarily capacitive circuit. The small bump to the right of the main peak is probably to be ascribed to a co-resonance. It can be seen from for example Fig. 4.3 that the frame contains another structure that is expected to have a similar mass and a similar spring constant and consequently a similar resonance frequency. The grey data around this feature in the graph have been omitted from the fitting procedure. They are considered to be too much influenced by the co-resonance at around $f = 6250$ Hz.

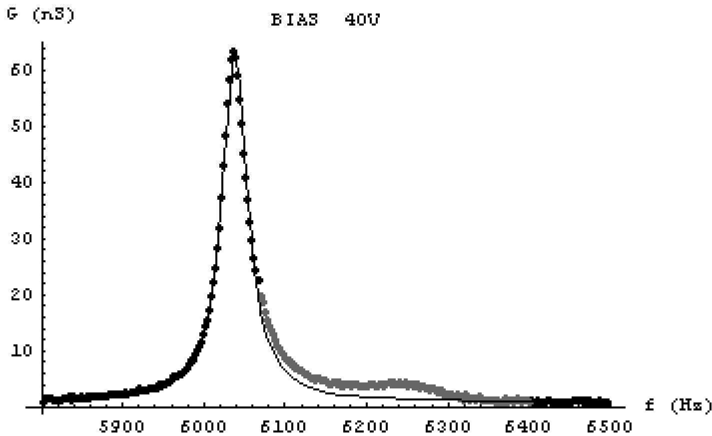


Fig. 4.35 Conductance (nanosiemens) vs. frequency (Hz) around the fundamental resonance frequency for Vbias = 40V. The small gray dots between 6070 Hz and 6400 Hz have been omitted for the fitting (thin line). The bump that these data trace out is probably due to a co-resonating structure.

From the fit a series resonance frequency of f_s = 6038.5 (±1) Hz = $\omega_s/2\pi$ has been deduced and peak width of Δf = 38.6 Hz, combining to a quality factor of Q = 156.5. Assuming a mass of 6.40 mg, this frequency would imply a *dynamic* (or electromechanical) stiffness K_{EM} = 9213 N/m. This is a bit lower than was calculated in subsection 4.3.1 (10.46 kN/m), but still the agreement is quite good. In fact, the intrinsic (mechanical) stiffness k should be somewhat higher because of the coupling between the electric and mechanical domain, see section 2.1. From the peak characteristics, the parallel lumped elements R, L and C can be deduced, which result in R = 16.2 MΩ, L = 66.7 kH, C = 10.4 fF. Though not as dramatic as stated in section 2.4, still it's especially the inductance assuming quite an exotic value.

4.3.3 Bias dependent resonance

Because the capacitive gap is a priori unknown, the measurement is to be repeated for different bias voltages, see Fig. 4.36. This is done for voltages from 20 to 40 V. We see that the height of the conductance peak gets lower when the dc bias voltage decreases. At zero voltage, no resonance at all can be seen. The reason is that the transduction factor Γ would then vanish, cf. Eqs. (2.103) and (2.119). We see that for higher voltage, the

resonance frequency decreases. It is derived that this is conform what can be expected. According to expectation, higher bias voltages decrease the dynamic resonance frequency. By combining

$$K_{EM} = k - \frac{\varepsilon_0 A}{z^3} V^2 \tag{4.13}$$

(see Eqs. (2.18) and (2.19)) with

$$C = \alpha V^2 + C_0, \qquad \alpha = \frac{\varepsilon_0^2 A^2}{2kz_0^4}, \qquad C_0 = \frac{\varepsilon_0 A}{z_0}, \tag{4.14}$$

(see section 2.1) the dependence of the gap z on the bias voltage V is taken into account.

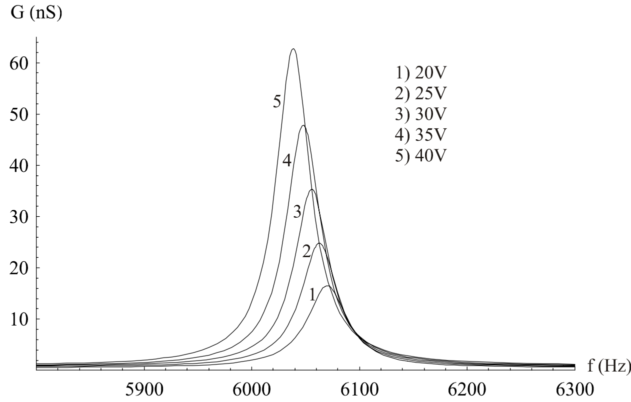


Fig. 4.36 Fits of conductance function for data at 20 V (lowest peak) to 40V (highest peak) bias.

For easier reading, let us introduce the parameter Λ, defined as

$$\Lambda \equiv \frac{\varepsilon_0 A}{kz_0^3}, \qquad [\Lambda] = 1/\mathrm{V}^2, \tag{4.15}$$

so that

$$\omega_s^2 = \frac{K_{EM}}{m} = \frac{k}{m}(1 - \Lambda V^2 - \frac{3}{2}\Lambda^2 V^4), \tag{4.16}$$

where terms of order in Λ higher than 2 have been neglected. ω_s^2 has been fit with $c_1 + c_2V^2 + c_3(V^2)^2$ in the variable V^2 in Fig. 4.37.

Comparison of the fit with Eq. (4.16) yields and over-determined problem, as three fitted parameters ($c_{1,2,3}$) are competing for two parameters in Eq. (4.16) (k/m, Λ). The intrinsic mechanical (damped) resonance angular frequency $\omega_0^2 = k/m = c_1$ is determined on 1459 krad²/s². Inserting the mass of 6.40 mg in Eq. (4.16) yields a spring constant of k = 9335 N/m. This can be said to be in quite good agreement with the calculated value of 10.44 kN/m. The latter value could be tuned down a bit if one realizes that the assumption of a triangular cross section leads to an over-estimate of the stiffness, because close inspection of the springs shows that the top of the triangle is already a bit flattened.

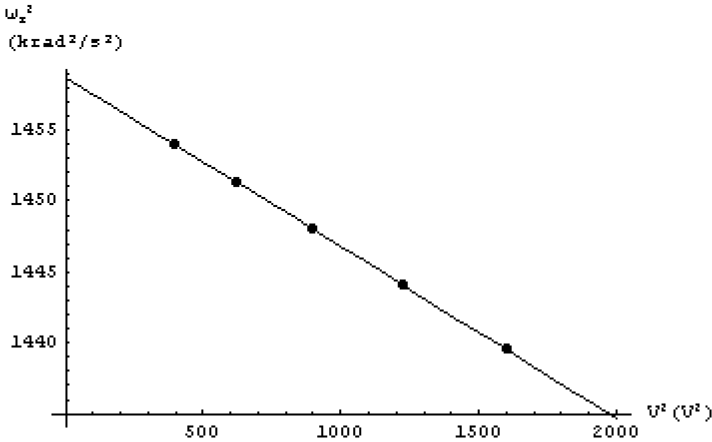


Fig. 4.37 Fit of dependence of series resonance on bias voltage. The points lie on a straight line to very good approximation, though the fit allowed for a curvature, which turned out upwards, as expected.

This misestimated spring thickness weighs in cubical, so small errors manifest easily in the result. Furthermore we observe that a bias of 40 V (cf. Fig. 4.36) alleviates the intrinsic stiffness by 122 N/m, which is a reduction by 1.3%. The main objective of the resonance experiments has thus been achieved: Determination of the spring stiffness, which agrees satisfactorily with prediction. Calculation of the Λ–parameter additionally yields the capacitive gap in these series and can be done from the fit either by $-c_2/c_1$ or $2c_3/3c_2$, which must lead to incompatible values. Of these two the former was chosen, as the Λ^2 term in (4.16), associated with c_3, contributes only a few ppm, which would require $\omega_s(V)$ to be determined with considerably higher accuracy. Given the fitted spring constant and $\Lambda = c_2/c_1 = 8.11\cdot10^{-6}\ \mathrm{V}^{-2}$, we obtain a capacitive gap $z_0 = 4.4\ \mu\mathrm{m}$, considerably greater than designed (1 μm). For the record, $2c_3/3c_2$ gives $\Lambda = 5.7\cdot10^{-6}\ \mathrm{V}^2$, positive value, which entails upward curvature as expected and is even quite close to the 'correct' Λ. However, this latter result should be considered fortuitous rather than enforced by experimental accuracy. The variation of the gap with voltage is so small for these structures that this effect is hardly noticeable, in spite of its cubical influence.
With the calculated parameters, the transduction factor (cf. Eq. (2.103)), written in a form that takes a bias-dependent gap into account,

$$\Gamma = kz_0(\Lambda V + \Lambda V^3)\,, \tag{4.17}$$

would for $V = 40$ V come out on $\Gamma = 13.4\ \mu$C/m. Using the values for L and C from Fig. 4.36 , this would predict a mass $m = 11.9$ mg and a dynamic stiffness $K_{EM} = 17.2$ kN/m, both over-predicted by the same factor 1.85. The results would agree for a transduction factor Γ that is a factor $\sqrt{1.86} = 1.36$ smaller, or if the maximal conductance would have been higher by a factor of 1.86. The complete results for the bias dependence of peak characteristics and its derived parameters are summarized in Table (4.1).

V (V)	f_s (Hz)	χ (nS)	Δf (Hz)	R (MΩ)	L (kH)	C (fF)	Q	Γ (μC/m)	M (mg)	K (N/m)	D (g/s)
20	6070.1	16.04	45.0	62.3	221	3.11	135.0	6.63	9.69	14.10	2.738
25	6062.8	24.18	41.3	41.4	160	4.32	146.9	8.30	10.99	15.94	2.849
30	6055.7	34.60	39.4	28.9	117	5.91	153.8	9.98	11.64	16.85	2.879
35	6047.8	47.06	39.1	21.2	86.5	8.01	154.6	11.7	11.79	17.02	2.896
40	6038.5	61.86	38.6	16.2	66.7	10.4	156.5	13.4	11.95	17.20	2.896

Table 4.1 Bias dependence of resonant parameters and derived quantities

Concerning the mass and the dynamic stiffness as derived from the motional inductance L and C respectively, it is observed that they approach m = 6.40 mg (a priori calculated) and k = 9213 N/m (fitted) increasingly better for declining bias voltages. For V = 20 V, the factor with which they deviate from these values is still the same for both parameters, but is smaller for decreasing voltages. This would entail either an incorrectly described dependence of the transduction factor Γ on the bias voltage, or, damping (affecting $\chi \equiv G_{\max} = 1/R$) has been poorly taken account of. A complete discussion of this latter topic would be far beyond the topics of this thesis, though a short digression is given in the next subsection. AI final issue that is not clearly understood is the increase of the quality factor with increasing voltages. Higher voltages lead to a lower stiffness and hence quality factor, at least for a constant damping factor. In the table this factor is indeed quite constant, but if it is downscaled with the same factor as M and K 'should' be, it is larger at lower voltages. The reason behind this supposed bias dependence of the damping factor is not understood.
Additional to these resonance measurements, a few $C(V)$ measurements have been performed. As these devices are very insensitive, capacitance changes are very small, but measurable.

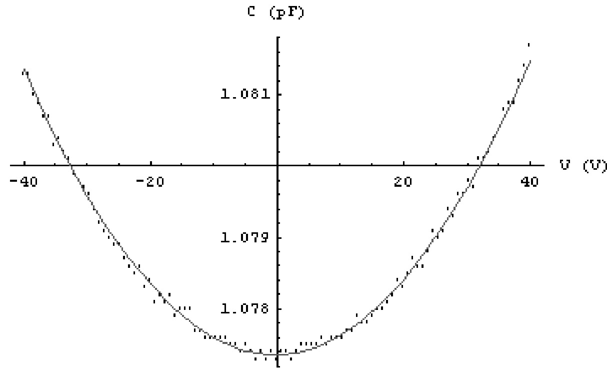


Fig. 4.38 C(V) curve, read out at 1 MHz.

If we allow for a small parasitic capacitance, inserting in the system of equations

$$C_{M0} = \frac{\varepsilon_0 A_s}{z_0}, \qquad C_0 = C_{M0} + C_p, \qquad \alpha = \frac{C_{M0}^4}{2k(\varepsilon_0 A_s)^2} \tag{4.18}$$

the fit parameters C_0 = 1.077 pF and α = 2.54 aF/V², yields a gap z_0 = 5.4 μm; somewhat larger than but not grossly inconsistent with what had been deduced from the resonance measurements. It has to be noted that the parasitic capacitance here is

–0.09 pF, small but, more importantly, negative. For this curvature to be possible, the capacitance should have been at least 10% larger. A possible explanation for this disparity between curvature and capacitance is non-parallel electrodes (see subsection 2.5.1) or, more general, a non-uniform gap.
At last, a 1 kHz-1MHz frequency sweep of the impedance had been done. For a phase θ = –89.2°, the impedance is dominatedly capacitive, which was calculated by $C_M = (2 \pi |Z| f)^{-1} = 1.078$ pF ($f = 1$ MHz, $|Z(f = 1 \text{ MHz})| = 148$ kΩ), consistent with the $C(V)$ curve.

4.3.4 Phase error

As stated in the beginning of this section, the admittance characteristics have been derived from impedance ($|Z|$, θ) data. Off-resonance, any small error in the phase will for a capacitance-dominated impedance ($\theta \approx -90°$) introduce huge relative errors in the conductance $G = |Z|^{-1} \cos \theta$, which can even affect its sign. Only on-peak for resonances with a very large phase shift can the conductance be derived with acceptably small relative uncertainty. The maximum phases for the resonance varied from $\theta = -69°$ for V = 20 V to $\theta = -21°$ for V = 40 V. In the former case, uncertainty of a few degrees introduces 10-20% error in the peak height, while in the latter case errors are in the order of 1%. This uncertainty in the peak height has noticeable consequences for the obtained values for R, and for the lumped-elements derived mass and stiffness. Furthermore, also the lower-parts of the peak are described with increasing uncertainty, which may affect the peak width and therefore quality factor and again derived mass and stiffness. Proper experiments on resonance should certainly be performed without a translation between impedance and admittance; either resonance properties should somehow be derived from impedance characteristics (for the series resonance frequency the maximum in the phase is a reasonable alternative; at $V = 40$ V, $\theta(\omega) = \theta_{max}$ for $\omega \sim 6050$ Hz, 0.2% higher than fitted from G), or admittance should be measured immediately. However, it is expected that this phase uncertainty has negligible effect on the frequency of the maximum conductance (series resonance), which was used to determine dynamic and intrinsic stiffness and capacitive gap.

4.3.5 Damping

The experimental series have been performed in air of a constant pressure 0.9 mbar. Though the probe station allowed for pressures below 10^{-4} mbar after protracted pumping, technical issues prevented this vacuum to be maintained with pumps switched off, which had to be because of otherwise introducing too much noise in the frequency measurements. A few possible damping sources will be mentioned.
It can be shown that a small damping force linear to the velocity has the effect of not only decreasing the amplitude of resonance (which would actually become infinite if the driving force operates at free resonance frequency), but also decreasing the frequency of free resonance, which may lead to the interpretation of the inertial (dynamic, resonating) mass M_{dyn} becoming larger than the gravitational mass m by $M_{dyn} = m(1 + Q^{-2}/4)$. The origins of damping linear in velocity are manifold, for example (small) losses via the springs, but more notably the viscous forces in the surrounding fluid (air). These viscous forces have not been studied further in this context, but are expected to dominate velocity-square dependent drag forces that are present when an object moves with high

velocity through a fluid and creates a pressure build-up in the front, a vacuum behind and turbulences. To get a feel for the expected amplitudes and velocities, let us recapitulate the amplitude of a damped oscillator (denoted '*Z'* rather than the common '*A*', to distinguish from area) under a harmonic driving force $F = F_0 e^{i\omega t}$:

$$Z^2 = \frac{F_0^2}{\left(m(\omega_0^2 - \omega^2) + D^2/4m\right)^2 + D^2\omega^2}, \tag{4.19}$$

where m is the intrinsic mass, $\omega_0^2 = k/m$ is the damped resonance frequency and D the damping coefficient. For $\omega = \omega_0$ and $Q^2 = mk/D^2 >> 1$ (small damping), Eq. (4.19) reduces to very simply

$$Z = \frac{F_0}{D\omega_0}. \tag{4.20}$$

Assuming that D in Table (4.1) is overestimated by the same factor as K and M, this would entail D = 1.56 g/s.
The voltage amplitude has been 0.5V, which for a capacitive surface of 0.71 mm² and a gap of 4.4 µm would deliver a force amplitude F_0 = 41.3 nN. At 40 V bias the motional amplitude would then become A = 0.7 nm (note that applying a static bias voltage of 0.5 V would lower the plate by only 5 pm, about one percent of the diameter of an aluminum atom in the top electrode and about a factor of 140 smaller than the amplitude).
Equating maximal kinetic energy with maximal mechanic energy by $\frac{1}{2}mv_{\max}^2 = \frac{1}{2}kZ^2$, we find $v_{\max} = Z\omega_0 = F_0/D$ = 30 µm/s. At these very low speeds viscous forces are expected to dominate velocity-square drag. The maximal Reynolds number $Re = \rho\, v_{max}\, L/\eta$ with mass density 1.2 g/m³ (air at pressure 0.9 mbar), speed 30 µm/s, length scale (for which the structure diameter was taken) L = 5 mm and air viscosity (room temperature) η = 18 µPa·s is Re = 10^{-5}, indicative for the viscous regime. This becomes explicitly apparent if the forces $F_{visc} = Dv_{\max}$ and $F_{drag} = \frac{1}{2}\rho\, v_{max}^2\, C_d\, A$ (C_d = drag coefficient = 1.28 for a flat plate; A = area) are compared. The former closely equals the maximal driving force, while the latter is many orders of magnitude smaller.

Acoustic damping starts affecting the resonance if the acoustic wave length $\lambda_{ac} = v_{sound}/f_s$ = 5.6 cm is smaller than the dimensions of the resonating structure, which is obviously not the case here.

At last, squeezed film damping can be of importance. In the 'worst case', the volume of gas trapped underneath the resonating structure is adiabatically compressed with no outflow of heat or gas molecules. From equilibrium position to down-state takes $1/(4f_s)$ = 40 µs. For adiabatic compression $P\Omega^\gamma$ = constant applies (P = pressure, Ω = volume = area × gap, γ = adiabatic ratio = 1.4 for air), so that after compression because of the pressure difference between underneath and above the plate a force

$$F_{sq} = \frac{\gamma AP}{z_0} Z, \tag{4.21}$$

which would outweigh the maximal driving force by a factor of 2.5. The fractional quantity could be interpreted to be a spring constant that contributes about 150 N/m, having the effect of letting the downcoming plate 'bounce off' the trapped gas layer and

pulling the uprising plate back. It is of a magnitude about equal to the modification the electric coupling applies to the stiffness at 40 V. The validity of this crude squeezed film damping model and the estimates of its effects hinge critically on the outflow of gas during compression of the trapped gas film; the high value calculated for Eq. (4.21) would let one suspect of a gross overestimation. There are many theoretical treatises on this subject that have right the description of this flow as central object of study, for example [10]. The squeezed film effect is preferredly formulated in terms of quality factors that contribute to damping (energy dissipation). In the context of damping it has to be noted that the compression process is invariably assumed to be *isothermal* rather than adiabatic (which would save the γ factor in Eq. (4.21)), so energy from the oscillation really irreversibly dissipates into the surrounding medium, underlining the damping characteristics of squeezed film, on top of the modification to beam stiffness.
A thorough study of the squeezed film effect would stretch the scope of this thesis too much. On a differently shaped sample, a single pressure-dependent experiment has been performed though. This sample, with the shape of a four-leaf clover, will be described in more detail in the next section, (4.4). The most important differences with the above described sample is that the clover sample is suspended on *four* beam spring pairs (rather than two; but of identical geometry) and has a considerably higher mass (23 mg). In a pressure range between 1 and 60 mbar the quality factor dropped roughly linearly from 48 to 19, indicating highly increased damping. Though damping is expected to slightly decrease the resonance frequency (though for $Q \sim 20$ less than a few per mill; a few Hz), it was found to *increase* by about 6%, for which an increased spring constant has to be accounted. Meanwhile, though Eq. (4.21) is now found to be a considerable overestimation of the effect of the squeezed film, the abundance of thorough studies of this effect in literature testifies to its importance to experiments optimized for resonance, which is underlined once more by this single experiment that not even optimized for his purpose.

4.4 TEMPERATURE DEPENDENT EXPERIMENTS ON DEVICE-LEVEL CHARGING EFFECTS THROUGH C(V) CURVES

This section enters the territory of device level effects related to charge trapping for the double beam samples. Though the structures encountered so far showed satisfactory behavior (that is: functioning, and in satisfactory concordance with expectation), it was deemed desirable to enhance sensitivity. This is required if one wants to study the possibly small effects of charge trapping. Sensitivity can be increased by thinnering the springs. Such a step would bring the structures better in accordance with the original design, because until this point a considerable discrepancy existed between design and product. In subsection 4.4.1 the sample used is mechanically characterized and tested for its immunity for the thermal variation in the large cryogenic temperature range (100-300 K) in which the $C(V)$-measurements are to be carried out. Subsection 4.4.2 will focus on the built-in voltage, the $C(V)$ parameter that is most directly associated with device level charge trapping effects.

4.4.1 Temperature dependent sample characterization

In order to gain sensitivity, some samples underwent an additional wet-etching step for further thinnering of the springs. In due course, the vertical columns interconnecting the corners of the L-spring structure described in section 4.2 were removed. Care was taken that aluminum was (re-)applied on both sides in order to minimize thermo-mechanical deformation during cooled experiments. The experiments described in this section were performed on the clover-shaped structure shown in Fig. 4.39. This structure has a large capacitive area compared to the other structures, which was countered by a bottom electrode to be an equivalent (series) area of A_s = 2.8 mm^2. A structure of similar shape has been used by [11] serving as an accelerometer in order to have a large mass and still relatively long springs in a compact structure. The mass of the structure presented in the current sub-section is calculated to be 23 mg. The suspension is relatively stiff because of the four pairs of springs that are relatively short (2.0 mm), especially compared to the L-springs sample described in section 4.2, but as already described in sub-subsection 4.1.5.3 that particular sample suffered from under-etching of the corners which entails additional risk to the mechanical stability of the structure. The clover structure is quite rigid and the thinnered springs proved to be sufficiently complying.

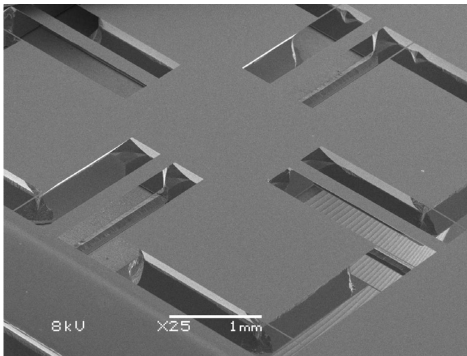


Fig. 4.39 Clover structure

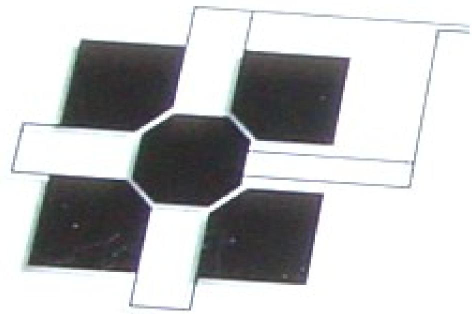


Fig. 4.40 Bottom electrode

The bottom electrode (Fig. 4.40) shows a central electrode and four interconnected clover leafs. The electrodes match the top structure (Fig. 4.39). The leads between the bottom clover leafs are laid such that the springs of the top structure do not electrically couple to them.

The clover structure used in the pressure dependent experiment mentioned in subsection 4.3.5 still had the original springs. Resonance measurements yielded f_s = 4980 Hz, from which a spring constant of 22.6 kN/m is deduced, a bit higher than twice the stiffness of the structure used for the bias dependent resonance measurements (section 4.3), which had two pairs of identically designed beam springs, rather than four. Some $C(V)$ - curves of that sample had been recorded as well, which yielded $u = \alpha / C_0^4 = 1.3 \cdot 10^{28}$ F^{-3}V^{-2}. This quantity (“intrinsic sensitivity”) normalizes the curvature to the capacitance and thus facilitates comparison between different structures concerning sensitivity, which is reciprocal to the stiffness of the suspension. For clover structures that underwent about 60 mins. of additional KOH wet – etching, this quantity increased by roughly a factor 30 to about $u = 0.4 \cdot 10^{30}$ F^{-3}V^{-2}. An example of a pair of additionally etched springs is shown in Fig. 4.41.

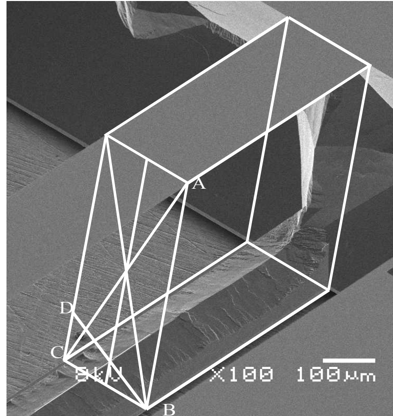


Fig. 4.41 Pair of thinnered springs. In this picture a geometrical construction can be made to estimate the height of the ridge along the center line of the lower beam. The distances AB and BC are known. The first provides the scale factor of the vertical dimension in the picture. DC can then be determined, which is twice the required height.

From construction, the spring height h was estimated 46 μm. This predicts k = 1348 N/m and f_s = 1218 Hz. Unfortunately, initial resonance experiments for this structure failed to provide useful results. Though 'something happened' in the neighborhood of the expected resonance frequency, a well-shaped resonance peak was lacking. Consequently, the prediction of the stiffness from construction cannot be corroborated by empiricism. From the $C(V)$ measurements, both the gap and the parallel capacitance are a priori uncertain parameters, so the isolation between experiment and theory remains in this case. No reason is found for this behavior. The top structure has a complicated shape, which potentially introduces several resonance modes, but this is an insufficient explanation, especially because with a similar structure satisfactory resonant behavior has been observed in a high vacuum system (section 4.5).
Nevertheless, series of $C(V)$-curves could be recorded in a large cryogenic temperature range, resulting in a meaningful inventory of the thermal dependent sample characteristics. The clover capacitance throughout a warm-up stage during the measurement series is shown in Fig. 4.42 and compared with an aluminum bridge HF sensor (section 3.3). The two devices have coincidentally comparable capacitance values.

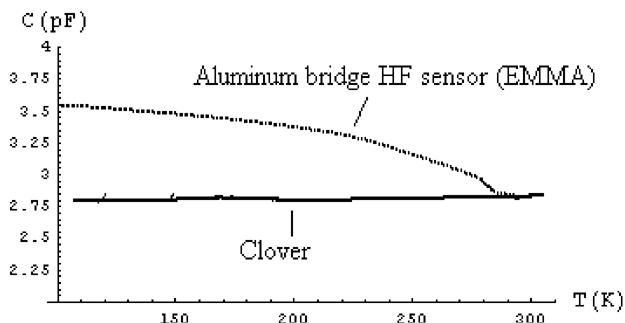


Fig. 4.42 Capacitance during the warming up phase of a clover sample and an aluminum bridge (section 3.3)

Clearly, the clover sample proves much more stable (variations in the order of 1%) than the aluminum bridge (more than 20% variation, compared to initial value) even in a not-stabilized temperature. The stability of the capacitance indicates that deformation of the top electrode is very small. Meanwhile, Table 4.2 displays the aforementioned quantity $u = \alpha / C_0^4$, which is a measure for the reciprocal stiffness of the device:

T (K)	50	100	200	300
u (10^{30} $F^{-3}V^{-2}$)	0.433	0.434	0.425	0.389

Table 4.2 Intrinsic sensitivity for different stabilized temperatures

The stability of the intrinsic sensitivity with temperature testifies that changes in stress are small during this range. The reader is referred to subsection 3.3.3 for a dramatic example of a temperature dependent spring constant (aluminum bridge sample), which varied for almost 2 orders of magnitude in the 4K-300K – range.

4.4.2 Built-in voltage

Long series of $C(V)$ measurements were parabolically fitted from which built-in voltages were deduced [12]. Temperature, ranging from 100 K to 300 K, was stabilized during the measurements series. Ambient air pressure could be maintained at about 10^{-3} mbar.
The drift in the built-in voltage was tempted to enhance by taking asymmetrical $C(V)$-curves (instead of symmetrical ones), that is $-40 < V < 0$ or $0 < V < 40$. This keeps the electric field in the same direction during the whole series, only varying in magnitude, like was done for the aluminum bridge samples in section 3.3.

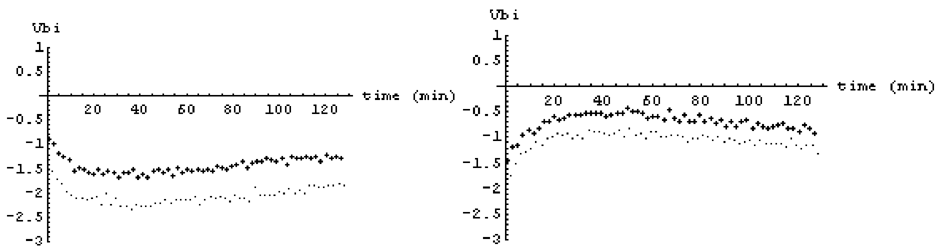


Fig. 4.43a (left) Built-in voltage for negative half-curves
Fig. 4.43b (right) Built-in for positive half-curves

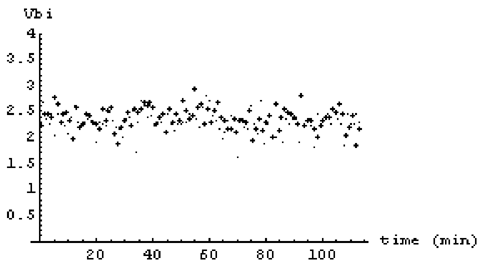
In the above presented measurement series (Fig. 4.43) a triangular voltage is applied (voltage running from 0 to ±40 V and back again). Thick dots represent built-in voltages (or ‘curve minimum shifts’) fitted from curves of ascending voltages, thin dots represent descending voltages. Horizontal is the time, spanning 128 minutes. From these measurements the last 80 (out of 120) curves are used to fit a (linear) drift rate. These measurements have been repeated for 200 K and 100 K, which showed decreasing drift rates with decreasing temperatures and drifted linearly during the complete series. Summarized in table 1 are the 300 K, 200 K and 100 K results, consisting of the built-in voltage together with their spread along a linear fit, and the drift rates.

Temp	Quantity	-40 < V < 0 ascending	-40 < V < 0 descending	0 < V < 40 ascending	0 < V < 40 descending
300 K	Vbi	-0.51 ± 0.057	-0.86 ± 0.051	-1.62 ± 0.050	-2.21± 0.053
	dVbi/dt	-0.258	-0.216	0.282	0.282
200 K	Vbi	-2.15 ± 0.21	-2.14 ± 0.17	0.51 ± 0.23	0.22 ± 0.18
	dVbi/dt	0.342	0.156	-0.180	-0.060
100 K	Vbi	-2.08 ± 0.078	-2.11 ± 0.115	2.44 ± 0.21	2.34 ± 0.22
	Vbi/dt	0.035	0.041	-0.099	-0.049

Table 4.3: Built-in voltages (V) and drift rates (V/m)

As discussed in section 2.3 and Chapter 6, it is not so much the magnitude of the built-in voltage, but rather its instability in time that is found of primary interest, although the possibility of the former influencing the latter is certainly conceivable. Previous to each measurements series a voltage of 0 V was applied in an attempt to create comparable starting values for the built-in voltage, but without success. From Table 4.3 it is clear that the magnitude of the built-in voltage drift decreases with decreasing temperature, which makes it harder to enforce a synchronization of starting values for the built-in voltage. We observe further that changing the polarity of the (unipolar) $C(V)$-curves changes the direction of the drift, which is understandable. However, the sign change between 200 K and 300K is less well understood.

The ‘built-in-splitting’ in Fig. 4.43 disappears at lower temperatures. At 200 K the two branches are still more or less separated, to become completely entangled at 100 K (Fig. 4.44). In Fig. 4.45 the built-in voltage is monitored during the warm-up stage mentioned above.



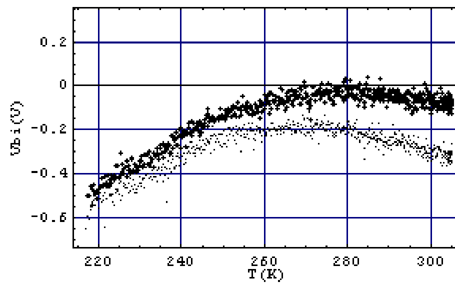


Fig. 4.44: Built-in voltage at 100K *Fig. 4.45 Temperature dependence of ‘built-in splitting’*

The influence of temperature of short-timed effect of built-in splitting is similar to what has been observed from the EMMA samples treated in chapter 3: The effect abates at lower temperatures, for aluminum electrodes between 150 K and 200 K, and almost disappears completely below 150 K. The suddenness of the (dis)appearance of this effect when set on the temperature scale varies somewhat. Especially for the EMMA samples quite sharp transitions have been observed.
Regarding the evolution of built-in voltage in time, a short simulation has been performed similar to what is described in subsection 2.3.1, see Fig. 4.46, be it here with an alternating voltage sweep between –10 V and 0 V and a spring constant of 3000 N/m. The discussion as was held in the mentioned subsection applies here quite well too. Again, though the model defects are clear, even this reproduces some important

characteristics: Built-in splitting and stabilization of the two levels in time, for ‘negative half-curves’ on the negative side. The model does not allow for fixing parasitic charges in deep traps, but for a thin oxide like native Al_2O_3, it is not expected that at room temperature large amounts will be remanent in the oxide. This is for an important part because, in such thin dielectrics (2-3 nm), the surface charges are very close to their images and feel encouraged to recombine with them, although in a ‘high-*k* dielectric’[25] like alumina, electric fields are weakened relatively much. Nevertheless, quite an amount of charge lingers in permanently in the oxide, because the fast bias sweeps does not allow for auto-decharging of the oxide. It has to be mentioned though that increasing the simulated sweep-period does not decrease the average stable built-in level (not only the decharging time, but also the charging time increases); it does however increase the splitting.

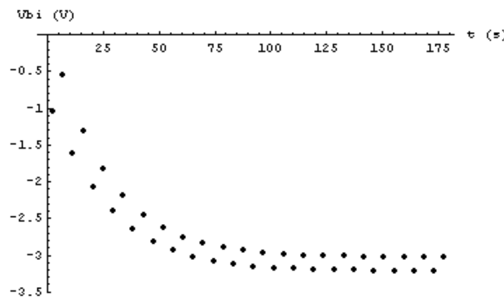


Fig. 4.46. Fitted Built-in voltages from series of simulated unipolar C(V) – curves.

4.5 HIGH VACUUM EXPERIMENTS

Another copy of the clover structure was subjected to experiments in a chamber in which a vacuum could be maintained considerably superior to the probe station in which all foregoing experiments in this chapter had been proceeded. The here reigning pressure of $5\cdot10^{-7}$ mbar potentially decreases the possible influence of ambient conditions on device-level charge-trapping effects. On the downside, this system did not provide for regulating the temperature, in particular no cooling could be supplied.
The vacuum could be attained only after at least 12 hours of heating, including the sample, which was deemed to withstand the conditions. A dedicated vacuum feed-through provided the electrical connections to the sample.

[25] High-*k* dielectrics are characterized by a relatively high dielectric constant (in these contexts denoted by *k*, rather than ε or ε_r) compared to silicon dioxide (SiO_2, 3.7). Examples include aluminum oxide (Al_2O_3, between 9 and 10), tantalum pentoxide (Ta_2O_5, ~25), hafnium oxide (HfO_2, ~30), which materials have recently entered electronic and semiconductor industry in order to reduce leakage currents, that manifest themselves increasingly strongly on with the downscaling of processor transistors, now 45 nm. Extreme values are reported for titanates, for example strontium titanate ($SrTiO_3$, 310), barium titanate ($BaTiO_3$, up to several thousand) and calcium copper titanate ($CaCu_3Ti_4O_{12}$, ~10^4) [13].

4.5.1 Sample characterization

From this copy of the clover sample, usable resonance frequency measurements could be drawn. At 1678 Hz there was a sharp peak with $Q = 480$ ($V = 24$ V), considerably higher than was reached in the milder vacuum of the probe station. For the high vacuum system, again parallel capacitance and gap were a priori uncertain parameters, which could be resolved by supplementing $C(V)$ recordings with resonance measurements. One representative curve out of a series yielded the parameters $V_{bi} = -0.53$ V, $C_0 = 7.10$ pF and $\alpha = 12.3$ aF/V^2. Because it was considered that, unlike the thick-spring clover sample, the bias voltage could have detuned the spring constant obtained from the resonance measurement, the system of equations

$$C_{M0} = \frac{\varepsilon_0 A}{z_0}, \qquad C_0 = C_{M0} + C_p, \qquad \alpha = \frac{C_{M0}^4}{2k\varepsilon_0^2 A^2} \tag{4.22}$$

is joined by the dynamic stiffness K_{dyn} now made dependent on all parameters and bias voltage:

$$K_{dyn} = k\left(1 - \Lambda V^2 - \frac{3}{2}\Lambda^2 V^4\right) = m\omega_s^2 = 2557\,, \qquad \Lambda \equiv \frac{\varepsilon_0 A}{kz_0^3}\,. \tag{4.23}$$

The relations close for $C_p = 4.60$ pF, $C_{M0} = 2.50$ pF, $z_0 = 9.9$ µm and $k = 2571$ N/m. Comparing the latter value with the dynamic value shows that with hindsight also in this case the bias dependency of the stiffness was redundantly taken into account. It is however almost a factor of two higher than as guessed from construction, see the preceding section. The from these parameters calculated intrinsic sensitivity of $u = 0.317 \cdot 10^{30}$ F^{-3}V^{-2} does not disagree too sharply with typical values encountered in the foregoing section. The dissimilarity could be understood if a spring constant of $k = 1.8$ kN/m is assumed there. To conclude: the spring constant derived from construction is somewhat less consistent with what is derived from measurements, compared to the previous section.

4.5.2 Built-in voltage

A few series of $C(V)$ - measurements with alternating upramping and down-ramping voltage sweeps (±15 V) have been carried out. These series invariably showed a relaxation behavior of the built-in voltage, as expected for no preference for one particular net polarity was applied in these series. An example is given in Fig. 4.47 which shows the first part of a 66 hour measurement series. Only the displayed data, truncated after five hours, were used for an exponential fit $e^{-t/\tau}$. Relaxation times τ are typically in the order of 2.5 – 3.5 hours for these high-vacuum series. For the clover sample in intermediate vacuum (1 mbar, probe station) comparable long-winded $C(V)$ - series had been recorded, showing relaxation in just over 2 hours. It has to be noted that the exact value of these times depend somewhat on the length of the tail of the data. The differences in relaxation times are therefore not thought to be significant, nor seems built-in splitting, which is present in both intermediate vacuum and high vacuum, to be significantly different. The latter effect does seem to be smaller than what is seen for the aluminum HF sensor (section 3.3), namely 0.07 V and 0.2 – 0.5 V (room temperature)

respectively. In the former case, the electric field is somewhat lower. Though higher actuation voltages are used, the particular configuration of the electrodes lowers the reigning gap voltage(s) again and capacitive gaps are larger, all compared with the latter case.

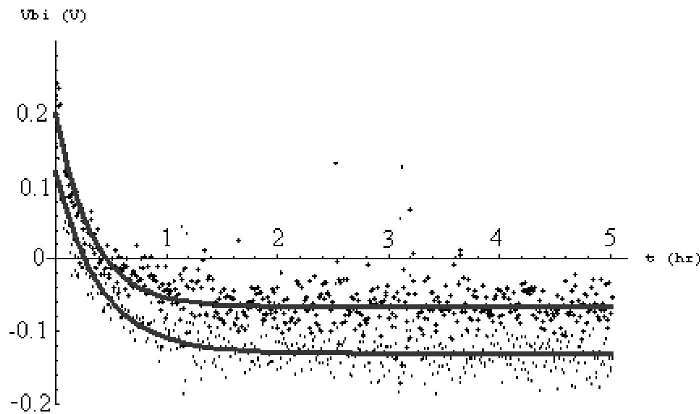


Fig. 4.47 Long term built-in voltage measurement. Built-in voltages for upramping voltage C(V) – curves are displayed in big dots, whereas small dots correspond to down-ramping voltage curves. The two branches, spaced about 0.07 V, are fit separately with relaxation times of 3.06 hours (upper) and 2.47 hours (lower).

4.6 CONCLUSIONS

The novel fabrication process applied to the top structure proved to be a feasible method to produce the springs as wanted (section 4.1). The resulting top structure devices show satisfactory and explainable static capacitance-voltage behavior (section 4.2). Mechanical characterization of complete devices (top and bottom electrodes) tested the suspension for resonant behavior in agreement with expectation, even if the devices are not optimized for vibrational purposes. A disagreement persists concerning the translation between electric and mechanical domain, though these results are not crucial for the mechanical characterization. (section 4.3). For enhanced sensitivity, some samples received additional wet etching treatment. This did work out as intended. The resulting devices proved to be considerably more stable in a large cryogenic temperature regime than the silicon nitride bridge and the aluminum bridge HF sensors from the EMMA project (as described in chapter 3) for the criteria of deformation (related to capacitance) and stress (related to $C(V)$-curvature), thus responding positively to the intentions declared in the design of the devices. The intensity of device-level dynamic characteristics related to trapped charges (built-in drift, built-in splitting) have been demonstrated to wane with decreasing temperature. (section 4.4) within the context of aluminum electrodes having its natural thin oxide as dielectric in ambient air pressure of 10^{-3} mbar (section 4.4). Some tests in high vacuum ($5{\cdot}10^{-7}$ mbar) indicate that the ambient air pressure does not have significant change in its potential influence on built-in relaxation, but perhaps on splitting. It is thought therefore that at least part of the observed built-in effects is intrinsic for the dielectric layer (section 4.5).

REFERENCES

[1] M. J. Madou, "Fundamentals of Microfabrication (2nd edition)", CRC Press, Boca Raton, Florida, USA (2002)
[2] M. Elwenspoek, and R. J. Wiegerink, "Mechanical MicroSensors", Springer, Berlin (2001)
[3] A. J. Nijdam et al., "Influence of the angle between etched (near) Si{111} surfaces and the substrate orientation on the underetch rate during anisotropic wet-chemical etching of silicon", *Journal of Micromechanics and Microengineering* **11** 499 (2001)
[4] A. J. Nijdam, "Anisotropic wet-chemical etching of silicon pits, peaks, principles, pyramids and particles", PhD thesis University of Twente (2001)
[5] H. Schroeder, E. Obermeier, and A. Steckenborn, "Micropyramidal hillocks on KOH etched {100} silicon surfaces: formation, prevention and removal", *Journal of Micromechanics and Microengineering* **9** 139 (1999)
[6] M. C. van Essen, R. J. Wiegerink, L. J. Fernández, H. Rogalla, and J. Flokstra, "Design and fabrication of stable capacitive test structures to study charge trapping phenomena", *proceedings of 15th MicroMechanics Europe (MME), Göteborg, Sweden (2005), p. 296*
[7] M. Vannoni, M. Trivi, and G. Molesini, "Phase-shift interferometry with a digital photocamera", *European Journal of Physics* **28** 117 (2007)
[8] David R. Lide (Editor in Chief), "Handbook of Chemistry and Physics (82nd edition)", CRC Press LLC, Boca Raton, Florida, USA (2001)
[9] D. R. França, and A. Blouin, "All-optical measurement of in-plane and out-of-plane Young's modulus and Poisson's ratio in silicon wafers by means of vibration modes", *Measurement Science and Technology*, **15** 859 (2004)
[10] P. Steeneken et al., "Dynamics and squeeze film gas damping of a capacitive RF MEMS switch" *Journal of Micromechanics and Microengineering* **15** 176 (2005)
[11] R. Puers, and S. Reyntjens, "Design and processing experiments of a new miniaturized capacitive triaxial accelerometer", *Sensors and Actuators A* **68** 324 (1998)
[12] M. C. van Essen, L. J. Fernández, J. A. Galán, R. J. Wiegerink, H. Rogalla, and J. Flokstra, "Charge trapping in cryogenically stable capacitive MEMS-devices", *proceedings of 7th Workshop on Low Temperature Electronics (WOLTE), Noordwijk, The Netherlands, 2006 pp. 179*
[13] A. P. Ramirez, et al., "Giant dielectric constant response in a copper-titanate", *Solid state communications* **115** 217 (2000)

CHAPTER 5

CONDUCTING AFM

In Chapters 2-4 theory and measurements of MEMS concerning temperature dependent effects of charge trapping on device level are presented. A different approach is treated in this chapter. The development of sensitive operation modes of non-contact AFM (frequency modulation) has opened the possibility of studying localized charges, with high spatial resolution. Section 5.1 follows the development of AFM, zooming in on the relevant technique to detect and image individual trapped charges. Relevant quantities in this field are provided. In section 5.2 a study is presented demonstrating images of trapped charges. For interpreting conducting AFM results it was considered indispensible to get a grasp of the interaction strength between the tip of an AFM and a sample with a localized charge. An available, widely used model (section 5.3) superficially yields satisfactory results, but upon closer inspection has implications of which some are quite suspicious. In this chapter, the relevant electrostatic problem is recalculated by a newly developed model (section 5.4) and found to predict very different results for this interaction. Generally, a much larger interaction strength is predicted by the method presented here. Moreover, it is shown in this chapter that the existing model contains a fundamental, intrinsic problem. The new "Multi mirror" model is in agreement with Finite Element Modeling (section 5.5). Direct model comparisons are offered in section 5.6. "Multi mirror" is brought into contact with practical results (section 5.7). For a thorough foundation and for the derivations of approximating analytical formulas be the reader referred to the Appendix.

5.1 PRINCIPLES OF AFM

5.1.1. The family of Scanning Probe Microscopy

The early eighties of the last century marks the beginning of a new era in the field of surface sciences, when Binnig and Rohrer announced [1] their invention of the Scanning Tunneling Microscope (STM) in 1981. The long coveted, but elusive Holy Grail of displaying separate atoms in real space had become reality. Within a year after its invention, this technique enabled captioning the Si(111) 7x7 surface [2], hereby resolving a long standing debate about its exact structure and properties, which helped to establish STM to a total success. STM is the earliest member of what has now become an extensive, growing and thriving family of Scanning Probe Microscopy (SPM). All techniques falling under SPM are based on sensing the surface of a sample with a small probe, specialized, dedicated and sensitive to the physical quantities of the surface the experimenter is aimed to map. In this respect, 'Microscopy' is unfortunately chosen terminology, as this is rather associated with optics, perhaps to be extended with Scanning Electron Microscopy (SEM), which uses electrons instead of light as reflecting / refracting substance.[26] STM is based on the existence of an electric tunnel current, which is strongly dependent on the exact distance between probe and sample surface.

Though STM is not a technique used for the research in this thesis, I will devote below a few words to this technique for it contains important principles upon which Atomic Force Microscopy (AFM) relies and has evolved from in a natural way: soon after the introduction of STM, Binnig realized that for the close distances between tip and sample surface significant forces of various nature are present, which inspired him, together with Quate and Gerber [3], to the invention and introduction of Atomic Force Microscopy in 1986. The AFM probe is sensitive to the small interactions between tip apex and a sample surface. The probe is attached to a cantilever, which movement is detected and interpreted in terms of topography and surface features of different nature. AFM will be discussed in a bit more detail below. Since then, many new SPM flavors have been added, such as for example Electrostatic Force Microscopy (EFM), Scanning Capacitance Microscopy (SCM), Kelvin Probe Force Microscopy (KPFM) and Magnetic Scanning Probe Microscopy (MSPM). The latter term is again covering various techniques, of which for example Scanning SQUID Microscopy (SSM) [4] recently achieved notable successes [5] with imaging the tiny magnetic fields of trapped flux vortices in superconducting materials with a resolution of about 2 micrometer.

[26] It is a debate about words, but the ancient Greek σκοπέω, "observe", might actually allow for a broader interpretation than "see by eye". I will leave this case to better versed classicists.

5.1.2. Scanning Tunneling Microscopy

Scanning Tunneling Microscopy owes its success to the strong relation between (magnitude of) tunnel current and tip, which allows for excellent resolution, in the first place in vertical, but also – and very importantly – in lateral direction, see Fig. 5.1.

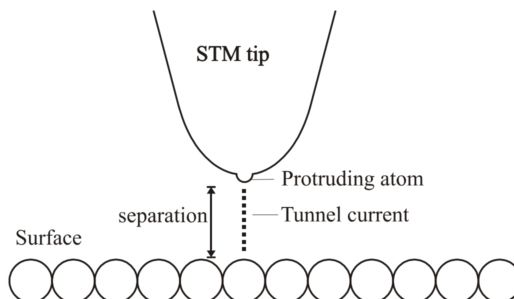


Fig. 5.1 Schematic interaction between STM tip and surface (not to scale). The strong correlation between tunnel current and separation distance allows for mapping a sample surface with atomic resolution.

In order to distinguish separate atoms in the lateral direction, it is imperative that the tip atom closest to the surface (the exaggeratedly drawn protruding atom in Fig. 5.1) assumes a significant part of the interaction with the sample. For tunneling currents, the tip does not need to be exceedingly sharp for this to be the case; the front atom can still be relatively embedded in the surrounding material. The tunnel current is small (typically 0.1-1 nA), but does not at all require highly esoteric electronics to be converted to a voltage signal, which is used as rudimentary data. The surface is scanned in raster-like fashion and for each point on a lateral (x,y)-grid a topographic height, directly derived from the current, is assigned. This provides a 3D surface plot of the sample. In its simplest form, the tip senses the surface from a fixed height and the varying current is recorded. However, as this risks the tip colliding with the surface in case of a sudden protrusion from the surface, the tip height is usually adjusted by piezoelectric elements, dictated by a feed-back circuit tuned to keep the tunnel current constant. This way, the tip is maintained at a constant distance from the surface. The voltage over the piezocrystals provides the topographic information and hence serves as data output. This feed-back set-up allows for recording larger ranges of height differences and does not require the sample surface to be extremely parallel to the plane traced out by the raster-scanning tip apex.

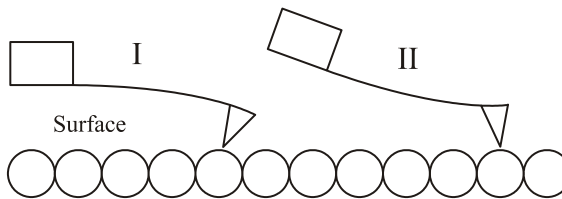
5.1.3. Atomic Force Microscopy

Until today, the resolving capabilities of STM are unsurpassed and its easy operability has helped it to be a standard instrument in the present day surface scientist's arsenal. There are however obvious limitations to this technique: only conducting samples can be imaged. By contrast, Atomic Force Microscopy can image surfaces of virtually any nature, including polymers and organic molecules. Operation environments include vacuum, ambient air and even liquids. Also, unlike STM, AFM is not restricted to topography, but can image many sorts of interactions and be employed for active

manipulation on the nano-scale. All this makes AFM to a very versatile tool indeed. By their introduction of the AFM in 1986, Binnig et al. expressed their expectations of the AFM having the potency of atomically resolved surface imaging. It took five more years however before it came to be [6], while the famous Si(111) 7x7 surface had to wait until 1995 for being exposed in real space by AFM [7]. Complicating factors in AFM to be a match for the STM's resolvability are firstly that the probe signal (movement) is only indirectly convertible to usable data, i.e. a voltage signal; this is much easier achieved in STM which probe signal is a directly usable current. Secondly, the type of forces the AFM senses are often less strongly distance-dependent. This requires AFM tips to be extremely sharp, in order to have the front atom optimally protruding from the neighboring tip atoms.
Atomic Force Microscopy technology has much in common with STM; the difference is mainly the type of probe. The tip of an AFM is attached to a cantilever, which shows a varying deflection during scanning. This movement is read out and interpreted in terms of topology and possibly different kinds of interaction. Initially, cantilever movement was detected by STM. Later it became common to invoke interferometry or the refraction of a laser beam, aligned on the cantilever apex, for movement detection.
AFM developed into two main branches of operation: A) static AFM (or contact AFM) and B) dynamic AFM (or non-contact AFM).

A) *Static AFM*

In case of static AFM, the tip is scanned over the surface and its deflection is read out. Usually, the tip is brought to contact with the surface (contact mode AFM), see Fig. 5.2, but this is not necessarily the case.



Fig, 5.2 Contact mode AFM, with either I) the tip being allowed to be attracted by adhesive forces and the cantilever mount gently trying to lift it, or II) the tip being pressed against the surface, when repulsive forces dominate the tip-sample interaction.

The feed-back parameter is the deflection of the cantilever, which height is piezoelectrically adjusted so as to maintain constant deflection. The piezoelectric signal provides a topographic image. Torsion around the cantilever beam axis (perpendicular to the deflection indicated in Fig. 5.2 can be utilized to study frictional forces.
Contact mode cantilevers are usually soft (around or below 1 N/m) in order to avoid too much deformation of both the tip and the sample. Operation in vacuum is not straightforward, because this kind of cantilevers are very sensitive to external vibrations, but ambient air and even liquids are suitable working conditions. Under special circumstances, atomic lateral resolution is achievable.

B) *Dynamic AFM*

Dynamic Atomic Force Microscopy is comprehensively reviewed in amongst others Garcia and Pérez [8]. With dynamic AFM, the cantilever is deliberately imposed a vibrational motion, usually at or near its free resonance frequency. Permanent physical contact with the sample is avoided. In most cases, only attractive forces are being probed. A counterexample is for instance an electrostatically biased tip probing a local trapped surface charge of equal polarity. Two examples of tip-sample interactions are provided below. One important force is the attractive Van der Waals force, which in its simplest form[27] is given by:

$$F_{vdW} = -\frac{HR}{6z_t^2}\,. \tag{5.1}$$

Here H is the Hamaker constant, depending on material properties of tip and sample (typically in the order of 1 eV [10]) R the tip radius of curvature and z_t the instantaneous distance between tip and sample. The Van der Waals force is a result of fluctuations in the electric dipole moment of individual atoms. The bonding energy of two atoms at mutual separation z is proportional to z^{-6}. Additivity and replacing the sum over the discrete (atomic) contributions by an integral leads to the simplified form Eq. (5.1). This relatively slow weakening of the force with distance enables the Van der Waals force to manifest at relatively long ranges. A well-known expression for short-range forces is the *Lennard-Jones* (LJ) potential:

$$V_{LJ} = -U_c\left(2\frac{z^6}{z_0^6} - \frac{z^{12}}{z_0^{12}}\right), \tag{5.2}$$

where U_c is a chemical bonding energy, z the instantaneous tip sample separation and z_0 the equilibrium distance. The first, adhesive term stems from a short-range Van der Waals interaction. The second term is repulsive and is only manifest at extremely short approach, where it will very quickly dominate.

Characteristic for dynamic AFM is the use of fairly stiff cantilevers (resonance frequencies ranging from 1 kHz to above 100 kHz), an important advantage of which is to help prevent “jump-to-contact” (JTC): the tip permanently sticking to the surface under influence by adhesive (Van der Waals) forces. The vibrational characteristics (resonance frequency, amplitude, phase shift) can be read out and manipulated very accurately. Also dynamic AFM is commonly operated using feed-back control, which any of the mentioned vibration characteristics are suitable for. When the tip is brought close to the surface, tip-sample forces modify these characteristics. The large range of types of encountered forces, and the complexity of properly treating the dynamics of large amplitudes (which are often used; up to 100 nm), make the interpretation of the data all but a trivial task.

On this basis, dynamic AFM has branched further into two classes, namely “Amplitude Modulation” (AM) and “Frequency Modulation” (FM).

[27] Valid for a geometry defined by a spherical tip and a flat surface

1) Amplitude Modulation

In AM-mode, the tip-sample distance is controlled so as to maintain a constant amplitude of the vibration, which is affected by tip-sample interaction forces. This distance is used as topology signal. Peculiar for amplitude modulation is that despite being a dynamic AFM technique, a variation developed that was successfully employed in the repulsive regime, i.e. by *intermittent contact* between tip and sample, for which Veeco® registered the term “tapping mode”.
In the following lines, a few notions about AFM dynamics will be briefly reviewed along the concept of a ‘weakly perturbed harmonic oscillator’. A full mathematical description of a vibrating cantilever with tip would be an arduous task, but for many purposes it suffices to consider the cantilever motion in terms of beam deflection theory. Often, the problem is further simplified by considering only the tip motion governed by the following equation for a damped mass-spring system:

$$m\frac{d^2z}{dt^2}+kz+\frac{m\omega_0}{Q}\frac{dz}{dt}=F_{ts}+F_0\cos\omega t\,. \tag{5.3}$$

Here, m is the dynamic mass, k the spring constant, ω_0 the undamped angular resonance frequency, z the position of the tip, Q the quality factor of resonance, F_0 the driving force amplitude[28], ω the angular driving frequency, and F_{ts} the collection of all tip-sample forces, which can depend in numerous ways of the separation distance with the sample, hence on z. For $F_{ts}=0$ we recognize the standard problem of a forced, damped harmonic oscillator. In this case, the amplitude A of the motion performed by the tip is given by (see also subsection 4.3.5),

$$A(\omega)=\frac{F_0/m}{[(\omega_0^2-\omega^2)^2+(\omega\omega_0/Q)^2]^{1/2}}\,, \tag{5.4}$$

which is maximized when excited at (damped) resonance frequency:

$$A_0=\frac{QF_0}{k}\,. \tag{5.5}$$

The damping has the effect of decreasing the resonance frequency by

$$\omega_D=\omega_0\sqrt{1-\frac{1}{2Q^2}} \tag{5.6}$$

and modifying the phase by

$$\tan\phi=\frac{\omega\omega_0/Q}{\omega_0^2-\omega^2}\,. \tag{5.7}$$

[28] Not to be confused with the motional amplitude of the vibrating cantilever obviously.

Here ϕ is defined as the phase in the steady component of the complete solution to Eq. (5.3) given by $A\cos(\omega t - \phi)$. In the case of forces additional to the driving force, the total force becomes for small displacements:

$$F = F_0 + (z - z_0)\left(\frac{dF}{dz}\right)_{z=z_0} . \qquad (5.8)$$

This affects the effective stiffness through

$$k_e = -\frac{dF}{dz} = k - \left(\frac{dF_{ts}}{dz}\right)_{z=z_0} . \qquad (5.9)$$

As a consequence, the angular resonance frequency is modified by:

$$m\omega_e^2 = k - \frac{dF_{ts}}{dz} . \qquad (5.10)$$

By now it has become clear how the force gradient affects the three vibrational parameters phase shift, resonance frequency and hence amplitude. It has to be stressed that the above treatment is strictly only useful to develop a qualitative notion of the influence of the force gradient. This simple harmonic approach fails clearly to reproduce experimentally observed quantitative results, for various reasons. Among them are the assumptions of the force *gradient* being independent of position Eq. (5.8) and the force gradient smaller than the cantilever stiffness Eq. (5.9). For the large amplitudes that are commonly employed in dynamic AFM, these conditions are never both met. For example, the force gradient can vary with orders of magnitude within the oscillation trajectory. Furthermore, in intermittent contact mode, energy and momentum is transferred between tip and sample, which is not considered here. The reader will recognize the difficulty of interpreting the data from AFM operated at amplitude modulation. An extra source of ambiguity is the existence of two stable oscillation modes of the cantilever when in interaction with tip-sample forces, upon which not will be digressed here further.

2) Frequency Modulation

Thermal noise puts a lower limit to the minimum in frequency shift that can be detected with amplitude modulation. This minimum is given by:

$$\delta(\Delta f) = \sqrt{\frac{f_0 k_B T B}{4\pi k Q \left\langle z_{osc}^2 \right\rangle}} . \qquad (\underline{\text{AM}}\text{-mode}) \qquad (5.11)$$

Here, Δf is the shift in resonance frequency due to tip-sample interaction, f_0 is the cantilever's free resonance frequency, k_B Boltzmann's constant = $1.38 \cdot 10^{-23}$ J/K, T the temperature in K, k the cantilever stiffness, Q the quality factor of resonance and $\left\langle z_{osc}^2 \right\rangle$ the mean-square amplitude. B, the bandwidth of operation, is typically the number of pixels scanned per second. Sensitivity could be enhanced by operating the AFM in

vacuum, which would increase the quality factor by typically two orders of magnitude compared to ambient air. However, the operability in vacuum of amplitude modulation AFM is problematic. While it is possible in principle, it is unmanageably slow. When a scan is started, the cantilever oscillation contains a transient term added to the steady state term mentioned underneath equation Eq. (5.7). This transient term is a decaying exponential characterized by a decay time $\tau = 2Q/\omega_0$. Any change in the conditions, for example moving the tip to the next grid point to scan, which may for example alter the tip-sample distance due to slope or corrugation, causes transient terms to re-appear in the cantilever motion. Given the high quality factors that are typical for vacuum (unlike for example air), it means that reliable amplitude measurement can commence only after a stabilization time in the order of 1 second. For the large number of grid points that are usual in contemporary AFM (at least 256 × 256), a scan would take much too long, which is undesirable not only for general reasons, but also will drift effects (e.g. thermal) render the process of reliable imaging highly problematic.
Nevertheless, Albrecht et al. [12] demonstrated a viable solution for satisfactory operation of Non-Contact AFM in vacuum. Instead of the amplitude, the resonance frequency was selected as feedback parameter. While amplitude measurements needs to wait for the transient motion to damp out, the resonance frequency, affected by tip-sample forces, adapts to the new conditions within typically one oscillation cycle. Prior to a scan, the tip is brought to oscillation at resonance frequency while still far from the sample. Then a fixed shift in resonance frequency is set. The tip is then lowered until the new resonance frequency, now influenced by interactions with the sample, differs exactly this amount from free resonance. During the scan, the tip height is adjusted to keep this frequency constant, which is again associated with the surface topology. A second feedback loop is coupled to the *driving force* to keep the amplitude constant. According to [12], the minimum detectable frequency shift in FM AFM is given by:

$$\delta(\Delta f) = \sqrt{\frac{f_0 k_B T B}{\pi k Q A_0^2}}\,, \qquad \text{(FM-mode)} \qquad (5.12)$$

which is apart from a numeric factor identical to (5.11). Unlike with AM, where an increase of Q would directly negatively affect the bandwidth, these quantities are uncoupled in FM. It is no surprise that it was exactly FM with which non-contact AFM was reported first [6] to have achieved atomic resolution, including again the famous Si(111) 7x7 mapping [7]. Later, even sub-atomic features, associated with chemical bond lobes, have been resolved [9]. Readers interested in the background of FM-AFM and many of its aspects are recommended to consult [10].

5.1.4 Frequency shift in FM AFM

It is thought worthwhile to mention a few more important expressions in the field of Frequency Modulated AFM, see also [11] and [12]. For small oscillation amplitudes and under the condition that the driving force compensates energy loss through damping and tip-surface interactions, the frequency shift is simply related to the interaction force gradient via:

$$\Delta f(z_c) = -\frac{f_0}{2k}\frac{\partial F_{ts}}{\partial z}\bigg|_{z=z_c} . \tag{5.13}$$

As interaction forces wane far from the sample, the frequency shift is *negative* for *attractive* forces. However, in general large amplitudes are being applied in NC AFM, in which case Eq. (5.13) ceases to be valid. For weakly disturbed, large oscillations, several authors provide the following dependence of the frequency shift on the interaction forces:

$$\Delta f(z_{\min}, k, A_0, f_0) = -\frac{f_0}{2\pi k A_0}\int_0^{2\pi} F_{ts}[z_{\min} + A_0 + A_0\cos\varphi]\cos\varphi d\varphi . \tag{5.14}$$

Here z_{min} is defined as the distance of closest approach. The square brackets denote functional dependence here.
A final quantity should be reproduced here. It has been demonstrated by Giessibl [13] that in the limit of the oscillation amplitude much larger than the range of force, the operational quantities in Eq. (5.14) can be extracted from the integrand, leaving a quantity that is dependent on the interaction force only. This *normalized frequency shift* is related to the ordinary frequency shift via

$$\gamma(z_{\min}) = \frac{kA_0^{3/2}}{f_0}\Delta f(z_{\min}, k, A_0, f_0) , \tag{5.15}$$

In concreto:

$$\gamma(z_{\min}) \approx \frac{1}{\pi\sqrt{2}}\int_0^{\infty}\frac{F_{ts}[z_{\min} + z]}{\sqrt{z}}dz . \tag{5.16}$$

This approximation is valid for interaction forces that drop off with distance with at least a quadratic rate or exponentially, which holds for most common forces. In those cases:

$$F = \frac{C}{z^n} \quad \Rightarrow \quad \gamma(z_{\min}) = \frac{1}{\pi\sqrt{2}}\frac{\Gamma(n-½)}{\Gamma(n)}\frac{C}{z_{\min}^{n-½}} , \tag{5.17a}$$

$$F = Ce^{-uz} \Rightarrow \quad \gamma(z_{\min}) = \frac{Ce^{-uz_{\min}}}{\sqrt{2\pi u}} . \tag{5.17b}$$

Here, C and u are constants. Γ is the ordinary gamma function, interpolating the discrete-valued faculty-function. These expressions Eq. (5.17) provide good agreement with experiment.

5.2 CHARGE IMAGING ON THIN OXIDE FILMS

As indicated in subsection 5.1.1, Atomic Force Microscopy is capable of investigating a large variety of physical quantities that can be defined for surfaces and thin layers of a very broad range of materials. Within the context of this thesis, we are interested in small-scale features of the electric properties of thin dielectric layers. For this, AFM can be supplemented with Kelvin Probe technology by the use of a conductive tip. Given the particular AFM operation mode commonly employed in the research described in this section, we speak of "Frequency Modulated Kelvin Probe Force Microscopy" (FM KPFM). With this technique, fluctuations in the work function of a given sample surface can be revealed with a lateral resolution better than 50 nm. Within the context of KPFM, the term "Contact Potential Difference" (CPD) is often preferred over 'work function', to stress the difference in potential of the conductive tip and the small sample surface directy underneath it. Generally, the interaction stemming from CPD probed by a tip biased at a voltage V with respect to the electrode underneath the dielectric layer can be derived from the electrostatic potential between tip and sample:

$$U = \tfrac{1}{2}C(V - V_{CPD})^2 . \qquad (5.18)$$

Where C is the tip-sample capacitance, which depends on their separation z. The resulting force and the gradient therein are obtained by once and twice differentiating with respect to the normal coordinate:

$$F = \frac{\partial U}{\partial z} = \frac{1}{2}(V - V_{CPD})^2 \frac{\partial C}{\partial z} , \qquad (5.19)$$

$$\frac{\partial F}{\partial z} = \frac{1}{2}(V - V_{CPD})^2 \frac{\partial^2 C}{\partial z^2} . \qquad (5.20)$$

Note in Eq. (5.19) the *positive* sign in $F = +\partial U/\partial z$. The reasons for this are firstly the fact that the cantilever tip is a voltage controlled system and secondly that charge and capacitance are linearly related. See the discussion in subsection 2.1.1, specifically Eqs. (2.1) to (2.5).
Like Van der Waals forces, electrostatic forces can be long-ranged and consequently contribute substantially to the genuine topographic signal. Repeated scans at different bias voltages V are needed to untangle the interactions, but this is not a straightforward procedure. A more elegant alternative is applying a modulated bias voltage at frequency ω_V: $V(t) = V_{DC} + V_{AC} \sin(\omega_V t)$. Equation (5.20) then becomes:

$$\frac{\partial F}{\partial z}(\omega_V) = \left[\frac{1}{2}(V_{DC} - V_{CPD})^2 + \frac{1}{4}V_{AC}^2 + V_{AC}(V_{DC} - V_{CPD})\sin(\omega_V t) + \frac{1}{4}V_{AC}^2 \cos(2\omega_V t) \right] \frac{\partial^2 C}{\partial z^2} \qquad (5.21)$$

We observe that the applied voltage results in a constant contribution (first two terms), a component at angular frequency ω_V dependent on V_{DC} and a component at angular frequency $2\omega_V$ independent of V_{DC}. The DC-voltage that removes the first harmonic is equal to the CPD. Obviously, this requires an additional feed-back loop. The

disadvantage of this method is the risk that artifacts of the bias modulation can show up in the ultimate CPD-image. We will not discuss the details of this further.
Imaging of local charges on thin Al_2O_3-films has been described in [14] and [15] and more comprehensively in [16], though the field of imaging surface charges has aged for almost 20 years already, beginning with the work by Terris et al. [17]. Many authors have reported important contributions since then.
The ultrathin (2.5nm) alumina layers were deposited by Atomic Layer Deposition (ALD) on p-type silicon (001). The conducting AFM operated in Ultra High Vacuum (UHV; $5\cdot10^{-11}$ mbar) in non-contact Frequency Modulated mode. The cantilever is read-out in near-resonance (45-115 kHz [15]) by detecting the deflection of a laser beam by a four-quadrant photo diode. When a bias voltage is applied to the tip, electric charges manifest as large protrusions or indentations, depending on the polarities of the bias voltage and the surface charge. For a detailed description of the set-up and the technique of imaging be the reader referred to the aforementioned thesis.
Protrusions caused by charges can be distinguished from proper topographic features by re-scanning the same area with different bias voltages. In particular, reversing the bias polarity causes the image of localized charges to reverse color, whereas the proper topographic map remains essentially unchanged. Reversing the bias voltage thus decides whether a particular feature is a local charge. If affirmative, the polarity of the charge can then also be immediately determined. An alternative for doing repeated scans with changed bias voltage is scanning with modulating bias, as proposed and demonstrated in [17]. The electrostatic interaction of a biased tip (V) with a conducting sample having a localized surface charge Q in its dielectric tip layer is of the form:

$$F_{el} = aQ^2 + bQV + cV^2 \,. \tag{5.22}$$

In Eq. (5.22), a, b and c are constants depending on the geometry of the electrostatic problem and material properties, such as the dielectric constant of the top insulating layer. In absence of any charge, only the last term of Eq. (5.22) survives. For a modulated bias voltage of the form $V(t) = V_0 \cos \omega_V t$, the electric force becomes in that case

$$F_{el} = \tfrac{1}{2} cV_0^2 + \tfrac{1}{2} cV_0^2 \cos 2\omega_V t \,, \tag{5.23}$$

so an interaction signal from the AFM will contain a component with angular frequency $2\omega_V$. Now in presence of a charge Q, the second term in Eq. (5.22) will give rise to an extra signal component with a frequency identical to the angular bias frequency ω_V. This allows for imaging charges with a single scan. Assuming V_0 positive, the signal from a positive charge would then be in phase with the applied voltage. A negative charge would be in anti-phase.
Here, I will restrict to reproducing a representative result (Fig. 5.3) done with varied constant bias voltage. Fig, 5.3 captions rather fortuitously two charges of opposite polarity in a single picture. It is reported that most charges encountered on Al_2O_3 are negative, with a density of 0.4 ± 0.2 μm^{-2}. Positive charges are much less abound, no reliable density has been established. Models for quantitative analysis of local charges (force on tip, depth and intensity of the charge) are presented in the following sections. A charge on a silicon dioxide film (standard test sample of the AFM set-up), captured using the same setup, is shown in Fig. 5.4.

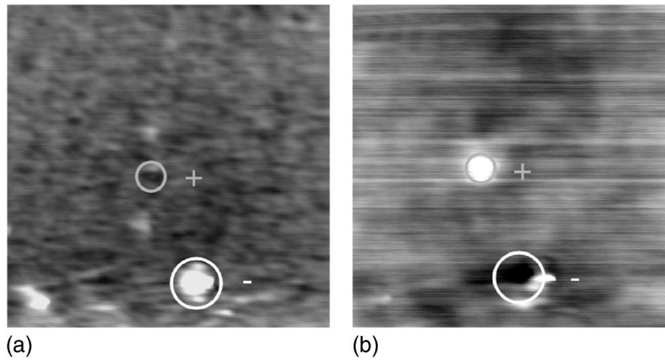


Fig. 5.3 Topographic scans (1.5 x 1.5 µm²) with a negative and positive charge. Bias voltage is -0.69 V (a) and +0.60 V (b) respectively (reproduced from [15]).

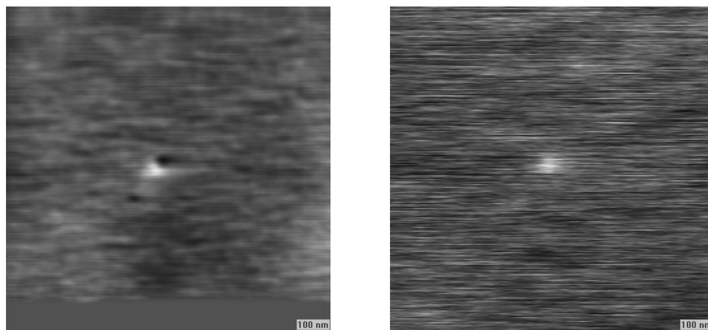


Fig. 5.4 Trapped surface charge on SiO_2 on Si. Left: Topography; Right: CPD. Bias: -0.5V; scanning window: 1 × 1 µm².

A different, interesting experiment concerning the dynamics of charge trapping is described in [18]. Here the tip of a conducting AFM was used to deposit charges on a dielectric layer, anodized barrier-Al_2O_3. After charge deposition and retraction of the tip, the tip scanned over the charge multiple times. The peaks in these scans, resulting from the charge-tip interaction, decreased in time. From the specific decay patterns in time, conclusions were drawn about the vertical distribution of traps in the dielectric layer. It was derived, that for oxides thicker than about 30 µm, the charge trap density becomes differentiated: a higher trap density close to the surface, and a lower trap density at greater depths, closer to the Al_2O_3 interface.

5.3 THE ELECTROSTATIC PROBLEM AND EARLIER MODELS

This section starts with the introduction of the electrostatic configuration, the relevant quantities and parameters. Then, an existing model by Ludeke is described, and its amendment by Lambert. Numerical results of this Ludeke-Lambert model are deferred to section 6.6, where this model is compared with the Multimirror model presented in section 6.4 and Finite Element modeling treated in section 6.5.

5.3.1 The electrostatic problem

The electrostatic problem under consideration is the one as depicted in Fig. 5.5, which shows the tip of a conducting AFM probing a localized charge in a dielectric layer on top of a conductor. The tip is represented as a conducting sphere, with radius *R* and put to a bias voltage *V* with respect to the sample. This bias is usually represented by an appropriate point charge ($q_V = 4\pi\epsilon_0 RV$) located in the centre of the conducting sphere. There is a space *a* between tip and sample surface. The dielectric is of thickness *d* and has a dielectric constant *ε*. The trapped charge *q* is located a distance *s* above the metal/dielectric interface.

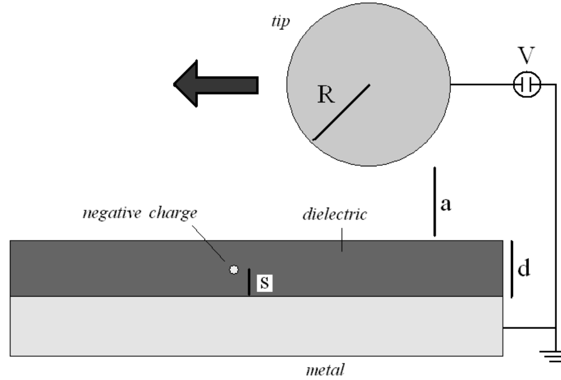


Fig. 5.5. Electrostatic configuration

It can be shown (see Appendix) that the electrostatic force must be of the form

$$F_{el} = \alpha_{VV} V^2 + \alpha_{qV} qV + \alpha_{qq} q^2 . \tag{5.24}$$

Here *V* is the bias voltage and *q* is the amount of trapped charge. All expected electrostatic interactions can be grouped into three classes; the first is the interaction of the bias-charged tip with the sample. All bias-related charge in the tip scales with *V* and so does the induced charge in the sample. This holds true also if one considers that this induced charge in the sample on its turn again induces charge in the tip, etc. Automatically, the summed interaction scales with V^2. The second class is the interaction between the trapped charge *q* in the sample and the bias charge q_V in the tip, which yields the second term. Finally, the trapped charge interacts with its own image in the tip, which the third term is accounting for. Non-electrostatic interactions are not considered in this context, nor any additional static charges or varying surface potentials (extra terms linear

in *V* or *q*). The coefficients *α* are functions of the dielectric constant and the geometric parameters: the oxide thickness, the tip radius, the tip spacing and the horizontal tip centre-charge distance, if there is any. The latter two (α_{qV} and α_{qq}) depend also on the vertical coordinate of the trapped charge (positioned at $x = y = 0$); α_1 does not, for the pure tip-substrate interaction cannot depend on the locations of charges that happen to be around.

As stated in section 5.1, the relevant parameter in Force Modulated Atomic Force Microscopy is the force *gradient*, from which the shift in cantilever resonance frequency can be determined.
The spatial gradient of the electrostatic force in the (vertical) z-direction is of similar form as Eq. (5.24):

$$\frac{\partial F_{el}}{\partial z} = \beta_{VV} V^2 + \beta_{qV} qV + \beta_{qq} q^2 , \qquad (5.25)$$

in which the all coefficients β are determined by $\beta = \partial\alpha/\partial z$. In this and the following sections, two models will be presented and compared, that aspire to solve the electrostatic problem, or at least to determine the interaction between tip and sample plus trapped charge. It is in these coefficients that the differences between the models, or calculation methods, treated here transpire.
To add further meaning to this, it is sought to compare the models on the level of experimentally relevant parameters. A localized bias-varied probing measurement principally yields two distinct, independent quantities. Firstly there is the shift in the minimum of the parabola, V_{min}, which is the voltage with which during a measurement the applied pre-set bias voltage is modified in order to minimize the force gradient.

$$V_{\min} = -\frac{\beta_{qv} q}{2\beta_{VV}} , \qquad (5.26)$$

It is immediately apparent from Eq. (5.26) that V_{min} scales with q. The second is the 'peak height', which is the difference between the interaction levels of a charge present or absent, which is equivalent of comparing the interaction strength of the tip in the charge's zenith with far away from the charge (in the horizontal direction). A linear scan in the x-direction will show a peak compared to the background bias interaction when passing over the charge:

$$\Delta F' \equiv \frac{\partial F_{el}}{\partial z}(x = x_q) - \frac{\partial F_{el}}{\partial z}(x \to \infty) = \frac{\partial F_{el}}{\partial z}(q = q) - \frac{\partial F_{el}}{\partial z}(q = 0) . \qquad (5.27)$$

Consequently,

$$\Delta F' = \beta_{qV} qV + \beta_{qq} q^2 . \qquad (5.28)$$

5.3.2 Earlier models

Earlier work on the solution of the above electrostatic problem (Fig. 5.5) has been published by Ludeke and Cartier[19], who consider only the interaction of the trapped charge with its image in the tip and the tip's bias charge. For the force gradient, the relevant quantity in frequency modulated non-contact conducting AFM, they give:

$$\frac{\partial F_z}{\partial \tilde{z}} = \frac{q^2 R}{4\pi\varepsilon_0 \varepsilon_{eff}^2} \frac{3\tilde{z}^2 + R^2 - l^2}{(\tilde{z}^2 + l^2 - R^2)^3} + \frac{qV_b R}{\varepsilon_{eff}} \frac{l^2 - 2\tilde{z}^2}{(\tilde{z}^2 + l^2)^{5/2}} . \qquad (5.29)$$

Ludeke defines $\tilde{z}$ (I reserve ordinary 'z' for the vertical distance from the metal plate) as the vertical distance between charge q and tip centre, $= d - s + a + R$. l is the horizontal distance between charge and tip centre and V_b is the bias voltage. In [19] the second term has a minus sign, as he puts the tip on ground instead of the sample.
The effect of the dielectric layer is accounted for by defining an 'effective dielectric constant'

$$\varepsilon_{eff} \equiv \frac{1}{2}(\varepsilon + 1)\,. \tag{5.30}$$

This factor accounts for the change in apparent strength of the source charge q as a result of polarization of the surrounding dielectric material, when observed through a interface between dielectrics; the assigned terminology seems thus a bit misleading as a 'dielectric constant' is rather a factor influencing field strength. In the section on elementary image problems we will encounter this factor again.
A refinement on this model is the inclusion of the interaction between tip and sample, for which Lambert and Saint-Jean [20], provide an extra contribution that I'll call $F_{sub,z}$ (substrate):

$$\frac{\partial F_{sub,z}}{\partial \tilde{z}} = \frac{qR\tilde{z}\left(-4\pi\varepsilon_0 R V_b + \dfrac{qR}{\varepsilon_{eff}\sqrt{\tilde{z}^2 + l^2}}\right)}{8\pi\varepsilon_0\varepsilon_{eff}(r + a + d/\varepsilon)^2(\tilde{z}^2 + l^2)^{3/2}} + \frac{\left(-4\pi\varepsilon_0 R V_b + \dfrac{qR}{\varepsilon_{eff}\sqrt{\tilde{z}^2 + l^2}}\right)^2}{8\pi\varepsilon_0(r + a + d/\varepsilon)^3}\,. \tag{5.31}$$

The total gradient is the sum of these gradients (1) and (3). Given this explicit expression for the force gradient, the β-coefficients of Eq. (5.25) are readily extractable. After integration of Eq. (5.29) and (5.31) over $\tilde{z}$, the α – coefficients come out as well. It is not thought functional to provide the explicit expressions for those coefficients here. N.B.: In the Ludeke-Cartier expression Eq. (5.29-31), $\beta_{VV} = 0$. The pure bias-interaction is not considered there. In itself, this is no serious deficit, as this interaction plays no role in the study of trapped charges. It does not appear in the experimental quantities Eqs. (5.26) and (5.28).
In [16], yet a different expression is suggested for the 'effective dielectric constant', in order to better accommodate the Ludeke – Lambert model to the observed electrostatic phenomena in the case of ultrathin (2.5 nm) oxides.

$$\varepsilon_{eff} = \frac{\varepsilon_{ox}(a + d - s)}{\varepsilon_{ox}a + d - s} \tag{5.32}$$

There are however a number of situations in which this model just fails to provide an accurate account of the interaction strength. A point in case is for example a charge on the interface of the base electrode and the dielectric layer. The Ludeke – Lambert model predicts a nonzero tip-charge interaction, while obviously it has to be zero, as the charge coincides with its mirror then. Especially for thin oxides the predictions of this model are quite suspect; an interface charge gives a significant value for the 'gradient minimizing voltage' (more on this later), which should be zero, and increasing by a mere 10% for a surface charge. In section 6.6 the problems with this Ludeke-Lambert model will be addressed in more detail.

5.4 THE MULTI MIRROR MODEL

In principle, the electrostatic problem in Fig. 5.5 could be solved by using the method of image charges. This method can be found in any textbook on electrostatics with the classic examples of a point charge in the presence of an infinite, flat, conducting surface and a conducting sphere. With appropriate modifications this method is also applicable for a charge facing a flat surface of a dielectric. These three standard situations are in fact the ingredients for the calculation method presented in this section.
If the charge faces two mirrors instead of one, the total number of image charges is infinite, except in some exceptional cases. The second face not only mirrors the original charge, but also the image of the charge in the first mirror. The first mirror picks up all images in the second mirror, etcetera. This is still known from electrostatics [21], but this infinite mirroring between tip and sample is newly applied to AFM modeling here. The tip apex is usually less than one tip radius spaced from the sample. Neglect of the infinite mirroring results in a sometimes dramatic underestimated prediction of the tip-sample interaction.
In the present configuration (fig. 5.5) there are in fact three mirroring surfaces present: The dielectric surface, the metal/dielectric interface and the AFM tip. This complicates matters considerably. Nevertheless, this method of images proves viable for calculating the electrostatic problem, as will be demonstrated in this chapter.

5.4.1 Imaging in tip

For this chapter, it was chosen to not display the complete build-up of the principles of the multi-mirror calculations all the way from its fundamentals. The reader is referred to Appendix 1 for the standard image charge situations and the mechanism of infinite reflections. It is demonstrated there that Poisson's equation is satisfied in all relevant regions in this case.
This section starts with the important assumption that the tip of the AFM can be aptly modeled by a conducting sphere. This is expected appropriate when the tip spacing is clearly smaller than the radius of curvature. The electric field is then essentially confined between the sample and the lower part of the tip. Related quantities like energy density, capacitance and force are well addressed.
For large spacings or sharp tips, the 'rest of the complete' tip gains influence with respect to the mere tip apex. In these cases, the sphere approximation is expected to underestimate the total interaction strength. In practice, further deviations are possible if the tip apex is not well spherical, but has a certain roughness.
The evaluation of the electrostatic interaction starts with the determination of the point charge in the dielectric and the location of all the mirrors and reflections as described by Eq. (A10) (see Appendix 1). The equations for values and positions of these charges concern those for an observer in the vacuum region and a trapped charge *inside* (or on) the dielectric.
The bias voltage on the spherical tip is represented by a charge at its centre. In this case, the equations for a charge floating *above* the sample apply. First, there is the (important) direct mirror in the surface of the *dielectric* layer. Furthermore, this charge results in two chains of virtual charges: one resulting from the original charge and one resulting from the image of the tip bias charge in the *metal*, i.e. the dielectric/metal – interface.
All charges outside the tip, real and virtual, cause images in this tip.

These "in-tip"-images can be grouped in two classes:

1) The image of the trapped charge, the images of all its infinite reflections and the images of the mirror of the charge in the metal/dielectric interface and its infinitely many reflections.
2) The image in the tip of the mirror in the dielectric layer of the original bias charge, and the images of all reflections of the bias charge caused by the two interfaces metal/dielectric and dielectric/vacuum (the surface of the dielectric), and the images of the mirror of the original bias charge in the metal, plus its corresponding infinitely many reflections.

At this point a first order approximation of the electrostatic potential in all regions can be evaluated, by summing the partial potentials of all individual real and virtual charges. It is stressed once more that for the calculation of the potential, it is essential to realize in what region the potential is calculated, i.e. where the observer is. In vacuum, a certain collection of charges with certain values is seen. Inside the dielectric, the apparent strength of those charges is different, and there are even charges felt *not* seen by the vacuum observer, namely those who (to the dielectric observer) appear to be in the vacuum region.
The force on the tip is principally just calculated by calculating the interaction of an "in-tip" charge (either real or virtual) with each of the "out-tip"- charges: the trapped charge with all its reflections, and the mirrors (in the sample) plus reflections of the original bias charge. This procedure is repeated for all "in-tip" charges after which all individual contributions are summed. Internal interactions between in-tip charges only are not considered: they do not contribute to interaction on the tip as a whole.
Only at this point, this calculation can be said to be 'complete' because a) all three basic imaging[29] principles have been addressed, and thus b) the tip interacts with the trapped charge even in absence of a bias charge, because the trapped charge is imaged in the tip.
However, technically this potential does not fulfill the boundary conditions: All newly assigned image charges in the tip should again have their (not yet described) counterparts in the sample. A better approximation is obtained in a cyclic manner. In the first cycle, the tip, with its collection of bias and image charges, is mirrored in the dielectric (Eq. (A9)). Furthermore, the tip itself sews chains as described by Eqs. (A10) and (A13). The dielectric mirrors and these chains are again mopped up by the tip according to the principle of spherical imaging as in situation 2). At this point the first cycle stops. The next cycle starts again by first reflecting the tip in the dielectric and then forming the multitude of charge chains and reaping them again with the tip. With subsequent cycles, a better approximation is obtained.

5.4.2 Force and gradient calculation; convergence

Like described above, the force on the tip is calculated by grouping all charges in the tip and let each of them interact with each charge in the 'outside world', which includes the trapped charge in the dielectric, all direct mirrors (in metal and dielectric) and all primary, secondary, tertiary etc. chains, mirrors of these chains, and so on. The z-gradient

[29] For the sake of distinguishing all groups of virtual charges, the following terminology will be maintained as good as possible: The image of a charge in a flat surface (the surface of the dielectric or the dielectric/metal interface) is called "mirror", and an image in the tip will be called "image". Also, if the context allows, mirroring in the 'dielectric' and 'metal' will be used as brief expressions for 'dielectric surface' and 'dielectric/metal interface'.

(the relevant AFM probing quantity) can be well determined by calculating the force for a certain tip spacing and repeating the process for the tip displaced a small vertical distance.
In a practical calculation, the infinite chains of reflections must be truncated, as well as the number of 'sew-and-reap' cycles[30]. But still then, one very quickly ends up with tens of thousands of charges, having hundreds of millions of interactions or many more. With a few tricks, these calculations are doable, though still considerably time-consuming.
First, it appeared that the length of the semi-infinite charge chains turned out to be of little importance. An illustration for this could be the following: Replace the chain of infinite reflected charges by a single continuous line charge, with the charges spread out over regions of width 4*d*. This line charge will take up an exponential form. For thin oxides, this is not too bad an approximation and in certain configurations it actually works quite well. Although the single charges do not cancel with those resulting from a mirrored source, these line charges almost completely do. Because the semi-infinite line charge of the original source charge has its head shifted 2d with respect to the line of the mirrored source charge, this 'overhang' of length 2*d* is not cancelled. But 'because' the tails do cancel, the length of the chains is not so important.
Hence, in the calculations this chain length is set to 1 (the effect of varying chain lengths has been studied); only in the last step of sewing out the charges from the tip to the sample, these chains are given lengths of typically 2-5 charges.
The number of repeat steps of sewing the charges and reaping them with the tip lets the number of charges very quickly get out of hand. A typical situation starts to settle after 8 repeat steps, which leads to impractical calculation times; this gives the force, and the same amount of time is required for the gradient, which is determined differentially (symbolic calculations are out of the question as the dependence of the magnitudes of the higher order charges on the coordinate gets impossibly complex). However, as one would expect, the interactions for successive amounts of repeat steps have an exponential trend and are well fittable as such; especially if one leaves out the first two steps. For 12 cycles this is demonstrated in Figs 6ab, which has been done for the a $\varepsilon = 9.1$ dielectric of 25 nm, with a negative unit charge 2 nm below the surface, probed by a 100 nm radius tip at -0.5V 15 nm above the surface. Interactions calculated with repeat steps 3-6 and fitting a trend for repeat-number towards infinity provides a quite reasonable estimate for the ultimate interaction.

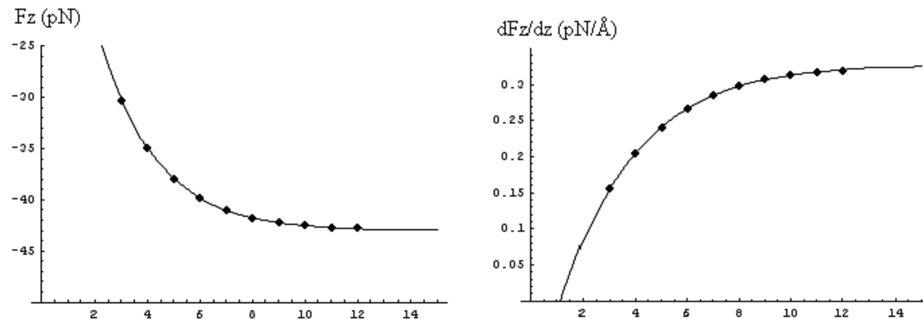


Fig. 5.6 .(left): Convergence of force vs. number of mirror cycles
Fig. 5.6b (right): Convergence of force.gradient vs. number of mirror cycles

[30] Without truncation, a lower bound estimate is $(\infty^{\infty})^2$ interactions, whose evaluation and summation would certainly be beyond 'impractical'.

5.4.3 Field line plots, electrostatic repulsion

In many cases, drawings of electric field lines in non-trivial electrostatic problems can be made only by 'educated guesses' or finite element simulations. In this particular problem the actual solution can be approximated arbitrary closely by adding more and more appropriate point charges. This means that everywhere the field can be calculated exactly for a given charge configuration. Multimirror potential contour plots, combined with some field lines, are given in Figs. 5.7. Of course, these plots hold no particular surprises, but they help to visualize the electrostatics.

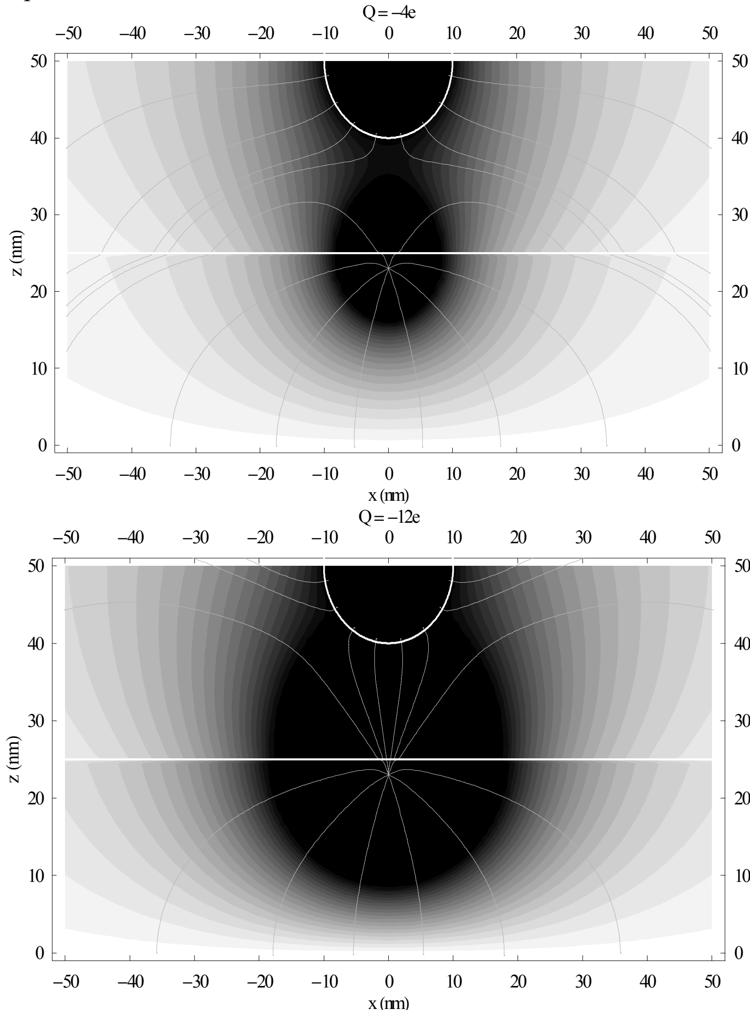


Fig. 5.7 Electrostatic potential and field plots for a 10 nm radius tip (white circle) spaced 15 nm above a 25 nm thick dielectric (below white horizontal line) with ε = 9.1. In the top figure there is a charge of -4e 2 nm below the surface of the dielectric, with the tip at -0.1V relative to the conductor (at z < 0). In this particular case, the tip experiences net repulsion. In the lower diagram a more intense charge (-12e) is present. The charge has a stronger image in the tip here. The attractive contributions outweigh the repulsive contributions.

5.4.4 Electrostatic repulsion

Most often, the net electrostatic force on a biased tip is attractive towards the sample. This is always the case in absence of a trapped charge and if the charge and the bias potential are counterpolar. A net repulsive force is only possible if the trapped charge and the bias charge in the tip have the same polarity. This repulsive force then has to outweigh two attractive contributions: the tip with the substrate (the normal capacitive attraction) and the trapped charge with its image in the tip. The balance for net repulsion can be delicate. The geometric parameters (dielectric constant ε, dielectric thickness d, tip radius R, tip spacing a (or the tip position $z_t = R + a + d$), lateral tip position x_t and charge position s) determine whether repulsion is possible at all. If this is the case, it can be reasoned that the intensity of the charge cluster must be between certain bounds. If it is too small, the attractive V^2 – interaction will dominate the repulsive qV – interaction. If it is too large, the trapped charge attracts its image in the tip too strongly: the q^2 – interaction dominates. To be more specific, consider the force F_{el} (cf. Eq. 5.24). The geometrical parameters α_{VV} and α_{qq} are always negative, while α_{qV} is always positive. F_{el} can have a positive value if and only if the geometrical condition $\alpha_{qV} > 4\alpha_{VV}\,\alpha_{qq}$ is met. Repulsion will then occur if the ratio of q and V is in a limited range of values:

$$\frac{-\alpha_{qV} + \sqrt{D}}{2\alpha_{qq}} < \frac{q}{V} < \frac{-\alpha_{qV} - \sqrt{D}}{2\alpha_{qq}}, \qquad D \equiv \alpha_{qV}^{2} - 4\alpha_{VV}\alpha_{qq} \tag{5.33}$$

The repulsive regime for two tip radii and a particular geometry otherwise is illustrated in Fig. 5.8.

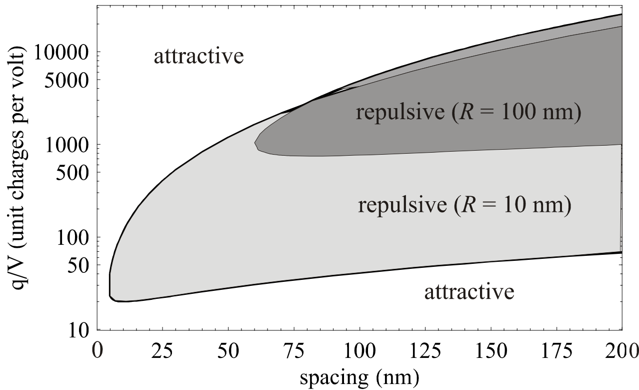


Fig. 5.8 Intensity of one charge cluster required to give a repulsive force on the tip, as a function of tip spacing. The charge cluster is located at the surface of a 25 nm thick dielectric with ε = 9.1. The tip, radius R = 10 or R =100 nm, is in the zenith of the trapped charge. The vertical axis is given here in the number of charges per volt. Here this does not denote a capacitance, as charge and voltage are independent, free parameters. The upper limit of the R = 10 nm region (light grey) crosses the upper limit of the R = 100 nm (dark grey) at a spacing of about 80 nm. The latter region is hence not totally encompassed by the former.

The first thing to learn from Fig. 5.8 is that below a certain spacing ($a \sim 0.5R$, here 4.5 and 60 nm respectively), the force on the tip is always attractive. A more remote tip enhances the chance to have repulsion, as well as a smaller tip. In general, a large z_t / R ratio[31] enhances the chance for repulsion. Another conclusion is that, given a tip radius and a bias voltage, there is an absolute minimum for the amount of trapped charge in order to have repulsion. For the R = 100 nm tip, this occurs at 81 nanometer spacing and requires 750 unit charges in the cluster. For a R = 10 nm tip, repulsion can occur for much smaller charges and smaller spacings. The chance for repulsion increases quite a lot, as is seen from Fig. 5.8. Beyond a certain spacing, the lower bound of the repulsive regime will increase (linearly) with spacing. This means that the tip will have three regimes for large charge clusters. At short distances, the dipole-dipole force of the trapped charge dominates. This attractive force drops off with the fourth power of the distance: $F_{el} \propto z^{-4}$. Receding from the surface enters the domain of the trapped charge with the bias charge (dipole-monopole). This repulsive force is proportional to z^{-3} and dominates the former. At very remote distances, the relatively weak but long range bias-bias (monopole-monopole, $F_{el} \propto z^{-2}$) interaction of the tip with the substrate itself dominates.
Repulsion is more likely to be possible for small tips. As remarked on the basis of Fig. 5.7, only a handful of unit charges suffice to repel a weakly biased tip of only 10 nm radius. Finally, more factors generally enhancing the chance of tip repulsion are thick dielectrics, with the charge as much as possible on the surface. A high dielectric constant shields the trapped charge and enhances the bias-bias interaction. Therefore, repulsion will occur easier in case of a low-ε dielectric.

5.4.5 Critical behavior

If the tip of the AFM is brought increasingly closer to the sample, the peak gradient $\Delta F'$ (Eq. 5.28) gets more and more negative, as the repulsion between trapped charge and bias charge is increasingly stronger. For even closer spacings, Multimirror predicts that attractive forces start to rapidly increase; one of them is the enhanced attraction between the trapped charge and its image in the tip, as remarked in subsection 5.4.4. It is thought useful to stress that certain parameters can suddenly change at close spacings, even in *sign*. Fig. 5.9 shows how the peak force (total force minus bias force) and the peak gradient critically change for a cluster of 25 charges at the surface of a 25 nm dielectric.

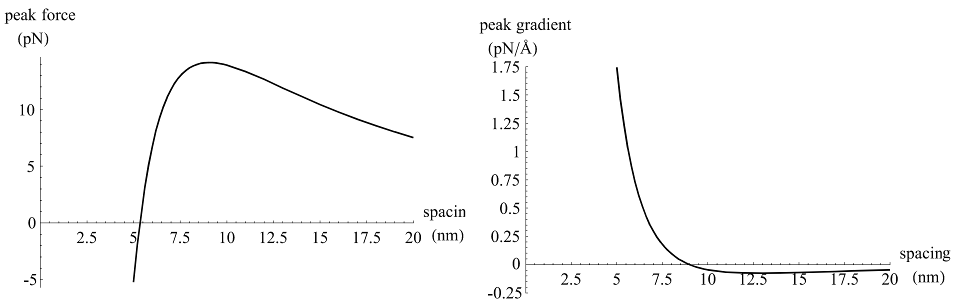


Fig. 5.9 Peak force and peak gradient $\Delta F'$ at close spacings for 25 unit charges at the surface of a 25 nm dielectric (ε = 9.1). Tip radius is 100 nm.

[31] See Appendix Eqs. (A66) and (A96), where is defined $p = 2z_t / R$.

The total force is attractive. But below a spacing 5.4 nm, the peak force itself is attractive (negative) too. The interaction of the trapped charge with its image in the tip is dominant. As stated in the previous subsection, it is short-ranged. At close spacings, it hence dominates the trapped charge – bias interaction and the bias-bias interaction, as these are of longer range. Above 5.4 nm, the repulsive charge-bias interaction takes over. The force reaches a maximum at 9 nm (14 pN). At larger spacings, also the repulsive force weakens. There is one more interesting tip spacing, as we have seen in the previous subsection. Above 60 nm, there exist combinations of trapped charge and bias voltage that make the force as a whole (including bias-bias interactions) repulsive. If we look at the peak gradient: this is the vertical derivative of the peak force. For close spacings it is positive. It crosses the axis where the force has a maximum, namely at 9 nm. It has a minimum at 13 nm spacing and approaches zero for even higher spacings. Finally, V_{min} behaves uniform over the complete range of spacings. It is highest at close spacings and disappears relatively slowly for higher spacings.

5.4.6 Resolution depth and magnitude

5.4.6.1 Bias variation

Electrostatically varied measurements

A measurement is aimed to yield magnitude and depth of a localized charge. As remarked, a typical experiment returns a ‘peak height’ and an offset bias voltage which minimizes the electrostatic force gradient when the tip is above the charge. In the multi mirror model the depth is not readily extractable from the parabolic coefficients because of their very complicated dependence on it. In principle however it is possible to deduce magnitude and depth of the charge graphically from the following graphs (5.10ab):

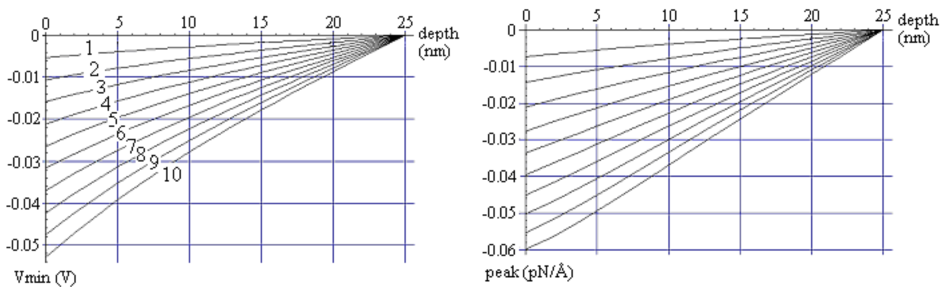


Fig. 5.10a (left) V_{min} as a function of depth of a charge
Fig. 5.10b (right) Peak value $\Delta F'(V_{bias} = -0.5\ V)$ for $q = 1 - 10$ electrons

The dependence of V_{min} and peak height on the depth of a negative unit charge are shown; the continuous looking lines are in fact discrete simulation points (depth steps 1 nanometer) connected by straight line pieces. This dependence is shown for 1 – 10 negative unit charges; in both cases the uppermost lines represents a unit charge. A measurement yields a certain value for V_{min}. We observe from the top graph that for $V_{min} = -0.03$ V only the lines for 6 – 10 charges come into consideration. In the first place this puts a lower limit for the amount of charge that has been observed, namely 6. In the second place, the approved traces are crossed by the –0.06 V line at certain depths. These

depths can be taken over in the lower graph, giving rise to a set of values of peak heights. Ideally, exactly one of these peak height values is equal to the measured value for this parameter. If there is no match, other types of interaction must have been in play, which can partly be dealt with. This topic will be left rested in this text. But even in the ideal case there is a very disturbing complication, namely that the set of peak values will be very closely packed, i.e. there is practically no spread in these values. The following three-dimensional plot seeks to clarify this somewhat more:

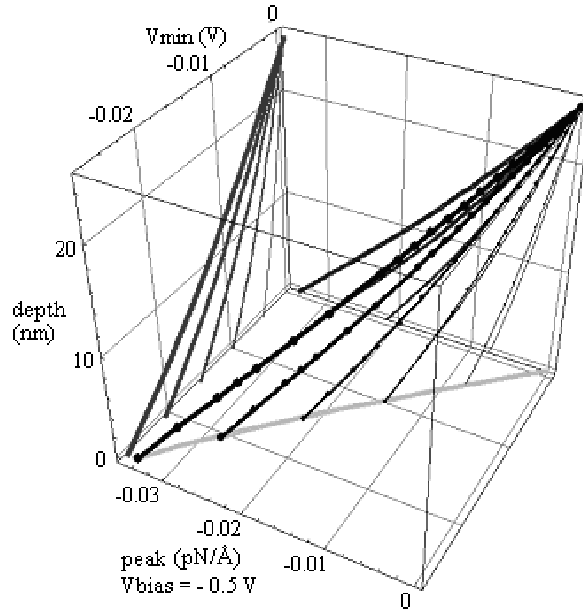


Fig 5.11 V_{min} and peak (ΔF') vs. depth.

For 1 – 5 charges (thin – thick lines) the depth of the charge (vertical axis) in the oxide is plotted against the peak height and the minimizing bias voltage. 'High in the cube' represents 'great depth', corresponding a charge practically 'at the bottom' of the dielectric, i.e. close to the metal/dielectric interface. In this figure, the projections of these lines on the three faces of the 'cube' are also given. Moving upwards from a certain point in the diagonal ground plane 'line' (actually a set of lines) lets one cross several graphs, representing different amounts of charge, at different depths. Two important remarks apply. Firstly, the vast majority of combinations of peak height and V_{min} are theoretically excluded, or claimed impossible to encounter in practice. Most probably this practice will not be so compliant and yield a point outside this ground plane line. In this case, one needs more information about the extra interactions. Secondly, more important is the observation that this ground plane line is in fact not a single line, but a bundle, as will emerge from this plot from the ground plane projection, see Fig. 5.12.

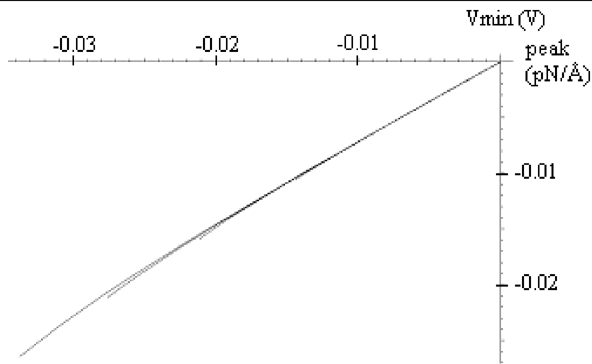


Fig. 5.12 Ground plane projection of Fig. 5.11

The indistinguishability of the multiple(!) lines in this ground plane renders the task of unambiguously deducing values for charge magnitude and depth from the experimental parameters V_{min} and peak height impossible. A small charge close to the surface is equivalent to a big charge at greater depth. At best a lower limit for the charge magnitude can be given. This ambiguity is partly resolved if one works with small bias values.

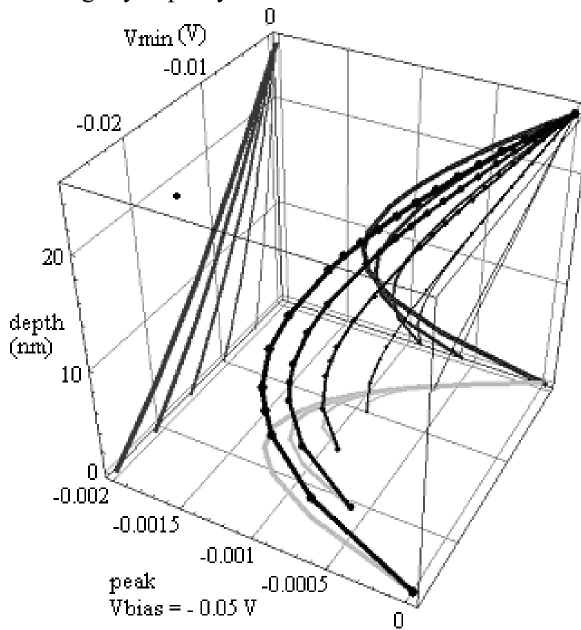


Fig. 5.13 Low bias probing

Quite surprisingly, one and the same charge at different depths can give the same peak height, provided that the bias voltage is relatively small (though nonzero), as in Fig. 5.13, where the bias voltage is reduced by a factor of 10 to 0.05V. A measured peak height of 0.001 pN/Å crosses the back plane projection (blue) of the q = 5e curve (thick) at depths of roughly 4 nm as well as 15 nm. Also the q = 4e and q = 3e curves are intersected twice. Of course, the total interaction is different for both intersection and for the V_{min} parameter there is no ambiguity. We do observe that in the case of a large negative V_{min} the ambiguity disappears (spread in ground plane lines) and only remains in case a small

negative V_{min} is observed. It has to be borne in mind however that reducing V_{min} scales down the peak height signal by roughly the same factor (neglecting the cq^2 term).
The background of this indistinguishability is that one is essentially looking at a dipole[32], consisting of the trapped charge and its image in the metal/dielectric interface. If looked from a large distance at a dipole $p = q \cdot 2s$, with q the charge and $2s$ the separation (the factor of 2 is introduced to maintain s as the distance metal interface – charge), its appearance is the same when q is increased by some factor and s decreased by that same factor.
Several other attempts to resolve this issue of indistinguishability have been tested. A reduced spacing (10 nm instead of 15 nm) yielded no improvement at all, though the last word has not yet been said; namely, for very close spacings (e.g. 5 nm, which is not unrealistic) critical behavior in the measured signals can occur, as is displayed in subsection 5.4.5. On intuitive grounds, it was suggested that the dominance of the charge-bias interaction over the charge – image in tip interaction might cause these interaction profiles to be so closely spaced. Reducing the tip radius by a factor of 10 to 10 nm reduces the bias voltage charge in the tip centre also by a factor of 10, in order to scale down the big charge cluster in the tip centre. Again, this yielded no gain in spread. This can be explained that reducing the tip radius while keeping the spacing constant helps nothing in conceding relative interaction strength in favor of the charge-image in tip interaction. The interaction distance between charge and bias voltage centre decreased considerably; as the force depends quadratically on this distance, this interaction in fact *grows* with respect to the charge-image interaction. Consequently a larger tip should also be considered; which is however experimentally not very realistic and undesirable for different reasons.

5.4.6.2 Geometrically varied measurements

The above described methods are al based on performing a couple of on-charge bias-varied measurements, perhaps compared to off-charge measurements. As these do not seem to provide detailed information about charge magnitude and depth, different methods have to be investigated. Adding geometry (tip position) as a variable parameter has the potency of helping to resolve the issue of 'degenerated' (ambiguous) solutions. Two methods come into consideration:

* Interaction profiles in the horizontal directions
* Variation in tip spacing

1) Horizontal scan

The most natural approach, as this is what is more or less automatically done in conducting AFM measurement practice. A charge on the surface is expected to give rise to a relatively sharply peaked profile. However, this is seems only marginally the case. For a 100 nm tip spaced at 15 nm from a 25 nm thick oxide it is assumed that a peak height of -0.015 pN/Å has been measured. Fig. 5.14 shows for 1 – 5 charges how the peak height depends on their depth. It is clear that a single charge could not have yielded this peak, but 2 and any higher number could have.

[32] It is also this dipole character that is responsible for humps that can exist next to the central peak of an interaction profile in horizontal scanning direction.

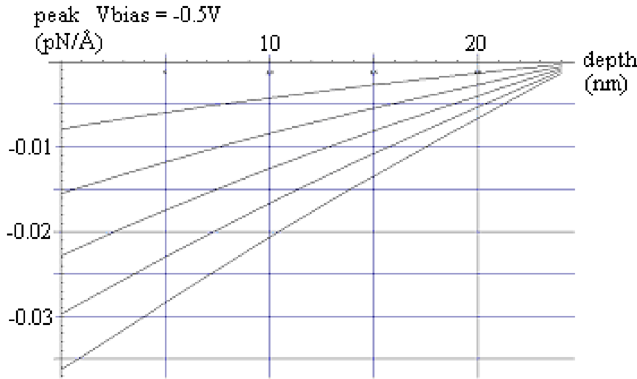


Fig. 5.14 Peak height dependency on depth

The x-profiles of a 2-charge cluster at a depth of 0.7 nm and a 5-charge cluster at a depth of 13.8 nm are compared in Fig. 5.15. These charge-depth combinations should yield the same peak height.

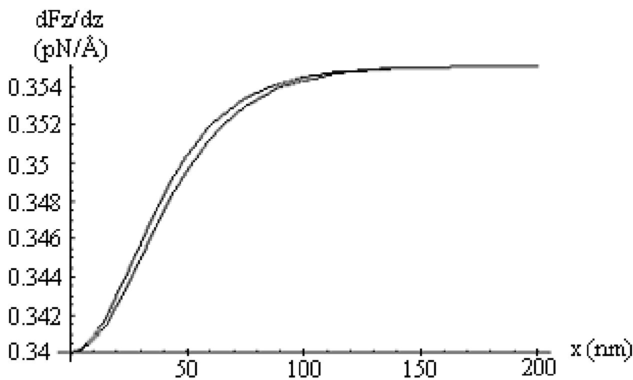


Fig. 5.15. x-profile comparison of charge clusters; the upper line corresponds to the more superficial charge

It transpires that the spatial peak *width* is only marginally different for these clusters. An explanation could be the following: For the above applied bias, the charge-bias interaction dominates over the charge-tip image interaction. Because of the large tip-radius and hence the long distance of this interaction (> 100 nm), the difference of the z-position of the charge clusters (~10 nm) is hard to notice.

Narrower tips however offer increased possibility of resolving depth and magnitude. In Figs. 5.16 and 5.17 interaction profiles are shown for 10 nm and 1 nm narrow tips, brought at close spacings in order to increase resolution power. The profiles are normalized for pure resolution comparison. The charge is at depths 0, 5, 10, 15 and 20 nm below the surface.

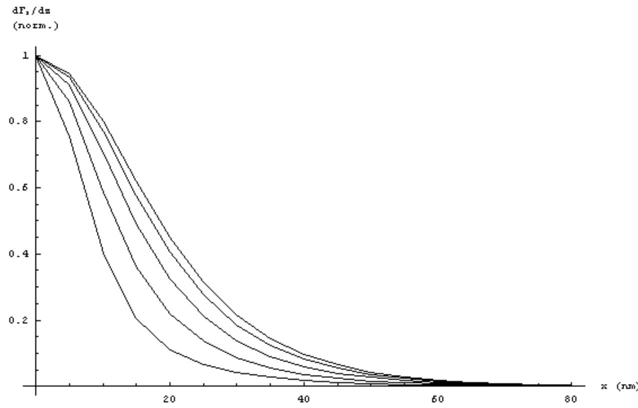


Fig. 5.16 R = 10 nm; *spacing = 5 nm*

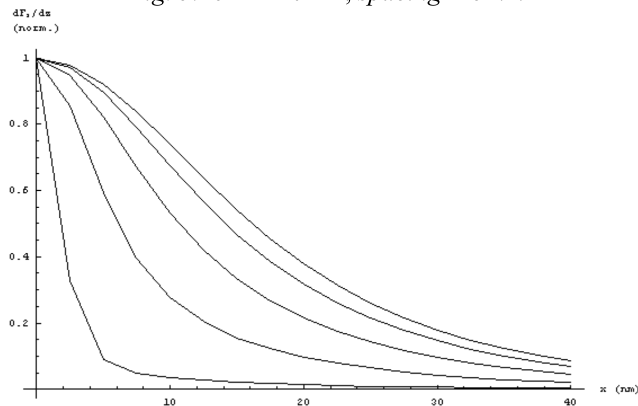


Fig. 5.17 *R = 1 nm; spacing = 2 nm.*

It has to be remarked that for these narrow tips, the validity of the spherical tip assumption is increasingly open to question. For these graphs however, the precise interaction strength is less relevant. The enhancement of spatial resolution is believed to be still well demonstrable under this assumption even for very sharp tips.

2) Varied tip-spacing

Performing bias varied measurements ('parabolas') at different heights is a second option. Both measurement series provide certain sets of charge/depth combinations that all explain the observed measurement parameters. The two sets should have exactly one element (charge-depth combination) in common. Again, a relatively superficial charge must be more incentive to interaction variation with height than a charge at a deeper position in the dielectric. In Fig. 5.18 the two black lines indicate two possibly measured values (-0.0114 V and -0.0140 V) for V_{min}. The light-grey curves are for a remote tip at 20 nm and the sharp, dark grey curves represent a near tip at 10 nm. Curves are given for 1-5 charges. Intersecting points of the black lines (measured values) with the charge curves provide several possible charge-depth combinations.

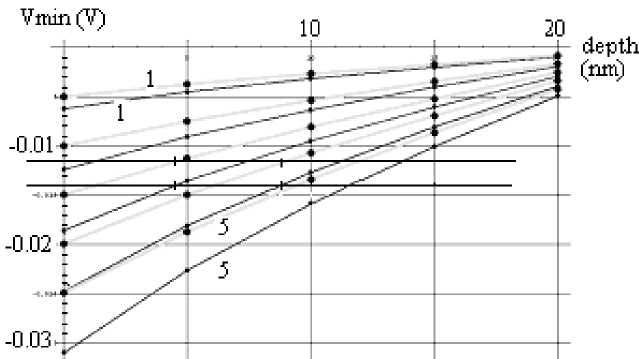


Fig. 5.18 Tip spacing (10 and 20) varied depth dependence of Vmin

Of course both tip spacings must point to equal amounts of charge, yielding numerous ‘intersection pairs’. The members of these pairs must lie above each other, for the depths must match. In the above graph this boils down to two candidate solutions of (q = –3e, depth = 4.5 nm) and (q = –4e, depth = 8.8 nm), though there may be more for higher amounts of charge. Additional information is retrieved from the peak heights belonging to this measurement. Though possibly mathematically more elegant, in practice this method has the disadvantage of being more cumbersome.

5.5 FINITE ELEMENT MODELING

Finite Element Modeling (FEM) is a powerful and versatile tool for performing calculations for a great variety of physical problems. It is especially applied in situations that are too complex to handle for analytical calculations. Typical problems consist of flow patterns of a gas or liquid through complex channels, the distribution of heat in cooling fins, deformation of materials under pressure, the magnetic field of currents, etcetera. FEM has become indispensable for contemporary engineering.

Concerning the research in this chapter, FEM is used for two purposes. Firstly for comparison of the outcome of the here treated analytical models and secondly to estimate the applicability of the spherical tip approximation used by these models, by supplementing the sphere with a conical part. Concerning the first application, it is remarked that the main results are presented in section 5.6 (model comparison).

5.5.1 Principles of Finite Element Modeling

In principle, any well-defined physical geometry can be implemented in Finite Element Modeling. Relevant material constants can be specified and appropriate boundary conditions. Once a certain problem is well-specified, the geometry is subdivided in elementary units, called ‘mesh elements’. Usually, for 2D geometries triangles are used for they are best suited to approximate any particular geometry[33]. Likewise, a 3D geometry is commonly meshed with tetrahedrical elements.

[33] For highly symmetric geometries, or parts of it, preference may be given to rectangles.

Once a mesh is defined, a solving algorithm attempts to assign a value for one or more relevant, 'leading' quantities to every point in space. In heat conduction through some material for example, this means specifying the temperature everywhere.
The solving entails numerical coupling of the mesh elements to each other to form an overall solution. The accuracy of the solution increases with the number of mesh elements. Sharp edges require relatively many and small elements in order to optimally approximate the local geometry. Also, the values of relevant physical quantities can vary a lot over short distances (have large gradients). More complex geometries demand more mesh elements. Needless to say, more mesh elements has to be paid with increased calculation times and is bounded by the internal memory of the system the simulation is running on.

A global solution can be plotted (graphically represented) in various ways. Also, from this solution, more quantities can be derived like temperature gradients, stress or forces. Even time dependent simulations are possible, in which parts move under influence by forces or in which potentials vary over time.
Here, COMSOL® Multiphysics[34] 2.3 was used for FEM of an electrostatic problem, which entails solving Possion's equation for all mesh elements, after which the electric potential of every point in space is specified, and with it the electric field (which is the negative gradient of the potential). An example of a potential contour plot and a meshing is given in Fig. 5.19:

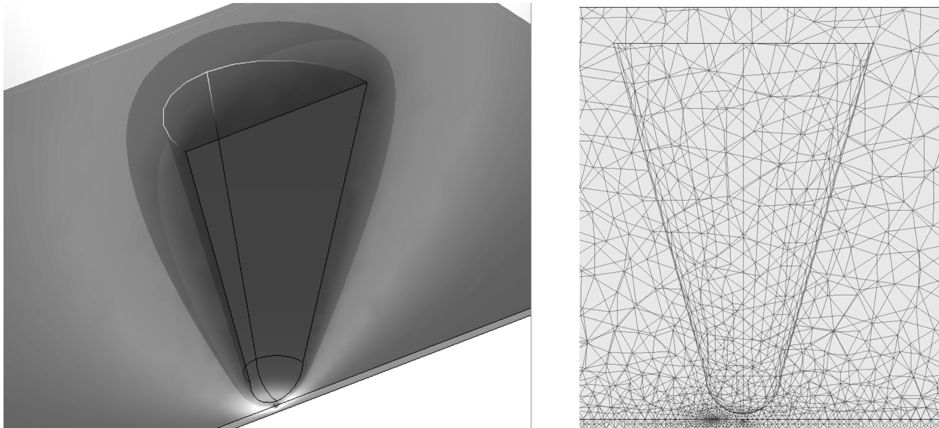


Fig. 5.19 (left) Contour slice plot of the potential. Shown here is a conical tip with apex radius of 100 nm radius tip (at bias -0.5V), spaced 15 nm above the surface (thin horizontal line) of a 25 nm thick dielectric (ε_r = 9.1). The small pointlike feature underneath is the potential around a (negative unit) point charge. The granularity of the mesh elements is visible at greater zoom only. The singular nature of the point charge is dealt with by making it to be the corner of a mesh element. (right) Mesh elements of conical tip and the dielectric with a localized charge.

[34] Formerly known as 'Femlab®'

5.5.2 Determination of force gradient

The information requested from these simulations is the force and the force gradient on the tip. The force on the tip can be determined in two ways. The first is to integrate the electrostatic energy density over the vacuum (bounded by a box much larger than the actual configuration). When this is done for a few different, closely spaced tip heights and the energy is plotted against height, a force is obtained. For the configuration described underneath Fig. 5.19, meshed with approximately 550.000 elements, the total energy is shown in Fig. 5.20. The message conveyed by these three plots is that the simulated vacuum energy has a very stable dependence on the tip spacing, while the dielectric energy clearly has not. Note that the differences are quite small, a hundredth of an attoJoule (= 0.06 eV) or less; a few percent in the dielectric region. The behavior of the energy in the dielectric region is alone responsible for the irregularities in the total energy (Fig. 5.22). Incidentally, it is the dielectric region that contains the singularity of a point charge. Still, a force can be reasonably determined in this case, but a gradient in the force (second derivative of $U(z)$) is not feasible. An additional difficulty is that the drawn dielectric has an enormous aspect ratio: 25 nm thick and several microns in length, and a charge-surface distance of only 2 nm. This is however mainly noticeable through the large number of mesh elements required.

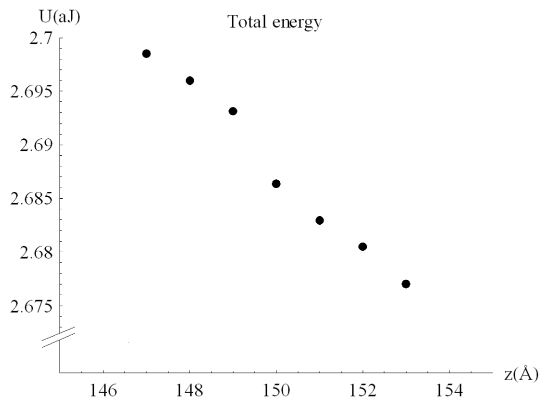


Fig. 5.20 Total electrostatic energy (atto-Joule) vs. tip spacing (Ångström).

It is interesting to add the results of the partial energies of the vacuum region (Fig. 5.21) and the dielectric region (Fig. 5.22).

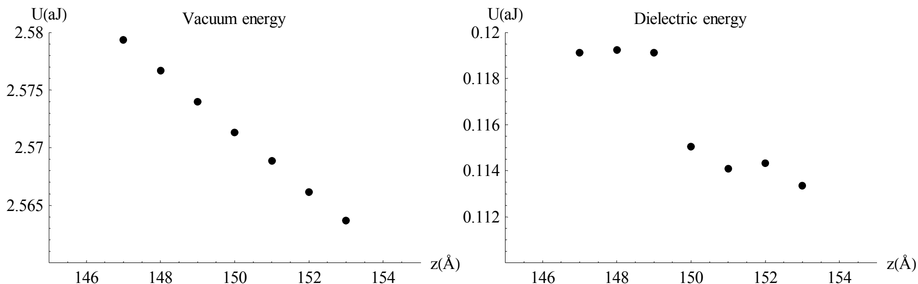


Fig. 5.21 Electrostatic energy of the vacuum region *Fig. 5.22 Electrostatic energy of the dielectric region*

Meanwhile, the force can be determined more directly by integration of the Maxwell stress tensor over the surface of the spherical tip. This can be done for a couple of tip spacings around the 'working point' (= the actual static tip spacing). A linear fit through these points yield the force and the gradient (Fig. 5.23).

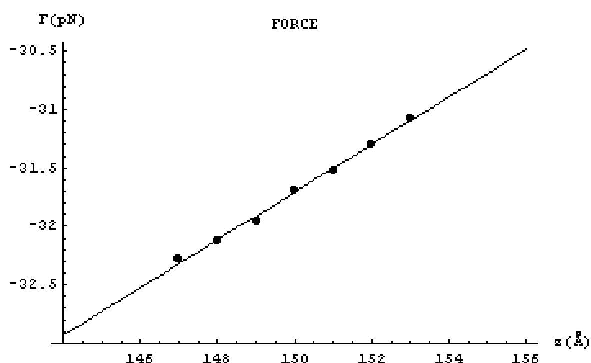
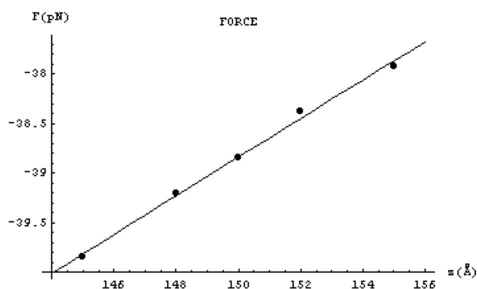


Fig. 5.23 Directly determined force vs. tip spacing

The force points are reasonably fittable, yielding a force of –31.71 pN and a slope (gradient) of 2039 µN/m. Determining the slope between two successive force points yields values spreading between 1569 µN/m and 2712 µN/m, which at least demonstrates that steps of 1 Å lead to unstable results. More and wider spread points are needed. At least five points are recommended, spread out over a not too narrow region in which the force can still be considered to be linearly dependent on the spacing; here typically about 1 nanometer.

5.5.3 Conical tip

The top half of the spherical tip was replaced by a cone-shaped piece of metal of 2 µm high. This produced the results presented below (Figs. 5.24-5.26).



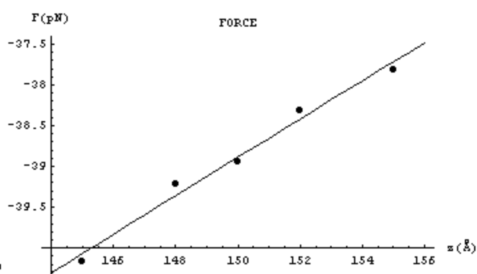


Fig.5.24: Force vs. separation for cone-shaped tip Bias –0.5 V, charge 2 nm underneath the surface. F = –38.8 pN; dF/dz = 1945 µN/m
Fig.5.25: Force vs. separation for cone-shaped tip Bias –0.5 V, no charge (bias background); F= –38.9 pN; dF/dz = 2347 µN/m

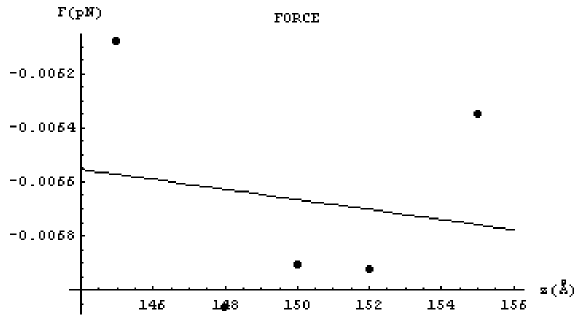


Fig. 5.26 Force vs. separation for cone-tip at V_b=0 V Bias 0 V; charge 2 nm underneath surface (charge – image interaction; F = –0.0067 pN; dF/dz = –0.19 µN/m

Considering the doubts about COMSOLs numerical accuracy, little more can be said about the total force gradients than that in case of the multi model they generally agree, but that quantities like ‘peak height’ cannot reliably be determined by the performed COMSOL simulations. In case of this cone-shaped tip the general value of 1945 µN/m should be a reasonable estimate of the force gradient, but the peak height of –402 µN/m (cf. the bias background) is much open to question. What *is* quite reliably calculated is the total force on the tip, if a bias is applied. Small changes in geometry are clearly reflected in a relatively smooth variation of the force, which for the lion’s share is determined by the strong bias interaction of the tip with the plate electrode. The COMSOL result of –38.8 pN force on the cone tip can therefore be taken with reasonable trust. This would mean that a spherical tip model underestimates the force by 18% compared to a (short) cone-shaped tip, given this configuration. Considering how much extra metal volume is added, this indicates that the sphere does not do too bad a job. But as in reality a tip is much larger than 2 µm high, this percentage should probably be somewhat higher even. How this estimate falls out for closer spacing has not been investigated. And also, how this difference in force works out on the difference in force *gradient,* is unsure.

The problem is firstly that the force data points must be aligned really well because for determining the peak height only coarsely the fit slope must already be accurate to better then 1%. Secondly, these Finite Element simulations have had great difficulty in taking account of the trapped charge.

Concerning the latter problem, an idea is to replace the point charge with a small spherical conductor with the appropriate potential on its boundary. This removes the singularity and encourages the mesh function to generate a higher(!) density around the charge (a singularity becomes a corner of a mesh element, not increasing the mesh density in its neighborhood very much). Concerning both problems: Reduce the problem to 2D by exploiting the rotational symmetry. At first the idea was to make a interaction profile by making the tip ‘scan’ a few points along the x-axis, but in many cases already a lot of information is retrieved by letting the tip float somewhere right above the charge, thus allowing to convert it to a 2D problem.

5.6 MODEL COMPARISON

In this section, the three models / calculation methods Ludeke-Lambert, Multimirror and FEM will be compared for various situations and various parameters. The first subsection tabulates some characteristic results, while the second subsection compares depth profiles, i.e. interaction parameters as a function of the depth of the charge in a 25 nm thick oxide.

5.6.1 Interaction parameters

The comparison addresses three different configurations: The first has an oxide of thickness 25 nm and a dielectric constant of 9.1, the library value for aluminum oxide in COMSOL. The second decreases the dielectric constant to 3.9 (silicon oxide). In both cases a single negative charge is placed 2 nm below the surface. The third situation is again aluminum oxide but reduced in thickness by a factor of 10 to 2.5 nm. Considered is a negative unit charge 0.5 nm below the surface in this case. The geometric ratios in this last case were deemed too disproportional (ultra-thin oxide) for COMSOL to handle properly in an straightforward way. In all situations applies: Tip radius 100 nm, spacing 15 nm, negative unit charge. The quantities F, dF/dz and ΔF’ are taken at V_{bias} = –0.5 V (tip negative with respect to the sample). The depth of the charge measured from the dielectric surface is indicated in the table headers (Table 5.1-3). For each table the capacitance (atto-Farad) as calculated by Multimirror for that situation is given.

Table 5.1 *MM:* C = 22.5 aF ε = 9.1 d = 25 nm. (depth 2.0 nm)				
	F (pN)	F’ (fN/Å)	ΔF’ (fN/Å)	V_{min} (mV)
Ludeke-Lambert	N.A.	8.5	-0.23	-9.4
Multimirror	-36.3	269	-5.0	-4.3
Finite Element	-31.9	206	-7.0	-8.3

Table 5.2 *MM:* C = 21.4 aF ε = 3.9 d = 25 nm. (depth 2.0 nm)				
	F (pN)	F’(fN/Å)	ΔF’ (fN/Å)	V_{min} (mV)
Ludeke-Lambert	N.A.	7.5	-0.26	-21
Multimirror	-28.5	170	-4.4	-7.0
Finite Element	-25.4	127	-16	-28

Table 5.3 *MM*: C = 24.6 aF ε = 9.1 d = 2.5 nm. (depth 0.5 nm)				
	F (pN)	F’ (fN/Å)	ΔF’ (fN/Å)	V_{min} (mV)
Ludeke-Lambert	N.A.	8.8	-0.22	-9.4
Multimirror	-48.2	490	-0.35	-0.18
Finite Element	N.A.	N.A.	N.A.	N.A.

Where applicable, Multimirror is in reasonable till good agreement with FEM for the interaction strengths (F, F’, ΔF’). In all occasions except one, these parameters are very much larger than as predicted by Ludeke-Lambert. V_{min} generally comes out considerably lower for Multimirror, compared with Ludeke-Lambert. The latter model agrees here better with FEM than Multimirror does. The meaningfulness of the latter mutual (dis)agreement is doubtful however. For the configuration of Table 5.1 this point will be clarified in the subsequent subsection (5.6.2). For the charge at slightly greater depth Multimirror in turn agrees much better with FEM than Ludeke-Lambert does.

5.6.2 Dependence on charge depth

It has been chosen to include in the model comparison the dependence of the interaction parameters Δ*F*' (or 'peak') and $V_{min,}$ for a number of reasons. In the first place, resolution of charge magnitude and depth has been one of the topics in section 5.4 devoted to the Multimirror model. Secondly, it illustrates a remark in the previous subsection about the matching of V_{min} between FEM and Ludeke-Lambert. Finally, and most importantly, a fundamental defect of the Ludeke-Lambert model is demonstrated, which is not present in the Multi-mirror model, see Figs. 5.27:

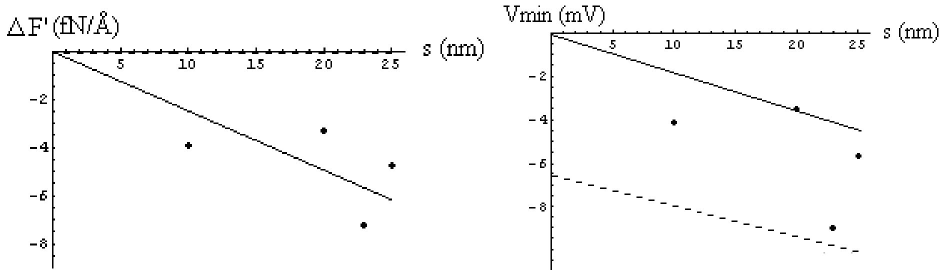


Fig. 5.27 ΔF' and V_{min} vs. z-coordinate of charge (s = d – depth). Dots are derived from FEM simulations, solid line is Multimirror and dashed line is Ludeke-Lambert.

These diagrams give rise to many important observations. Firstly, it transpires that the FEM simulations did not yield a stable and smooth dependence of the interaction parameters as a function of the location of the trapped charge; the data points are rather scattered. As was remarked already in section 6.3, presumably the singular nature of the trapped charge poses some problems here. Secondly, The differences between Multimirror and Ludeke-Lambert are once more shown to be enormous. Concerning the peak height Δ*F*', Multimirror agrees well with FEM. Ludeke-Lambert is hardly present in this diagram. Thirdly, the fact that Ludeke-Lambert agrees much better with FEM for the value of V_{min} at s = 23 nm can be considered rather coincidental, because of the scattering of the latter's data points. At s = 20 nm for example, Multimirror on its turn agrees much better with FEM than Ludeke-Lambert does.

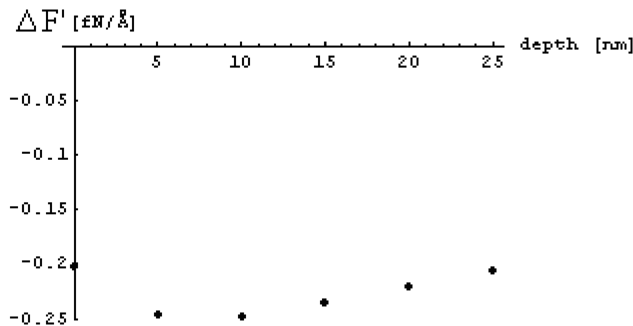


Fig. 5.28 Peak height according to Ludeke-Lambert. Here the horizontal coordinate is the depth of the charg. The value is not changing much over the whole oxide.

The fourth point to establish here concerns the just mentioned defect: according to the Ludeke-Lambert model, the trapped charge maintains a considerable influence, even when at the interface of the metal and the dielectric (s = 0). Actually, theoretically it should be zero. The depth dependence of ΔF' according to the Ludeke-Lambert model is presented in Fig. 6.24. For thinner oxides the predicted dependence becomes even weaker; in the case of a 2.5 nm aluminum oxide, the interaction parameters subside by merely ~10% when a charge is sunk from surface to interface.

5.7 RELATION TO EMPIRICISM

Charge imaging measurements have been performed on Al_2O_3 and SiO_2 by Marko Sturm et al. [15, 16], as described in chapter 5. His work is chosen to endow the multimirror model with empiric relevance in two ways:
Firstly, together with relation (5.12), the model predicts a verifiable value for the frequency shift of the AFM. Secondly, the voltage offset in the total force gradient (depending quadratically on the voltage) is compared with contact potential difference pictures from localized charges. In this way, the two experimental parameters peak height and Vmin are both covered.

Frequency shift

As digressed on in section 5.1, force modulated AFM maps frequency shifts. The models in this chapter provide force gradients. With Eq. (5.11), the gradient is readily translatable into a frequency shift, provided the oscillation amplitude of the cantilever is sufficiently small. However, as force modulated AFM is generally operated using large amplitudes, it is rather the force itself that is important and specifically its profile along the z-axis. Both the multi mirror model and the Ludeke-Lambert model (integration of (5.29-5.31) over z) are fed with a cantilever stiffness of a platinum iridium tip k = 45 N/m (www.nanosensors.com), a resonance frequency f_0 = 75 kHz, a closest approach z_{min} = 5 nm and an amplitude A_0 = 10 nm. Furthermore, oxide thickness 2.5 nm, ε = 9.1 and tip radius R = 100 nm are used. A single negative charge was located on top of the surface, which has no visible influence on the curves below (Fig. 5.29ab):

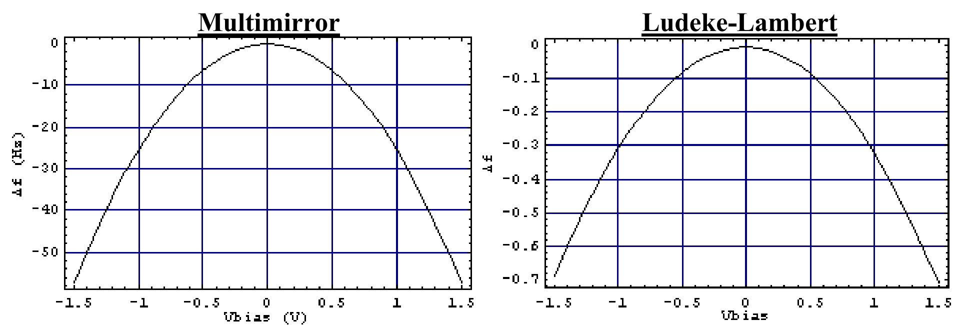


Fig. 5.29a Model comparison on predicted frequency shift *Fig, 5.29b*

A bias dependent frequency shift carried out by Sturm [16] showed:

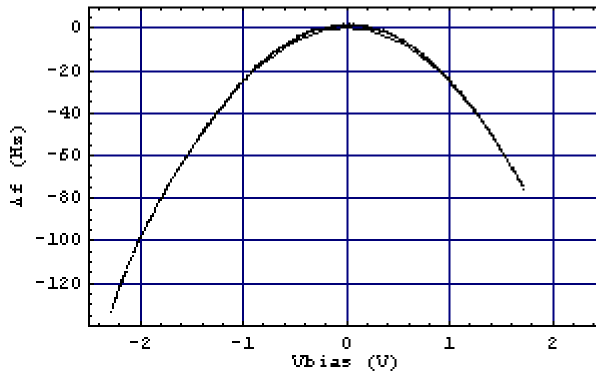


Fig. 5.30 Measurement of shift in resonance frequency of AFM cantilever as a function of tip-sample bias voltage [16] together with the multi mirror prediction

The numerical agreement of the Multimirror calculations with the experiment is clear. It has to be mentioned that the frequency shift depends quite strongly on the parameters, which adds an element of coincidence to the good agreement, but the order of magnitude is at the very least correct, whereas the Ludeke-Lambert model has no chance of explaining the measured frequency shift, due to a serious underestimation of the tip-sample interaction.

Contact potential difference

The CPD, contact potential difference, is defined as the tip-bias voltage that minimizes the electrostatic force gradient between tip and sample. If the sample dielectric contains fixed charges, this gradient is minimal for some nonzero voltage. Even at this voltage a (small) net gradient will remain in general. According to the Multimirror model, this CPD should be in the order of millivolts. This is in contradiction with typical measurements that show CPD's up to 1 or 2 *deci*volt:

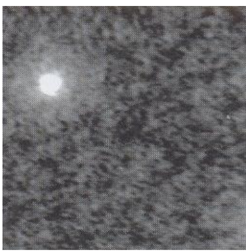


Fig. 5.31 CPD of a charge [16]

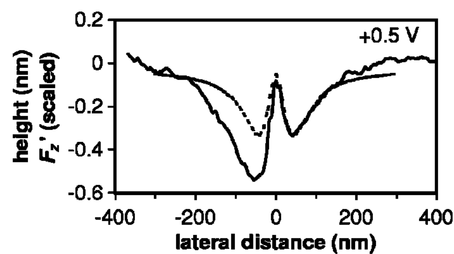
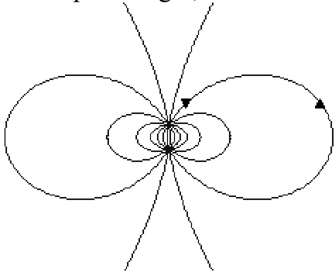


Fig. 5.32 ΔF' scaled

The 'dips' flanking the central peak signal the dipole nature of the interaction, which is especially pronounced for thin oxides, where the trapped charge is close to its mirror in the metal. The electric field E of a pure, solitary, mathematical dipole $\mathbf{p}$ at position x_p is given by:

$$\mathbf{E}(\mathbf{x}) = \frac{3\hat{n}(\mathbf{p}\cdot\hat{n}) - \mathbf{p}}{4\pi\varepsilon_0 \left|\mathbf{x} - \mathbf{x}_p\right|^3}, \tag{5.34}$$

In which $\hat{n}$ is a unit vector from $\mathbf{x}_p$ to $\mathbf{x}$. This field is similar to the pattern shown in Fig. 5.33, where it is plotted for a *physical* dipole (nonzero length). At length scales larger than the dipole length, the field lines are well comparable.



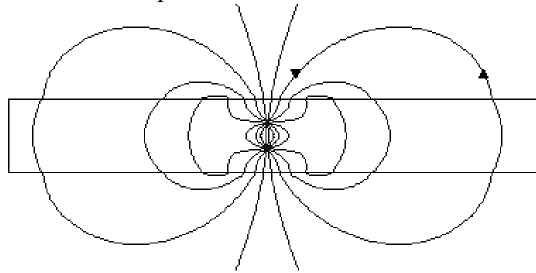


Fig. 5.33 Field of a dipole *Fig.5.34 Field of a dipole in a dielectric slab*

Now, if one approaches the vertical axis parallel to (and off) the horizontal axis, like a charged scanning tip, the vertical component of the electric field charges sign. This means that at first, a test charge would be downward attracted far from the dipole and repelled upward when above the dipole. This behavior is qualitatively conserved if the dipole is wrapped in a dielectric slab, as shown in Fig. 5.34. This can be seen already from Eq. (5.34), as with a horizontally more remote tip, the vector to the dielectric becomes increasingly more disparallel with the dipole vector, which decreases the term containing $\mathbf{p}\cdot\hat{n}$. It should be noted that the interaction between a tip and a charge (cluster) is more complicated, as there is the additional interaction of the charge with its image in the tip. This is always attractive, regardless the relative polarity. Especially for large charge clusters, the latter can become a significant contribution.

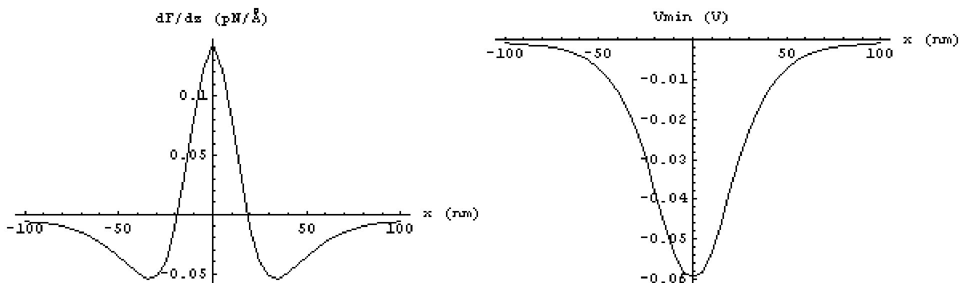


Fig. 5.35, x-profile of force gradient and CPD for native alumina, 2.5 nm thick, of a cluster of 50 electrons 0.5 nm below, probed with a 100 nm tip at V = –0.5 V at 10 nm.

It turns out however, that in the given configuration, the multimirror model can reproduce this profile *only* by assuming sufficient (more than one) charges per cluster. The graph in Fig. 5.35 is for 50 negative unit charges. As $V_{\min}$ is proportional with q, this would restore the CPD from the multi-mirror model to a desired value. This idea of clustered charges may initially be counterintuitive because of their mutual repulsion. The following could be an explanation: Charges get trapped typically in lattice defects. Provided these traps are (energetically) sufficiently deep, the mutual electrostatic repulsion will not cause the charge cluster to fall apart. Large tips will not be able to spatially distinguish two single charges so close together and count them as one charge with double value. Only narrow and closely spaced tips will notice a broadened x-profile.

5.8 CONCLUSIONS AND DISCUSSION

The electrostatic interaction between the tip of a conducting AFM and a dielectric layer containing a localized charge on top of an electrode had been taken account of by a model from Ludeke, amended by Lambert. Presented in this chapter is a model (Multimirror) that incorporates various types of repeated, multiple images and reflections. This and the Ludeke-Lambert model share the assumption that the tip of a conducting AFM can be aptly represented by a conducting sphere hovering above the surface. The validity of this assumption has been confirmed by Finite Element Modeling (FEM) for a tip of radius 100 nm.

Multimirror predicts much higher values for typical interaction parameters, such as total force gradient and peak interaction, than the Ludeke-Lambert model does. Depending on the parameter and the configuration, the difference can go up to about two orders of magnitude. Concerning this difference, Multimirror is supported by FEM in most cases, especially for the total force between tip and substrate. The FEM simulations had some trouble with accounting for the singular nature of the charge, considering the spread of the points in depth dependent studies. For the gradient minimizing voltage, or Contact Potential Difference (CPD), Ludeke-Lambert in turn predicts considerably larger values than Multimirror does. FEM is neither very supportive nor dismissive towards either model.

A more fundamental difference between the models is how parameters such as $\Delta F'$ and $V_{\min}$ change for a charge sinking in the dielectric towards the electrode/dielectric interface. Multimirror predicts these parameters to drop off to zero, while according Ludeke-Lambert not more than only a moderate decrease is to be noticed, especially for ultrathin dielectrics.

For resolving magnitude and depth of a trapped charge, various methods have been proposed. A static tip and varying the bias is not expected to convey any information concerning this. Scanning in horizontal direction over the charge works for sharp tips ($\leq$ 10 nm); only then will shallow charges give rise to significantly narrower interaction peaks than deep charges do.

An additional prediction of the Multimirror model is critical behavior of $\Delta F'$ when the tip approaches the substrate very closely, for certain configurations.

The Multimirror model, together with the frequency shift calculation (5.12), predicts a shift of the AFM cantilever resonance that is in excellent agreement to measurements, while the Ludeke-Lambert model falls short by two orders of magnitude in this case.

For a single charge, Multimirror predicts a Contact Potential Difference (CPD) that is orders of magnitude smaller than Ludeke-Lambert and what is typically measured in practice. The assumption of more unit charges in a single cluster is supported by that only then, Multimirror shows depressions next to the central peak in a x-profile scan. An additional consolation is that a charge density resulting from a 0.4 μm^{-2} cluster density (section 5.2) with only a single unit charge per cluster would be surprisingly low. Though not quite comparable situations, charge densities encountered in practice are usually much higher. Assuming more intense charge clusters resolves this.

All in all, Multimirror can be concluded to account for the parameters characterizing the tip – sample + charge interaction much different than the Ludeke-Lambert model does. Fundamental issues, FEM and empiricism are all supportive towards the former.

REFERENCES

[1] G. Binnig, H. Rohrer, Ch. Gerber, and E. Weibel, "Surface Studies by Scanning Tunneling Microscopy", *Physics Review Letters* **49**, 57 (1982)

[2] G. Binnig, H. Rohrer, Ch. Gerber, and E. Weibel, "7 x 7 Reconstruction on Si(111) Resolved in Real Space", *Physics Review Letters* **50**, 120 (1983)

[3] G. Binnig, C. F. Quate, and Ch. Gerber, "Atomic Force Microscope", *Physics Review Letters* **56**, 930 (1986)

[4] J. Kirtley, M. Ketchen, C. Truei, J. Sun, W. Gallagher, L. S. Yu-Jahnes, A. Gupta, K. Stawiasz, and S. Wind, "Design and Applications of a Scanning Squid Microscope", *IBM Journal of Research and Development* **39**, 655 (1995)

[5] K. H. Kuit, J. R. Kirtley, W. van der Veur, C. G. Molenaar, F. J. G. Roesthuis, A. G. P. Troeman, J. R. Clem, H. Hilgenkamp, H. Rogalla, and J. Flokstra, "Vortex Trapping and Expulsion in Thin-Film $YBa_2Cu_3O_{7-\delta}$ Strips", *Physics Review B* **77**, 134504 (2008)

[6] F. J. Giessibl, and G. Binnig, "Investigation of the (001) Cleavage Plane of Potassium Bromide with an Atomic Force Microscope at 4.2 K in Ultra-High Vacuum", *Ultramicroscopy* **42-44**, 281 (1992)

[7] F. J. Giessibl, "Atomic Resolution of the Silicon (111)-(7x7) Surface by Atomic Force Microscopy", *Science* **267**, 68 (1995)

[8] R. García, and R. Pérez, "Dynamic Atomic Force Microscopy Methods", *Surface Science Reports* **47**, 197 (2002)

[9] F. J. Giessibl, S. Hembacher, H. Bielefeldt, and J. Mannhart, "Subatomic Features on the Silicon (111)-(7x7) Surface Observed by Atomic Force Microscopy", *Science* **289**, 422 (2000)

[10] F. J. Giessibl, "Advances in Atomic Force Microscopy", *Reviews of Modern Physics* **75**, 949 (2003)

[11] F. J. Giessibl, "Forces and Frequency Shifts in Atomic-Resolution Dynamic-Force Microscopy", *Physics Review B* **56**, 16010 (1997)

[12] T. R. Albrecht, P. Grütter, D. Horne, and D. Rugar, "Frequency Modulation detection using high-Q cantilevers for enhanced force microscope sensitivity", *Journal of Applied Physics* **69** 668 (1991)

[13] F.J. Giessibl, H. Bielefeldt, S. Hembacher, and J. Mannhart, "Calculation of the Optimal Imaging Parameters for Frequency Modulation Atomic Force Microscopy", *Applied Surface Science* **140**, 352 (1999)

[14] J. M. Sturm, A. I. Zinine, H. Wormeester, R. G. Bankras, J. Holleman, J. Schmitz, and B. Poelsema, "Laterally Resolved Electrical Characterization of high-*K* Oxides With Non-Contact Atomic Force Microscopy", *Microelectronic Engineering* **80**, 78 (2005)

[15] J. M. Sturm, A.I. Zinine, H. Wormeester, B. Poelsema, R.G. Bankras, J. Holleman, and J. Schmitz, "Imaging of Oxide Charges and Contact Potential Difference Fluctuations in Atomic Layer Deposited Al_2O_3 on Si", *Journal of Applied Physics* **97** 063709 (2005)

[16] J. M. Sturm, "Oxide Growth on Silicon: Interface Formation and Nanoscale Electrical Properties", Ph.D. thesis University of Twente, Enschede, The Netherlands (2006)

[17] B.D. Terris, J. E. Stern, D. Rugar, and H. J. Mamin, "Contact Electrification Using Force Microscopy", *Physics Review Letters* **63**, 2669 (1989)

[18] N. Felidj, J. Lambert, C. Guthmann, and M. Saint Jean, "Charge Stability on Thin Insulators Studied by Atomic Force Microscopy", *European Physical Journal AP (Applied Physics)* **12**, 85, (2000)

[19] R. Ludeke, and E. Cartier, “Imaging of trapped charge in SiO_2 and at the SiO_2–Si interface”, *Applied Physics Letters,* **78** 3998 (2001)
[20] J. Lambert, C. Guthmann, and M. Saint-Jean, “Relationship between charge distribution and its image by electrostatic force microscopy”, *Journal of Applied Physics* **93** 5369 (2003)
[21] E. Durand, “Electrostatique Tome II: Problèmes généraux conducteurs”, Masson et Cie, Paris (1966).

CHAPTER 6

CHARGE TRAPPING LIMITING CAPACITIVE MEMS-BASED SENSORS

The influence of trapped charges on characteristics of capacitive MEMS-based sensors is addressed in section 2.3. Experiments concerning charge trapping in MEMS structures are reported in Chapters 3 and 4. Not yet considered is the question how charge limits the performance of complete MEMS devices.
Designs of competitive MEMS-based sensors are guided by stringent demands concerning accuracy and reliability. The presence of trapped charges could be a troublesome factor. In this chapter, it is evaluated to what extent parasitic charge can contribute to the noise before the demands on the sensor are violated. Two capacitive devices are considered: a gravity gradiometer and an aluminum bridge RF power sensor encountered already in Chapter 3.

6.1 CHARGE TRAPPING IN A GRAVITY GRADIOMETER

Since many decades, planetary research takes a keen interest in the internal structure and composition of celestial bodies. Direct observation of the Earth's interior has been a dream of Jules Verne in "Voyage au centre de la terre" (1864), but so far the deepest hole ever drilled reached 12 km. Instead, volcanism unveils the chemical composition of regions deep in the Earth's interior. The general internal stratification of the Earth has been derived from seismic waves. For example, it can be deduced that there is an abrupt transition at a depth of 2890 km, the boundary between the highly viscous earth mantle and the liquid outer core.
In some cases, temporal variations in rotational speed indicate the existence of a liquid layer inside a celestial body, as has been the case for the planet Mercury [1]. Furthermore, planetary geologists study features at the surfaces of bodies. Cliffs, lines of fracture and volcanoes are signs of a moving crust. When a meteorite strikes a moon or a planet, material from the interior is ejected and redeposited. A spectacular example is NASA's project involving the space craft Deep Impact [2,3]. In July 2005 it inflicted an impact on the comet Tempel-1 by firing a massive projectile, weighing 370 kg, into it. The spatial spreading of the ejecta was captured on camera and analyzed. Also the magnetic field of a celestial body can hold important clues. Magnetometer measurements by the Galileo satellite fortify the suspicion that the Jupiter moon Europa holds a liquid ocean underneath its ice crust [4].
Another method of research of internal structure is gravimetry, which has become an increasingly important metrological component in planetary research. Spatial variations in the gravitational acceleration of a celestial body reveal information about the local mass distribution and hence the interior of the studied object. Generally this is done by decomposing the measured field into spherical harmonic functions, the associated Legendre polynomials. One method of gravimetry, "Line-of-Sight" (LoS), is based on the gravity that satellites in orbit experience [5,6]. By tracking the position and fluctuations in the velocity relative to earth of satellites with extreme precision, details in the "gravity map" of a planet or moon can be derived. A disadvantage is that the line-of-sight method has limited resolving power. Only relatively large-scale features in the gravity field can be mapped. Another disadvantage is that the satellite may not always be visible to the observer. A solution to this problem is to put two satellites into orbit. They monitor each other's position and velocity. With the GRACE (Gravity Recovery And Climate Experiment) project [7], the expectation is expressed that this improves the resolution of the measurement by orders of magnitude.
Single satellites in orbit are in free fall and can therefore not measure the gravity field directly on board. Gravity *gradients* are feasible however. The gravity gradient is a tensor quantity defined by $\Gamma_{ij} = \partial g_i / \partial x_j$, with g_i the gravitational acceleration vector. One is often only interested in one or a few tensor elements. The gravity gradient is commonly measured in Eötvös: 1 E = 10^{-9} (m/s^2) / m = 10^{-9} s^{-2}. For gravity gradiometry to convey useful information, a sensitivity of typically 1 E is required.
Instead, good progress is made in the development of dedicated gravity gradient sensors: *gravity gradiometers*. In [8], a superconducting gravity gradiometer is presented which achieves a sensitivity of 0.02 E/√Hz. At the time of writing, the ESA satellite GOCE (Gravity field and steady-state Ocean Circulation Explorer) is planned to be launched in 2009 [9]. This satellite is aimed at amongst others detecting the displacement of quantities of water, in service of the study of oceanic circulations. For this, the diagonal

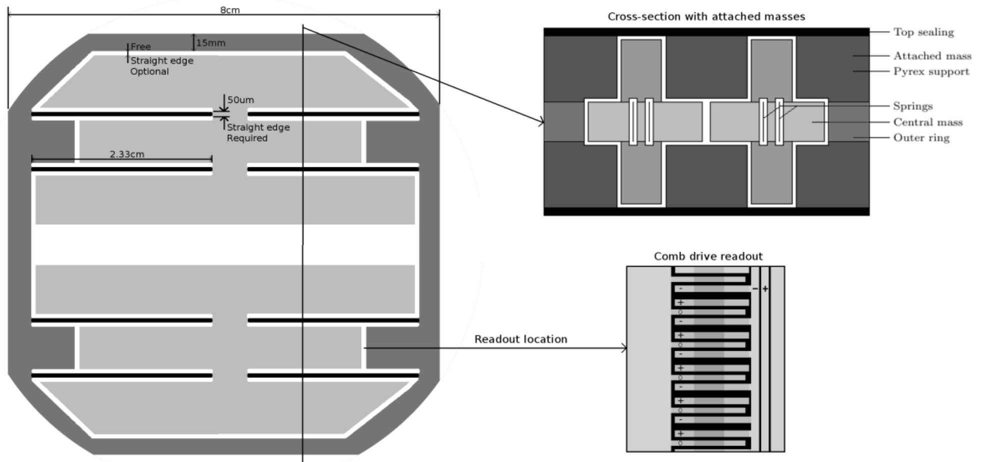


Fig.6.1 Left: MEMS gravity gradiometer (top view). Light grey areas represent masses. Thick black horizontal lines are weak springs that hold the masses in the surrounding frame. The sensitive direction is vertical in this picture. Top right: Cross section (side view). Bottom right: fragment of a comb drive (reproduced from [12])

elements of Γ are measured by three mutually orthogonal *gradiometers*. A gravity gradiometer consists of two extremely high-precision capacitive accelerometers separated by a certain distance: the baseline. The signal difference between the two accelerometers is a measure for the gravity gradient.
The high accuracy of this apparatus comes at the price of considerable weight (180 kg). For more distant planetary missions, such as the Jupiter moon Europa, this much weight is not feasible. A project currently running [10,11] involves the design of a miniaturized gravity gradiometer: a single, compact sensing system weighing about 1 kg including electronics. Its designed sensitivity is $\delta_\Gamma = 0.1$ E/√Hz. See Fig. 6.1 for a schematic.
The gravity gradiometer consists of two accelerometers separated a certain distance: the baseline. A heavy test mass and a large baseline are beneficial for sensitivity. These properties are obviously not trademarks of a miniaturized device. The structure is made by bulk micromachining of a 10 cm. wafer. The barycenters of the masses are only 5 cm apart. The masses can be endowed with extra weight, but not more than a few decagrams. Consequently, the readout has to be carried out with very high precision in order to create a sensor with sufficient sensitivity. One important step in achieving this is the use of a *comb drive* (Fig. 6.1, lower right): this is a large array of finger-shaped electrodes grasping into each other. This way, a large capacitive area is achieved within a small volume. Each accelerometer is equipped with two pairs of comb drives, in order to detect and compensate for any rotation of the gradiometer. This refinement is neglected in the discussion following below. A simple, possible schematic to read out this gravity gradiometer is shown in Fig. 6.2.
A real design contains feed-back electronics to keep the mass in its equilibrium position. For the current purposes, it suffices to show only a read-out circuit. The grey rectangles represent the accelerometer masses *m*, which are identical for the left and right accelerometer. The masses are separated by a baseline distance *b*. The dashed square on the left encompasses one such accelerometer. On opposite sides of one mass, there are two comb drives attached.

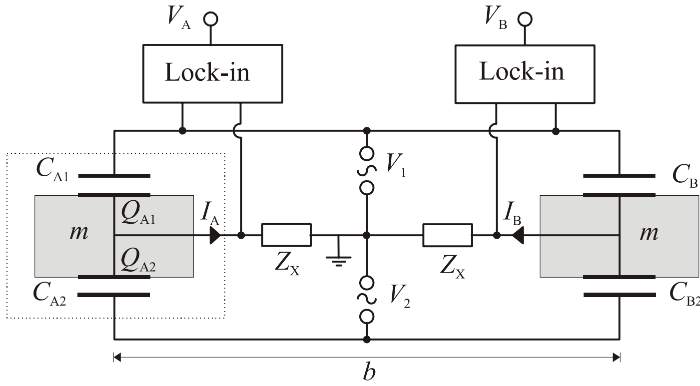


Fig. 6.2 Basic gravity gradiometer read-out scheme

For the left accelerometer, these comb drives are represented by C_{A1} and C_{A2}. The comb drives on the right are denoted C_{B1} and C_{B2}. The capacitances C_{An} (n = 1,2) and C_{Bn} are initially identical. The two sources V_1 and V_2 produce a small ac-voltage in mutual anti-phase (Eq. 6.1), at a frequency ω_V well above the mechanical resonance frequency of the mass-spring system:

$$V_1 = -V_2 = V_0 e^{i\omega_V t}. \tag{6.1}$$

The changing charges on the capacitors Q_{A1} and Q_{A2} contribute to a current I_A on the left. It will be shown that no current will run unless the capacitances C_{A1} and C_{A2} are dissimilar. This asymmetry is produced if the mass is displaced, which changes the capacitive gaps. This makes for a sensitive differential measurement. A similar situation on the right produces a current I_B, which will be present if and only if the capacitances C_{B1} and C_{B2} are unequal. The charges on those capacitor plates are not explicitly denoted. The current I_A can be read out by measuring the voltage V_{XA} across an impedant element Z_X. This voltage is fed to a lock-in amplifier, together with a possibly phase-shifted version of the source voltage. The lock-in amplifier produces a dc output signal V_A. Because a lock-in amplifier selects only signals with a frequency from a very narrow band around the operation frequency, noise with a broad frequency band is much reduced.
Like I_A on the left, on the right side the current I_B is transformed to a signal to a lock-in amplifier, which yields a dc output signal V_B. Comparing the values V_A and V_B reveals the variation of a gravity field over the length of the gradiometer. Applying Kirchhoff's laws to all closed loops learns that the voltage over the left impedant element Z_x is given by:

$$V_{XA} = \frac{C_{A1} - C_{A2}}{C_{A1} + C_{A2} + 1/i\omega_V Z_X} V_0 . \tag{6.2}$$

This shows that V_{XA} will have a nonzero value only for unequal capacitances. Obviously, the same goes for the lock-in output $V_{A,}$ which is proportional to the product $V_0 V_{XA}$ of the individual voltages.
A displacement Δz of the mass causes an increase of the gap z of one capacitor and an equal decrease of the gap of its counterpart. The capacitances C_{A1} and C_{A2} are hence not independent. Let

$$C_{A1} = \frac{\varepsilon_0 A}{z_0 - \Delta z_A} \qquad \text{and} \qquad C_{A2} = \frac{\varepsilon_0 A}{z_0 + \Delta z_A}, \tag{6.3}$$

in which z_0 is the unactuated gap, with an associated null-capacitance

$$C_0 = \frac{\varepsilon_0 A}{z_0}. \tag{6.4}$$

In these capacitances, A is the total capacitive area of the comb drive[35]. Eliminating Δz_A shows that the capacitances C_{A1} and C_{A2} are related via

$$C_0 = 2\frac{C_{A1}C_{A2}}{C_{A1} + C_{A2}} \equiv 2C_s, \tag{6.5}$$

which demonstrates that the *series* capacitance C_s is constant. For small displacements, which is certainly satisfied in case of a feedback loop, the sum of the capacitances can be considered constant. Deviations are only in the order of $(\Delta z_A/z_0)^2$. Thus we have:

$$C_{A1} + C_{A2} \approx 2C_0. \tag{6.6}$$

The difference between the capacitances is proportional to the displacement:

$$C_{A1} - C_{A2} \approx \frac{2\varepsilon_0 A}{z_0^2}\Delta z_A = 2\frac{C_0}{z_0}\Delta z_A. \tag{6.7}$$

Now consider a gravity force F_A on the left test mass m, displacing it a small distance Δz_A. The experienced gravitational acceleration g_A is thus:

$$g_A = \frac{F_A}{m} = \frac{k\Delta z_A}{m}. \tag{6.8}$$

Here k is the spring constant. After substituting (6.8) into (6.7) for Δz_A and together with (6.6) merging it with (6.2), we observe that V_{XA} is proportional to the gravitational acceleration g_A:

$$V_{XA} = \frac{mg_A / kz_0}{1 + 1/2i\omega_V Z_X C_0} V_0. \tag{6.9}$$

Hence the lock-in output V_A is proportional to g_A. The same path can be followed for the right part of the schematic Fig. 6.2, by replacing the subscripts A → B. We can now define

[35] Not to be confused with the subscripts under the quantities specified in the left part of the schematic.

the *common mode* acceleration $g_c = ½(g_A + g_B)$ by averaging the outputs of the individual accelerometers:

$$V_c = H_X(i\omega_V)g_c \,. \tag{6.10}$$

Here $V_c = ½(V_A + V_B)$ is the average voltage and $H_X(i\omega_V)$ the transfer function linking the recorded dc signal V_c to the measured quantity g_c. It is defined by

$$H_X(i\omega_V) = \frac{\lambda}{kz_0}\frac{m}{\left(1+1/2i\omega_V C_0 Z_X\right)}V_0^2 \,. \tag{6.11}$$

Here λ is a factor with dimensions V^{-1}, introduced to encompass all not further specified parameters of the specific electronic circuit, such as amplifier gain. It is seen that a high mass, a low spring constant and a narrow capacitive gap are beneficial. For a gravity gradient, we have to compare the currents of two accelerometers:

$$V_d = H_{X,b}(i\omega_V)\Gamma \,. \tag{6.12}$$

Here, V_d is the differential voltage $V_B - V_A$ and

$$\Gamma = \frac{g_B - g_A}{b} \tag{6.13}$$

the gravity gradient, which is the fluctuation of the gravitational acceleration over the length b of the detector. The *differential mode* transfer function $H_{X,b}(i\omega_V)$ is related to the common mode transfer function by $H_{X,b}(i\omega_V) = bH_X(i\omega_V)$.
The sensitivity of a practical design is limited by various noise sources. First, there is (a) the thermal noise (mechanical noise), which depends on the thermal movement of the atoms in the masses. This noise source is unavoidable and forms the absolute lower limit of the resolving capability of the sensor. It can however be reduced by operating the sensor at low temperatures and by choosing appropriate design parameters.
A further contribution comes from the read-out electronics. As can be seen in Fig. 6.2, this is in our case a capacitive read-out. Only one contribution will be studied, which is caused by charge trapping. Any other noise contributions from the read-out circuit are considered to be lower than the mechanical noise (a). The charge trapping noise will be compared to the mechanical noise. The process of charge trapping has a stochastic nature. It contributes to the total noise in two ways: (b1) current noise and (b2) force noise. The noise sources (a), (b1) and (b2) are discussed below.

(a) Thermal noise

The intrinsic *thermal* noise power in the gravity gradient for a pair of accelerometers is given by [8]:

$$S_{\Gamma,T} = \frac{8k_B T\omega_d}{mb^2 Q_d} \qquad [\mathrm{E}^2/\mathrm{Hz,\ or\ s}^3]. \tag{6.14}$$

Here k_B is Boltzmann's constant, T is the absolute temperature, ω_d and Q_d are the angular resonance frequency and the quality factor of the differential mode. Advantageous for a low noise are low temperature and low angular resonance frequency. The latter asks for weak springs and large masses. Furthermore a long baseline is clearly beneficial. Concrete design parameters are tuned such that the thermal noise does not exceed the required sensitivity. For the present sensor, this target sensitivity δ_Γ is 0.1 E/√Hz. It is thus required that $S_{\Gamma,T} \leq \delta_\Gamma^2 = 0.01$ E^2/Hz = 10^{-20} s^{-3}. Chosen values for the mass and the spring constant are m = 0.02 kg and k = 1.3 N/m. They give an angular resonance frequency $\omega_d = \sqrt{(k/m)}$ = 8.1 rad/s. so that $f_d = \omega_d/2\pi$ = 1.3 Hz. This resonance frequency is an upper limit to the measurement bandwidth. The baseline b is 5 cm. The sensor is operated at an temperature T = 77 K. A realistic value for the quality factor is $Q = 10^5$, provided that moving parts are made of silicon and that the device is operated in vacuum. The thus expected thermal noise $S_{\Gamma,T}$ = 0.014 E^2/Hz can be said to meet the target sensitivity.

(b) Charge trapping noise

As mentioned above, additional noise can be expected from charge trapping. Let us consider one capacitor in some more detail, see Fig. 6.3:

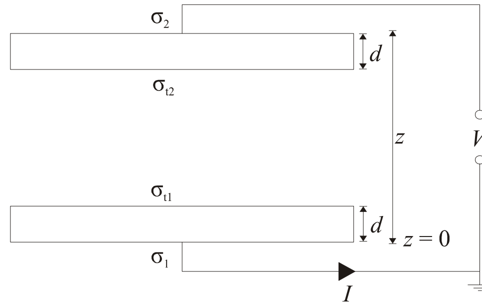


Fig. 6.3 Capacitor with dielectric layers, and surface trapped charge

Only the dielectric layers (thickness d, dielectric constant ε) are shown. The total charge density on the metallic electrodes is denoted by σ_1 and σ_2. The trapped charge, for thin electrodes assumed to be predominantly on the surface, is indicated by σ_{t1} and σ_{t2}. For feedback purposes, a *static* voltage V may be applied across the electrodes, on top of the ac read-out voltage. In principle, there are two ways in which noise in charge trapping can make itself felt: in (b1) direct trapping current noise and in (b2) force noise.
The current I (I_A in Fig. 6.2) is dependent on the charge $Q_{A1} = A\sigma_1$ and Q_{A2} (not drawn in Fig. 6.3) in the capacitors. Any noise in the latter quantities, by changes in the amount of trapped charge, affects the current. The electrode charge densities σ_1 and σ_2 depend on the trapped charges. They are given by:

$$\sigma_1 = \frac{-\varepsilon_0 V - (z-2d)\sigma_{t1} - d(\sigma_{t1}+\sigma_{t2})}{z-2d(\varepsilon-1)/\varepsilon}, \quad (6.15a)$$

$$\sigma_2 = \frac{\varepsilon_0 V - (z-2d)\sigma_{t2} - d(\sigma_{t1}+\sigma_{t2})}{z-2d(\varepsilon-1)/\varepsilon}. \quad (6.15b)$$

They can be derived from the requirements that the displacement field D makes a jump $\sigma_{t1,2}$ at the corresponding charge sheets and that the potential drops over the gap and the dielectric layers must add-up to V. It is easily seen that a current $I = A \cdot d\sigma_1/dt$ is in principle affected by changes in trapped charges $\sigma_{t1,2}$. Let us consider this in some more detail. For the following argument to be valid, it is permissible to assume that at every time the parasitic charge densities are at any time equally large and counterpolar: $\sigma_t \equiv \sigma_{t1} = -\sigma_{t2}$. This simplification cancels the last terms of Eqs. (6.15ab). Then

$$I = A\frac{d\sigma_1}{dt} = -\varepsilon_0 A \frac{d}{dt}\left(\frac{V}{z_{eff}}\right) - \frac{d}{dt}\left(\frac{z-2d}{z_{eff}} A\sigma_t\right). \tag{6.16}$$

Here $z_{\text{eff}} \equiv z - 2d(\varepsilon - 1)/\varepsilon$. The approximation $(z - 2d)/z_{\text{eff}} \approx 1$ is tempting, because deviations are of the order of $2d/\varepsilon z$, which is about 2×10^{-4} for our parameters. However, its time derivative is still of significance, because it yields a built-in voltage term. The approximation is therefore only made outside the time derivative. Thus, Eq. (6.16) becomes:

$$I = \underbrace{-\frac{\varepsilon_0 A}{z_{eff}}\frac{dV}{dt}}_{\substack{=CdV/dt,\\ \text{capacitor}\\ \text{current}}} + \underbrace{\frac{\varepsilon_0 A}{z_{eff}^2}\left(V - \underbrace{\frac{2\sigma_t d}{\varepsilon\varepsilon_0}}_{V_{built-in}}\right)\frac{dz}{dt}}_{\substack{\text{motional}\\ \text{current}}} - \underbrace{A\frac{d\sigma_t}{dt}}_{\substack{\text{trapping}\\ \text{current}}} \tag{6.17}$$

The total current can be viewed of consisting of three parts. The first part is the current through the MEMS capacitor, viewed as an ordinary, stationary capacitor. The second part results from the motion of the flexible electrode, which drives charges through the circuit. The third part is the current resulting from trapping and de-trapping of charges on the surface of the dielectric layers. This process changes the potential on the capacitor plates. The voltage source practically instantaneously restores the voltage over the capacitor by pumping charges to and from the plates. This charge flow is the contribution to the current and is in our case virtually equal to the charge trapping process.

The current I is in fact the total measurement signal (including noise), before it is converted to a form (voltage) that is suitable for the lock-in amplifier. Before considering charge trapping noise in this signal, it is helpful to illustrate the relevant frequencies that are present in I, see Fig. 6.4.

Fig. 6.4 Important frequencies on an approximately logarithmic scale

The mechanical differential mode resonance frequency f_d = 1.3 Hz divides the total frequency axis in two parts. The cross-hatched block in the low-frequency regime indicates the measurement band. It is bounded by f_1 = 1 mHz and f_2 = 1 Hz. The lower frequency limit follows from the use of the MEMS gravity gradiometer for space. The latter frequency is virtually equal to the measurement bandwidth $B = f_2 - f_1$. This

frequency region characterizes the slowly changing quantities in the current I. These are firstly the capacitive gap z_{eff} and its time derivative $dz_{eff}/dt = dz/dt$. They are affected by the slowly changing gravity forces, by a slowly changing dc component of the voltages V and by charge trapping. Because the power spectral density of charge trapping has typically a $1/f$ – character [13], it is especially strongly represented in the slow measurement band. Therefore, σ_t and $d\sigma_t/dt$ belong to this frequency category. In the high-frequency regime there is the read-out frequency $f_V = \omega_V/2\pi$ with which the voltage V and its time derivative dV/dt oscillate. It must be clearly above resonance frequency f_d in order to avoid that the flexible MEMS electrode moves along. Typical frequencies are of the order of 100 kHz, which suffices to suppress most $1/f$ – noise contributions from the electronic circuit. Two extra tick marks are drawn to indicate a $2B$ – region around f_V which is important in the signal processing. It is finally noted that z_{eff} is not affected by a fast moving voltage, other than by a small, constant amount, as stated by Eq. (2.79).
Noise in σ_t manifests itself in the current I in a direct way (trapping current + built-in voltage motional current) and in an indirect way: z_{eff} is affected by the parasitic force resulting from σ_t. These two mechanisms are discussed below.

(b1) Direct trapping current noise

The lock-in technique is very suitable to extract a desired measurement signal from a noisy background. This procedure is schematically represented in Fig. 6.5.

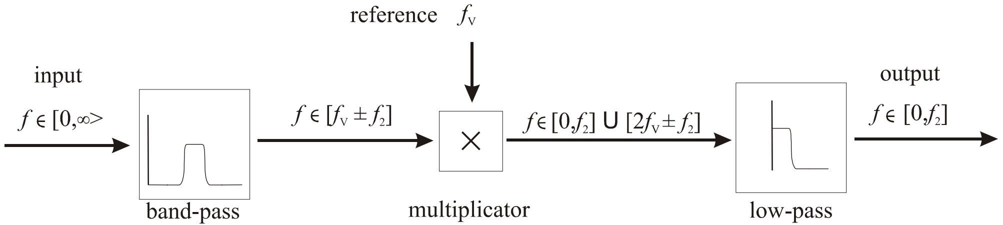


Fig. 6.5 Lock-in filtering procedure

The amplifying stage of the lock-in processing is not drawn here. First, the signal is led through a band-pass filter, which is tuned to the $2B$ – region around f_V. For a strict theoretical lock-in amplifier this stage is not necessary, but in practice it is of great help to block noise signals from especially low frequencies. The passing signal is then multiplied by a possibly phase-shifted version of the reference signal. This transforms the signal to a double frequency signal plus a signal of the original low frequency. For this reason, the multiplicator is commonly called a *demodulator*. The resulting signal is led through a low-pass filter, after which only the “cleansed” desired measurement signal is obtained.
As stated, the quantities σ_t, $d\sigma_t/dt$, z_{eff} and dz/dt are characterized by terms proportional to $\sin(2\pi f_B t)$, where f_B represents the frequencies from the measurement band B: $f_1 < f_B < f_2$. If we have a look at Eq. (6.17), we observe that the built-in part of the motional current and the trapping current contain only low frequency components. Furthermore, they are the only terms featuring σ_t or $d\sigma_t/dt$. These low-frequency signals are blocked by the band-pass filter. Therefore, the direct influence of charge trapping is not seen in the measurement.

(b2) Force noise

The capacitor current and the first term of the motional current are modulated by *V*. As discussed above on the basis of Fig. 6.5, these terms survive all through the procedure and form the desired measurement signal. It remains now to be discussed how trapped charges σ_t influence a measurement (*C* or *z*) via the electrostatic force. This will first be done for a static situation. Then a first approach to a spectral analysis is provided.
Trapped charges pull on the capacitor plates and thus change *z* or z_{eff}. The electric field in the gap between the electrodes depends on their difference:

$$E = \frac{V - d(\sigma_{t1} - \sigma_{t2})/(\varepsilon_0 \varepsilon)}{z - 2d(\varepsilon - 1)/\varepsilon}. \qquad (6.18)$$

By applying Maxwell's stress tensor on either electrode we get the total force:

$$F = \frac{\varepsilon_0 A}{2z^2}(V - V_{bi})^2, \qquad V_{bi} = \frac{d(\sigma_{t1} - \sigma_{t2})}{\varepsilon \varepsilon_0} \qquad (6.19)$$

Note again that *d* is the thickness of the *dielectric layer*. In Eq. (6.19), *d* has been neglected with respect to *z*, which is justified for the here considered device, as *d*/*z* is of the order of 10^{-3}. The trapped charge only contributes to the force noise in the VV_{bi} and the V_{bi}^2 – terms. For all but the smallest voltages (millivolts), the former term dominates. Invoking Eqs. (6.8) and (6.13) leads to the maximal admissible trapped charge density σ_{max}, which is half of the charge difference. It is strongly dependent on a static bias voltage *V*, see Eq. (6.20):

$$\sigma_{max} \equiv \tfrac{1}{2}(\sigma_{t1} - \sigma_{t2})_{max} \approx \frac{\varepsilon \varepsilon_0}{2d}\left(-V + \sqrt{V^2 + \frac{2mbz_0^2 \Gamma_{min}}{\varepsilon_0 A}} \right). \qquad (6.20)$$

Here Γ_{min} is the required sensitivity for the gravity gradient. The approximation made here is that the variable gap *z* is replaced by the constant unactuated gap z_0. For very low bias voltages, Eq. (6.20) takes a form negatively linear in the voltage:

$$\sigma_{max} \approx -\frac{\varepsilon \varepsilon_0}{2d} V + \frac{\varepsilon z_0}{2d}\sqrt{\frac{2\varepsilon_0 mb\Gamma_{min}}{A}} \qquad (V \text{ "small"}). \qquad (6.21)$$

The comparison of the voltage is with respect to the Γ_{min} term in the square root. But for already not too large voltages, the linear force term prevails clearly over the quadratic term, when the maximal admissible charge is reciprocal to the voltage:

$$\sigma_{max} \approx \frac{mbz_0^2 \varepsilon}{2dA}\frac{\Gamma_{min}}{V} \qquad (V \text{ "large"}) \qquad (6.22)$$

Typical parameters for the gradiometer design with a comb drive of 1000 fingers are listed in Table 6.1.

Quantity	Description	Value
Γ_{min}	Target gravity gradient detection treshold	0.1 Eötvös
A	Capactive area	1000 fingers × 10 × 100 μm²
z_0	Capacitive gap	2 μm
d	Thickness dielectric layer	2 nm
ε	Dielectric constant (silicon carbide)	10.2 (bulk value)
m	Mass	0.02 kg
b	Baseline	0.05 m
f_1	Lower measurement frequency bound	1 mHz
f_2	Upper measurement frequency bound	1 Hz

Table 6.1 Target specification and typical design parameters

The minimum force difference between the accelerometers that should be detectable is:

$$\Delta F_{\text{min}} = \Gamma_{\text{min}} \times b \times m = 0.1 \text{ pN}.$$

Now Eq. (6.20) can be used to calculate the amount of parasitic charge to cause a force difference corresponding with the value of the target specification. In the 'advantageous' case that no bias voltage is present, this surface charge density is:

$$\sigma_{\text{max}}= 6.8 \ \mu\text{C/m}^2 \sim 42 \text{ electrons }/\mu\text{m}^2.$$

This amount of charge would give a built-in voltage of

$$V_{\text{bi}} = 0.30 \text{ mV}.$$

Plotted in Fig, 6.6 is the parasitic surface charge density that is needed to provide a *static* force equal to the targeted minimum detectable force, for three different target sensitivities of the gradiometer. This quantity drops by an order of magnitude when a few millivolts of bias are applied, compared to the $V = 0$ situation. Note that the indicated charge densities are *excess* charges, as is indicated in Eq. (6.20). The true charge densities on one particular electrode may be higher, if it is compensated on the other electrode. Meanwhile, the built-in voltage stands at it is. Discussion of the results is deferred to section 6.3.

A detailed spectral analysis is less straightforward. Here, only a first sketch is provided for a treatment of this topic. It is remarked that charge movement is expected in the frequency regime of the measurement band. Diffusion times of charges in a dielectric reported by Felidj et al. [14], are typically 100 seconds, which corresponds to 10 mHz. This value is well within the measurement band.

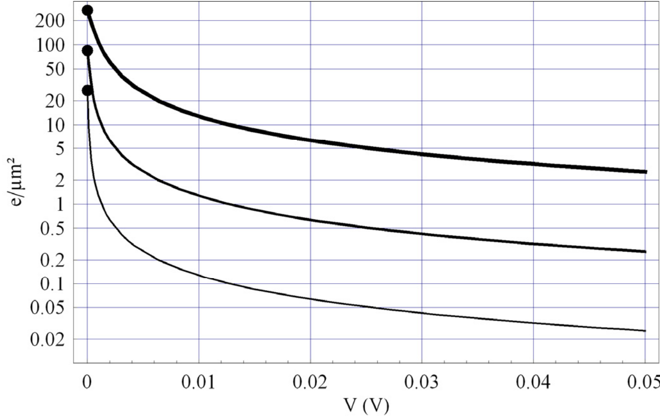


Fig. 6.6 Maximum admissible electron density vs. bias voltage for a 1.0 E (thick), 0.1 E (medium) and 0.01 E (thin) gravity gradiometer. Dots indicate the values for zero bias voltage.

Let us consider the extremes for the bias voltage. For $V = 0$ in Eq. (6.20) or Eq. (6.21), we see that the noise[36] $\delta(\Gamma)$ on Γ constrains the admissible noise $\delta(\sigma^2)$ in σ^2:

$$\delta(\sigma^2) = \frac{mbz_0^2\varepsilon^2\varepsilon_0\delta(\Gamma)}{2d^2A}. \tag{6.23}$$

In case of a bias voltage, Eq. (6.22) provides the direct noise constraint is on σ.

$$\delta(\sigma) = \frac{mbz_0^2\varepsilon\delta(\Gamma)}{2dAV} \equiv \frac{a}{V}\delta(\Gamma). \tag{6.24}$$

The symbol a is introduced for convenience to encompass all geometrical parameters. The voltage dependence of the noise has been left explicit. The gradient noise constraint is given as white noise, i.e. independent of frequency. As stated before, charge trapping noise power $\delta^2(\sigma)$ however is typically associated to be $1/f$ – type. This hypothesis reads:

$$\delta^2(\sigma) = \frac{c}{f}, \tag{6.25}$$

with c a constant and f the frequency. Furthermore, we assume that the measurement bandwidth $B \equiv f_2 - f_1$ is restricted on both sides, with f_1 the lower frequency bound and f_2 the higher frequency bound. Three possibilities are listed to resolve this incompatibility of the white noise limit with the theoretical $1/f$ - noise spectrum.

[36] The usual symbol for 'noise', 'standard deviation', 'error' etc., σ, would obviously lead to confusion in this context.

Strictly general noise limitation

The strongest demand is when the spectral charge trapping noise power density has to be lower than the target noise power density over the entire bandwidth. The charge trapping noise power density is highest at the lower frequency bound f_1 of the measurement band. In this case,

$$c \leq \frac{a^2 \delta^2(\Gamma)}{V^2} f_1 . \tag{6.26}$$

Logarithmic mean frequency

A less strong demand is if we allow the charge trapping noise power density to be higher than $\delta^2(\sigma)$ below some intermediate frequency f_{int} between f_1 and f_2. A possibility is to choose this frequency as the "logarithmic mean" of f_1 and f_2, i.e. $f_2/f_{\text{int}} = f_{\text{int}}/f_1$, which implies $f_{\text{int}} = \sqrt{f_1 f_2}$. The demand on c is then

$$c \leq \frac{a^2 \delta^2(\Gamma)}{V^2} \sqrt{f_1 f_2} \tag{6.27}$$

Integral mean

A final logical choice would be to demand that the charge trapping noise power density integrated over the bandwidth be equal to the total target noise power:

$$\int_{f1}^{f2} \delta^2(\sigma) df = \frac{a^2 \delta^2(\Gamma)}{V^2} (f_2 - f_1) ,$$

so that

$$c \leq \frac{a^2 \delta^2(\Gamma)}{V^2} \frac{f_2 - f_1}{\log(f_2 / f_1)} . \tag{6.28}$$

This demand on the noise is the least stringent of the three. The three spectral limitations to charge trapping are shown in Fig. 6.5. At f = 1 mHz, we have as trapping noise (square root of the values in Fig. 6.7):

$$\delta_1(\sigma) = 0.10\ \mu\text{C m}^{-2}\ \text{Hz}^{-1/2} = 0.6\ \text{el.}\mu\text{m}^{-2}\ \text{Hz}^{-1/2},$$
$$\delta_2(\sigma) = 0.57\ \mu\text{C m}^{-2}\ \text{Hz}^{-1/2} = 3.6\ \text{el.}\mu\text{m}^{-2}\ \text{Hz}^{-1/2},$$
$$\delta_3(\sigma) = 1.23\ \mu\text{C m}^{-2}\ \text{Hz}^{-1/2} = 7.7\ \text{el.}\mu\text{m}^{-2}\ \text{Hz}^{-1/2}.$$

The subscripts $_{1,2}$ and $_3$ correspond to the limitations 1,2 and 3 in Fig. 6.7. The first is comparable to the V = 0.01 V value of the middle line (0.1 E, ~1.5 el./μm²) in Fig. 6.6. The other two are higher.

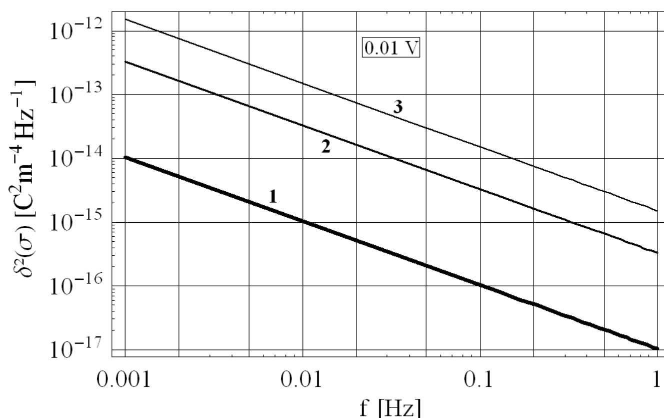


Fig. 6.7 Admissable charge trapping noise power density for 1) "strictly general noise limitation", 2) "logarithmic mean frequency" and 3) "Integral mean" for a bias voltage of 0.01 V. For larger voltages, the density is proportional to V^{-2}. In case of 0.1 V bias, all lines shift two segments down.

6.2 CHARGE TRAPPING IN AN RF POWER SENSOR

A finished project at the University of Twente was the development of a sensitive RF power sensor [15]. A common way of measuring power through a signal line is dissipating (some of) it in a resistor. However, in many applications the power *through* the line is important to know. This can be done with a capacitive MEMS sensor, which has the advantage over dissipative systems that it is independent of frequency over a very large band width. Therefore, it is suitable as a broad band (10 MHz – 1 GHz) through power sensor[37].

Several devices have been successfully realized (see Chapter 3 for more details). In Fig. 6.8 a design of an aluminum bridge is shown.

A signal with a power P and an rms voltage V_{rms} of a frequency far above resonance frequency (in the kHz range) is sent through a Coplanar Wave Guide (CPW). The power is related to the rms voltage via the characteristic impedance of the CPW ($Z = 50\ \Omega$) as in Eq. (6.29):

$$P = \frac{V_{rms}^2}{Z}. \tag{6.29}$$

The rms voltage causes a downward electric force F_{el}

$$F_{el} = -\frac{C}{2z}V_{rms}^2. \tag{6.30}$$

[37] Potentially, the extra capacitive element introduces reflection losses. This problem can be managed however by proper impedance matching further down the circuit.

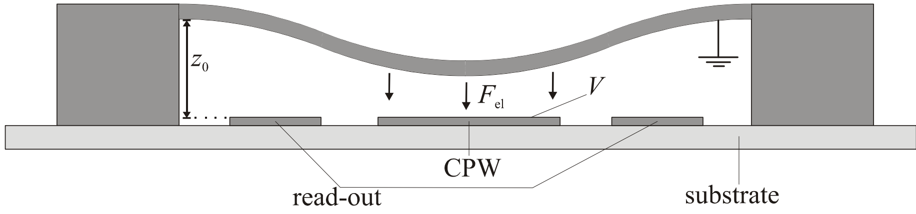


Fig.6.8 Capacitive MEMS-based RF power sensor

This force results in a small displacement $\Delta z = F_{el}/k$ of the overhanging bridge, with k the stiffness. The movement of this bridge is detected by readout electrodes parallel to the CPW[38]. This displacement in turn changes the capacitance of the sensor. The capacitance change ΔC is now related to the power through the CPW:

$$P = \frac{2z}{CZ} F_{el} \approx -\frac{2z_0}{C_0 Z_0} k\Delta z \approx \frac{2kz_0^2}{C_0^2 Z_0} \Delta C = \frac{1}{\alpha Z_0} \Delta C \,. \tag{6.31}$$

The first approximation sign indicates that the variable z and C have been replaced by their fixed, unactuated values z_0 and C_0. The same holds in fact for the CPW impedance Z, which depends on the capacitance. The second approximation sign relates displacement to change in capacitance, which are proportional for small displacements. For the last equality sign, Eq. (2.11) has been used. The reported sensitivity $\Delta C/P = \alpha Z_0$ of this sensor is a capacitance change of 100 fF per W [15]. Powers as small as 0.1-0.5 mW could be detected.
The theoretical thermo-mechanical noise δ_P of this device is given by [15,16]:

$$\delta_P = \frac{2z_0^2}{\varepsilon_0 A Z_0} \sqrt{\frac{4k_B T\sqrt{mk}}{Q}} \qquad [\mathrm{W}/\sqrt{\mathrm{Hz}}]. \tag{6.32}$$

Here A is the capacitive area, T the temperature, m the mass of the bridge and Q the quality factor of the bridge's fundamental resonance. Realistic values for the featuring parameters put this noise floor at around 1 nW/√Hz, to which the target specification of this sensor was set.
Now it will be calculated how much static trapped charge is needed to produce a displacement that corresponds to 1 nW power. Typical parameters for the power sensor design are listed in Table 6.2

The voltage resulting from the specified minimum detectable power is:

$$V = \sqrt{P \times Z} \; = 0.22 \text{ mV}.$$

[38] A more sophisticated design has feedback electrodes adopted, for keeping the bridge in a fixed position. This however entails a considerably more complex fabrication process.

Quantity	Description	Value
δ_P	Target power noise	1 nW/√Hz
A	Capacitive area	100 × 1800 μm²
z_0	Capacitive gap	1 μm
D	Thickness dielectric layer	2 nm
ε	Dielectric constant (silicon carbide)	9.1 (bulk value)
L	Bridge span width	300 μm
Z	CPW impedance	50 Ω

Table 6.2 Target specification and typical design parameters

This voltage is caused by a surface charge density trapped in the native oxide of the aluminum:

$$\sigma = \frac{\varepsilon\varepsilon_0 V}{2d} = 4.4\ \mu\text{C/m}^2 = 28\ \text{el./}\mu\text{m}^2.$$

The achieved sensitivity of the sensor, limited by the actual mechanical properties of the membrane and the quality of the readout electronics, is in the order of 0.1 mW. In that case

$$V = 70\ \text{mV},$$
$$\sigma = 1.4\ \text{mC/m}^2 = 8700\ \text{el. / }\mu\text{m}^2.$$

These numbers are compared with measurements in the next section (6.3). As stated, these are static values. Meanwhile, the RF power sensor can detect slow changes in characteristics of the high-frequency signal. The tempo of these changes has a speed limit at the mechanical resonance frequency of several kHz. As remarked in the previous section (6.1), charge trapping is especially abundant in the low frequency domain. In the same way as for the gravity gradiometer, this can cause trouble. For the dynamic interpretation of charge trapping, there is an approach possible similar to the one carried out for the gravity gradiometer.

6.3 DISCUSSION AND CONCLUSIONS

In this chapter it has been calculated what amount of charges is admissible for a device under design (gravity gradiometer, subsection 6.1) and a completed device (power sensor, subsection 6.2). These two capacitive MEMS sensors have in common that they push the limits of until then achievable sensitivities. It has been demonstrated in this thesis that charge trapping is well noticeable for this kind of devices, even with extremely thin oxide layers. It has been regarded sensible to determine what quantities of trapped charge are allowed such that the device can still function according to the demands specified for it.
The densities of trapped electrons for both devices are comparable, roughly 30-40 electrons / μm^2, if calculated for their most stringent demands. These figures are excess charge densities. On one particular electrode the charge density can be much higher, provided the charge on the counter electrode is such that the parasitic force resulting from the net , or excess charge remains unaltered.
For comparison: AFM Measurements by Sturm (section 5.2) indicate about 0.4 el./ μm^2 naturally occurring on aluminum oxide. This might sound reassuring, were it not for some objections. The first is that it is not clear to what extent the samples prepared for the AFM study are comparable to complete MEMS devices. The former undergo different fabrication processes, and under even more stable and cleaner conditions than MEMS devices. The second objection is that, according to what has been argued in Chapter 5, the counted charges may not be single charges but clusters of charges. Each cluster should contain many charges.
A more reliable guide can be found when a comparison is made with measurements on actual capacitive MEMS devices, such as presented in Chapters 3 and 4 in this thesis. Most notably section 3.3 contains measurements on voltage offsets caused by trapped charges, performed on aluminum bridge power sensors. The structures in Chapter 4 are less sensitive, but the voltage offsets are nevertheless present. The voltage offsets can immediately be compared. The current section (6.3) indicates that voltage offsets in some range around 1 mV are permitted. Instead, offsets of several decivolts have been systematically encountered for aluminum electrodes with natural oxide. This crosses the noise limits by at least two orders of magnitude. Finally, for a sensor of 0.1 mW power resolution, the calculated amount of permissible charge gives rise to a voltage that is typical in MEMS practice. In this case, charge trapping may have had a serious part in limiting the sensor's sensitivity.
Spectral quantities of the admissible noise are less straightforward to compare. What can be said is that at the lower bound of the measurement band (1 mHz), which is closest to a static case, and at a bias voltage of 0.01 V, the spectral quantity is in the same order of magnitude as the static value for maximal admissible charge. This depends on the exact way that the (white) target noise for Γ is interpreted along the $1/f$ – charge trapping noise.
To conclude: charge trapping is a truly limiting factor for the devices discussed in this chapter. In order to achieve the aimed at sensitivities, serious attention *must* be paid to charge trapping. This is preferably done in the design. Dielectrics on electrodes should be avoided as much as possible. Even ultrathin native oxides can harbor more than sufficient charges to detune the sensor. If dielectrics are unavoidable, the ambient conditions and operation of the devices should be optimized.

REFERENCES

[1] J. L. Margot et al., “Large longitude libration of Mercury reveals a molten core”, *Science* **316** 710 (2007)
[2] M. F. A. Hear et al., “Deep Impact: Excavating Comet Tempel 1”, *Science*[39] **310** 258 (2005)
[3] Deep impact: “Observations from a world-wide campaign”, *Science* **310** 265 (2005)
[4] M. G. Kivelson et al., "Galileo magnetometer measurements: a stronger case for a subsurface ocean at Europa", *Science*, **289** 1340 (2000)
[5] W. Freeden, H. Volker, and H. Nutz “Satellite-to-satellite tracking and satellite gravity gradiometry”, *Journal of Engineering Mathematics* **43** 19 (2002)
[6] J. P. Barriot, and G. Balmino, “Estimation of local planetary gravity fields using line of sight gravity data and an integral operator”, *Icarus* **99** 202 (1992)
[7] http://www.csr.utexas.edu/grace
[8] M. Vol Moody, Ho Jung Paik, and E. R. Canavan, “Three-axis superconducting gravity gradiometer for sensitive gravity experiments”, *Review of Scientific Instruments* **73** 3957 (2002)
[9] http://www.esa.int/esaLP/LPgoce.html
[10] MicroNed programme, Micro-machined electrostatic accelerometer/gradiometer
[11] R. Cuperus, J. Flokstra, R. J. Wiegerink, P. N. A. M. Visser, A. Selig, M. Smit, J. Bouman, "MEMS Based Gravity Gradiometer for Space Applications", 6th ESA Round Table on Micro & Nano technologies for space applications, ESA/ESTEC, October 2007, Noordwijk
[12] J. Flokstra, R. Cuperus, R. J. Wiegerink, and M. C. van Essen, “A MEMS-based gravity gradiometer for future planetary missions”, *proceedings of Workshop on Low Temperature Electronics (WOLTE) 8* (Jena) pp. 115 (2008)
[13] M. B. Weissman, “1/f noise and other slow, nonexponential kinetics in condensed matter”, *Reviews of Modern Physics*, **60** 537 (1988)
[14] N. Felidj, J. Lambert, C. Guthmann, and M. Saint Jean, “Charge Stability on Thin Insulators Studied by Atomic Force Microscopy”, *European Physical Journal AP (Applied Physics)* **12**, 85, (2000)
[15] L. J. Fernández, R. J. Wiegerink, J. Flokstra, J. Sesé, H. V. Jansen, and M. Elwenspoek, “A capacitive RF power sensor based on MEMS technology”, *Journal of Micromechanics and Microengineering* **16** 1099 (2006)
[16] C. T.-C. Nguyen, “Micromechanical Signal Processors”, Ph.D. Thesis, Berkeley University, Dec. 1994

[39] Special issue devoted to the Tempel 1 research

APPENDIX 1

FUNDAMENTAL MIRRORING PRINCIPLES AND POTENTIALS

The electrostatics involved in the calculations of the interaction of a conducting AFM tip with a charged sample (Chapter 5), are considered in more detail by dissecting them in more fundamental building blocks. First, three elementary image charge problems are recapitulated that are used as building blocks for the ultimate calculations. Then the issue of infinite reflections is arrived at. This is followed by recalling the principal conditions any solution to the electrostatic problem must comply, against which important principles of the multi mirror model are measured.

A1.1 BASIC MIRRORING PRINCIPLES

As appeared in Chapter 5, the electrostatic problem under consideration is shown in Fig. A1. A conducting AFM tip, modeled by a biased (*V*) sphere (radius *R*), hovers in the neighborhood of the zenith of a local charge in a dielectric layer on top of a conductor. The bias voltage is generated by a charge $q_V = 4\pi\epsilon_0 RV$ located at the centre of the conducting sphere. There is a space *a* between tip and sample surface. The dielectric is of thickness *d* and has a dielectric constant *ε*. The trapped charge *q* is located a distance *s* above the metal/dielectric interface.

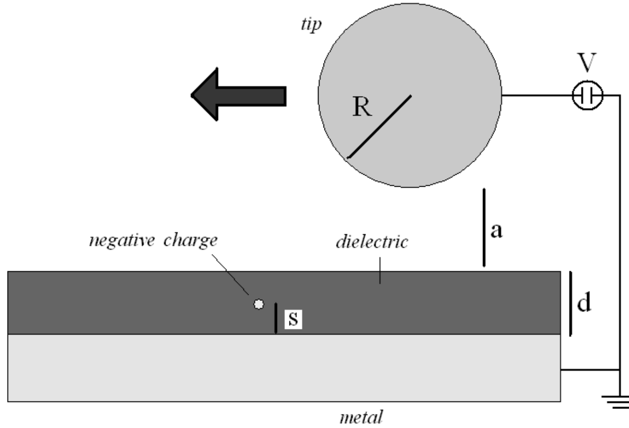


Fig. A1. Electrostatic configuration

The calculations for this problem involve three elementary image charge situations:

1) Charge *q* floating a distance *s* above a grounded plain conductor

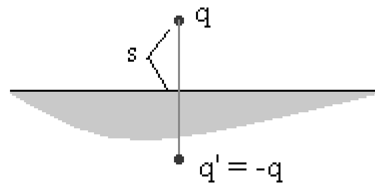


Fig. A2. Charge above plain conductor

The simplest case and a classic textbook problem [1]. In the area above the conductor, all electrostatic phenomena can be described by substituting the conductor by a charge valued $-q$ at a distance *z* below the surface of the conductor. This is justified, for in the region z > 0 the boundary conditions ($V = 0$ at $z = 0$ and at infinity) are identical in both situations, as is the source. Hence the solution to Poisson's equation is identical. In particular, for a charge *q* at (0, 0, *s*):

$$V = \frac{q}{4\pi\varepsilon_0}\left(\frac{1}{\sqrt{x^2+y^2+(z-s)^2}} - \frac{1}{\sqrt{x^2+y^2+(z+s)^2}}\right). \qquad \text{(A1)}$$

2) Charge *q* at a distance *s* from the centre of a grounded conducting sphere of radius *R*

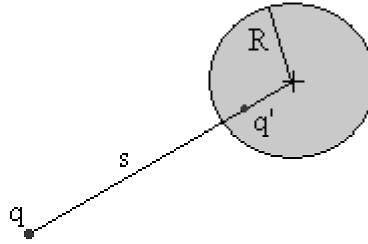


Fig. A3. Charge facing spherical conductor

A charge $q' = -Rq/s$ placed at a distance R^2/s from the centre of the sphere on the line connecting the centre with the charge ensures a spherical null equipotential right at the boundary of the physical tip, guaranteeing the equivalence for Poisson's equation and its solution for all space outside the sphere.
N.B. a sphere maintained at some potential V is the same except for an additional charge of $4\pi\varepsilon_0 RV$ at its centre.

The derivation of these results is not in all texts provided, but is straightforward. Consider two charges along the z-axis, q at the origin and $-\eta q$ at $(0,0,h)$:

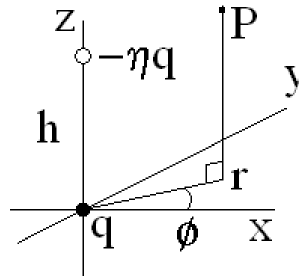


Fig. A4 two antipolar charges

For any $\eta > 0$, there exists a collection of points at which the electrostatic potential vanishes. In general, the potential at point P is given by:

$$V(P) = \frac{q}{4\pi\varepsilon_0}\left(\frac{1}{\sqrt{r^2+z^2}} - \frac{\eta}{\sqrt{r^2+(z-h)^2}}\right), \tag{A2}$$

Where $r = \sqrt{x^2+y^2}$. For $\eta = 1$, $V(P) = 0$ is satisfied for $z = \frac{1}{2}h$, when we arrive at a charge above a grounded plane, as described above. For $\eta \neq 1$, $V(P) = 0$ is obtained in case of:

$$r^2 + \left(z - \frac{h}{1-\eta^2}\right)^2 = \left(\frac{\eta h}{1-\eta^2}\right)^2. \tag{A3}$$

One recognizes the equation of a spherical shell with radius R and centre $(0,0,z_c)$:

$$R = \frac{\eta h}{1-\eta^2}, \qquad z_c = \frac{h}{1-\eta^2}. \tag{A4}$$

Consequently, *any* pair of unequal, counterpolar charges produces a spherical surface at which the electrostatic potential reduces to zero. Given R and z_c, as in the case of a charge meeting a grounded sphere, (A3) is inversed to:

$$\eta = \frac{R}{z_c}, \; h = z_c - \frac{R^2}{z_c}. \tag{A5}$$

Q.E.D. Finally, if we let u the distance between the edge of the sphere and the origin, we have $z_c = u + R$. In the limit of an infinitely large radius ($R \to \infty$), the charges become equal in magnitude ($\eta \to 1$) and the image charge will be located at $h \to 2u$, when we have again a charge 'above' a grounded conducting plane.

3) Two semi-infinite dielectrics with a charge

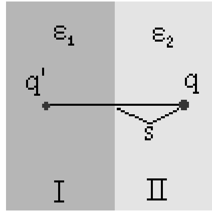


Fig. A5. Charge in adjecent semi-infinite dielectrics

In both regions I and II meaningful electrostatics can and must be described. It is demanded that the following boundary conditions are met:

$$\varepsilon_1 \left.\frac{\partial V}{\partial z}\right|_{z\uparrow \text{int.}} - \varepsilon_2 \left.\frac{\partial V}{\partial z}\right|_{z\downarrow \text{int.}} = \sigma_{\text{int}}, \tag{A6a}$$

$$\left.\frac{\partial V}{\partial x}\right|_{z\uparrow \text{int.}} - \left.\frac{\partial V}{\partial x}\right|_{z\downarrow \text{int.}} = 0. \tag{A6b}$$

The first (A6a) states that any discontinuous 'jump' in the displacement field perpendicular to the interface results from a sheet of interface charge σ_{int} (not induced) at that interface, which will be taken zero from here on. The second (A6b) demands the parallel electric field to have continuous magnitude at the interface. Here x is taken to be the parallel coordinate. The motivations for these boundary conditions are derived further down the Appendix.

In region II, where the source, or 'free' charge resides, the electrostatic problem is solved by assuming an extra single point charge in region I, valued

$$q' = -\left(\frac{\varepsilon_1 - \varepsilon_2}{\varepsilon_1 + \varepsilon_2}\right) q \,, \tag{A7}$$

at the spatially mirrored position, just like a charge above a plane conductor. An observer in region I will not sense this image charge. One can never add charges in a region where one is calculating the potential, as this comes down to unduly modifying the source in Poisson's equation. For this observer it is however possible to add or alter charges in region II. Indeed, by letting him see a charge of magnitude

$$q'' = \left(\frac{2\varepsilon_1}{\varepsilon_1 + \varepsilon_2}\right) q \tag{A8}$$

at the position of the original charge, the potential in region I connects to that in II (with the two charges) as prescribed by conditions (A6ab).

Reflections

At this point, all sample-related reflections can be described. Tip-related reflections are further largely left at rest in this text, apart from the already treated elementary image charge problem.
Started with is the problem of a charge above a dielectric layer on a conductor, see left-hand side of Fig. A6 (the situation is turned 90° clockwise in this figure).

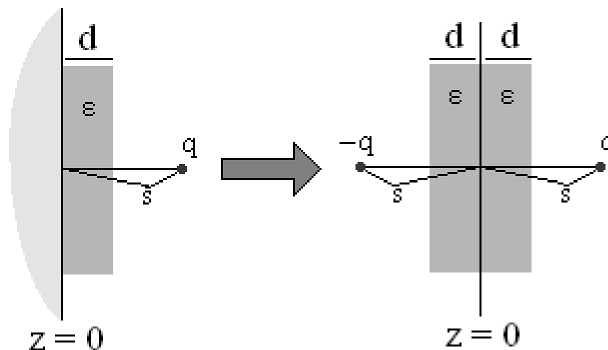


Fig. A6 Transformation hypothesis

This situation is relevant, for it is the starting point for describing the interaction between a biased tip and a grounded sample (without parasitic charge). The hypothesis is that the description of the electrostatic problem everywhere for which $z > 0$ (in and above / right of the dielectric slab) can be equivalently done by considering the situation depicted on the right side of Fig. A6, which are two antipolar charges at both sides of a dielectric slab of twice the original thickness. This is motivated by that in both cases the $z = 0$ plane constitutes a zero-equipotential. For the region $z > 0$ this boundary condition is still the same. A little more on this later. If swift inspection is not sufficient to convince the reader of the admissibility of this procedure, it should be borne in mind that a piece of dielectric material can be thought of as a large collection of positive and negative charges, and which shows polarization in response to an electric field. This complete collection of

charges is mirrored. In fact, the metal mirror of Fig. A2 is just applied here for many charges.

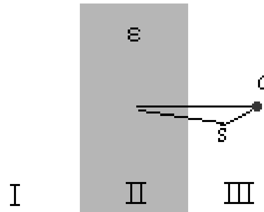


Fig. A7. Charge outside dielectric slab

After the transformation proposed in Fig. A6, the problem can be subdivided further by considering each of the two charges (original and 'virtual') apart. Starting with the stem charge q (Fig. A7), there are two types of reflection of importance:

1) Direct dielectric mirror

The special case of region II (Fig. A5) being vacuum corresponds to the case under investigation. This puts $\varepsilon_2 = 1$, which reduces Eq. (A7) to:

$$q' = -kq \quad \text{with} \quad k \equiv \frac{\varepsilon - 1}{\varepsilon + 1} \quad \text{and} \quad \varepsilon \equiv \varepsilon_1 . \qquad \text{(A9)}$$

In this special case, we recognize the Ludeke-Cartier definition (Eq. (5.30), Chapter 5) in Eq. (A7) by identifying $q'' = q/\varepsilon_{eff}$, which illustrates the statement that ε_{ff} is modifying apparent source strength rather electric field. We also see that situation Eq. (A2) is in fact a special case of situation Eq. (A5) with $\varepsilon_2 = 1$ and $\varepsilon_1 \rightarrow \infty$ in Eq. (A6) or just $\varepsilon \rightarrow \infty$ in Eq. (A8). This virtual charge Eq. (A8) can be located inside the dielectric, region II, but an apparent geometrical location in I is entirely possible for thin dielectrics or large separations of the source charge with the dielectric, which is does not of influence on its apparent strength.

2) Infinite reflections

The second effect is more complicated and leads to infinitely long chains of virtual charges. For visualization it might be helpful to recall what one is seeing while sitting in a train, representing the dielectric slab. A passenger inside sees his many reflections in the windows, that act as semi-permeable mirrors. This effect applies also if this passenger takes an object outside the train as source of reflection. An observer standing on a platform, looking into the train (Fig. A8), will see a direct reflection in the front window Eq. (A8), and if he looks deeper into the rear window, he will see many more.

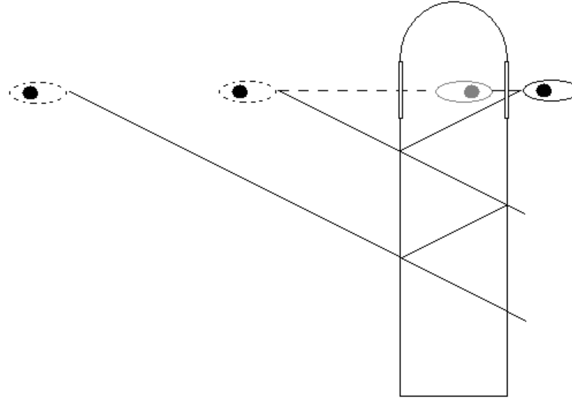


Fig.A8. Observer with reflections in two train windows

We start with an observer in region II (Fig. A7, the dielectric slab), who perceives a charge of apparent strength $(1 + k)q$ (in Eq. (A7), put $\varepsilon_2 = 1$) at the location of the original source, as described above. For this observer, this charge is mirrored in the I/II interface, again corresponding to the dielectric mirroring principles mentioned above. See Fig. A9. This observer sees this new virtual charge again mirrored in the II/III interface. This is picked up by the I/II interface again until observer II has infinitely many charges at both hands. An observer in III cannot see the charges projected in the III, but he does notice the charge chain in I, be it with an extra factor $(1 + k)$ compared to how II sees them.

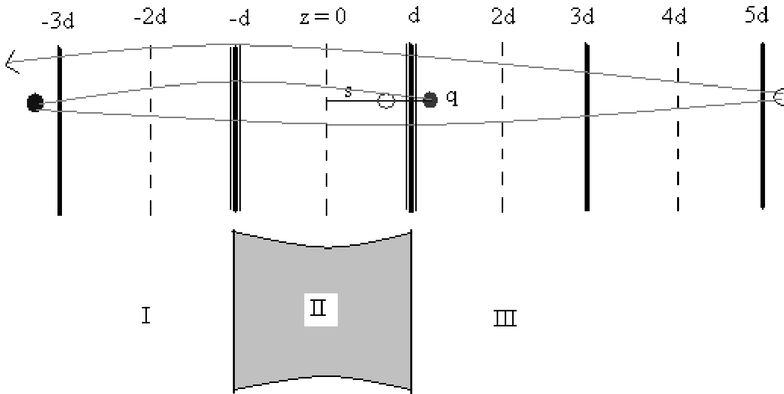


Fig. A9. principle of infinite mirroring

According to an observer in III, these charges have values

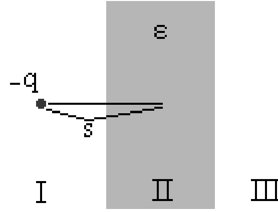
$$q_n = (1-k^2)k^{2n-1}q\ ; \quad n \geq 1, \tag{A10}$$

so are similar in sign to the original source charge. They reside on positions

$$z_n = 2d - 4nd - s; \qquad n \geq 1 \tag{A11}$$

with d the thickness of the physical dielectric slab (so region II has width $2d$) and s the coordinate of the source charge. In Fig. A9 the first of these is represented by the leftmost black dot. The following are located at west-ward positions with a periodicity of $4d$.

Metal mirror



Fig, A10. Mirrored charge outside dielectric slab

A charge, valued $-q$ (*minus q*), in domain I will be observed in III with a double 'filter':

$$q_{\mathrm{III}} = (1 + k)(1 - k)(-q)_{\mathrm{I}} \, . \tag{A12}$$

Also for this charge goes that it gives rise to infinite chains of virtual charges. Again, in III only the chain in I is observable. For III, these charges assume values

$$q_n = (1 - k^2)k^{2n}(-q)\,; \qquad n \geq 0 \tag{A13}$$

and positions

$$z_n = -4nd - s; \qquad n \geq 0 \tag{A14}$$

The original source Eq. (A12) is included in these series for n = 0. We have now demonstrated how a charge in III, e.g. bias voltage, mirrors directly in the dielectric, forms virtual charge chains and mirrors in the metal.

Charge in dielectric

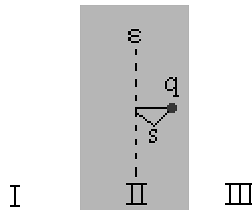


Fig. A11. Charge inside dielectric slab

A charge trapped in the dielectric (Fig. 8) is the object of research. It is the most complicated case: for an observer in II this produces *two* infinite charge chains, see Fig A12.

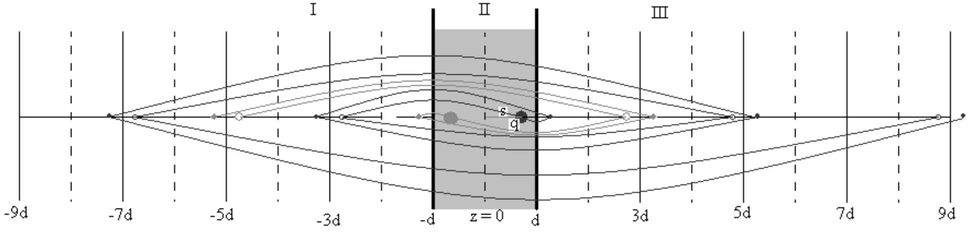


Fig. A12. Bilateral infinite mirroring

The start to read Fig. 9 is at the charge at position $z = s$, inside the dielectric slab on the right of the dotted $z = 0$ line. For an observer in the slab (II), this charge is mirrored in the I/II interface as well as in the II/III interface. These two mirrors are each the start of charge chains in an already known manner. The source charge at $z = s$ has also a mirror in the $z = 0$ plane because in reality there is still the metal electrode. This 'metal mirror' charge at $z = -s$ starts two infinite charge chains of its own. Again, for an observer in III it looks a bit less 'messy', as he sees only the semi-infinite charge chains at $z < 0$.
For the source charge at $z = s$ we have:

$$q_n = (1-k)k^{2n}q \; ; \qquad n \geq 0 \tag{A15a}$$

for one chain (which includes the original source charge for $n = 0$), and

$$q_m = (1-k)k^{2m+1}q \; ; \qquad m \geq 0 \tag{A15b}$$

for the other, at positions

$$z_n = -4nd + s \; ; \qquad n \geq 0 \tag{A16a}$$

and

$$z_m = -2d - 4md - s; \qquad m \geq 0 \tag{A16b}$$

A related problem, the potential function of a charge kept between parallel conducting plates (standard capacitor), has been described amongst others in [2]. In that case, the distribution of electrostatically opaque regions is reversed and there was not any dielectric taken into consideration.

As remarked, this charge is metal-mirrored in the $z = 0$ plane; its resulting chains have charges:

$$q_n = -(1-k)k^{2n}q \; ; \quad n \geq 0, \tag{A17a}$$

$$q_m = -(1-k)k^{2m+1}q \; ; \quad m \geq 0 \tag{A17b}$$

at positions

$$z_n = -4nd - s \; ; \qquad n \geq 0, \tag{A18a}$$

$$z_m = -2d - 4md + s; \quad m \geq 0 \tag{A18b}$$

The potential of these charges satisfies Poisson's equation and the appropriate boundary conditions, provided that m and n approach infinity, as will be demonstrated later.

It has to be noted that an observer in II perceives different charges and of different magnitudes than an observer in III. Obviously, the potential function in II has to be derived from the charges as observed in II.
It is insightful to investigate what happens in limiting cases. Let us consider the situation in Fig. A5. If there is no dielectric ($\varepsilon = 1$ so that $k = 0$), but still a metal electrode, the dielectric mirror Eq. (A9) of the charge disappears, so do the chain charges Eq. (A10). What remains is the metal mirror charge Eq. (A12), which reduces to its canonical value of $-q$. The other extreme is the dielectric slab being a metal itself, which corresponds to ε approaching infinity and k becoming 1. The dielectric mirror Eq. (A9) becomes a metal-mirrored image, while again the charge chains Eq. (A10) and the original metal mirror Eq. (A12) leave the stage.

A1.2 POTENTIAL FUNCTIONS

On a fundamental level, the solution to the electrostatic problem of the charged dielectric probed by a tip must meet Poisson's equation. The total problem can be chopped up into separate problems, as demonstrated above: charge above a dielectric slab (representing a bias charge), charge behind a dielectric slab (the metal mirror of the bias charge) and a charge inside the dielectric (the trapped charge). The metal mirror of the latter is in the same region, only of opposite sign and on the mirrored location.
In all cases, the electrostatic potential function satisfies:

$$\nabla^2 V = \frac{\partial^2}{\partial x^2}V + \frac{\partial^2}{\partial y^2}V + \frac{\partial^2}{\partial z^2}V = -\frac{\rho}{\varepsilon\varepsilon_0}, \text{with } \rho = \sum_i q_i \delta(x - x_i, y - y_i, z - z_i)\,. \qquad \text{(A19)}$$

Here, q_i is the charge, located at (x_i, y_i, z_i) which can be a bias charge, its metal mirror, or a trapped charge or one of its reflections. ε is the relative permittivity of the medium in which Poisson's equation is considered. The verification that the potential functions given below obey Eq. (A19) involves cumbersome mathematics that are neither special nor very enlightening and not thought worth reproducing, as in fact the provided potentials are just those of a collection of point charges.
Then there are certain boundary conditions that must be met. Firstly, the potential of a bounded collection of charges of finite net magnitude must vanish at infinity. The total potential functions are summations of single point charge potentials that individually obviously drop off to zero at inifinity (provided 'zero' is chosen as reference value there of course).
That the infinite set of reflected charges that is encountered in all situations is not in disharmony with the condition of 'finite net magnitude' might not be immediately clear.
Let us consider the situation of a charge q above the dielectric, as treated underneath Fig. A7. This charge has a direct mirror in the dielectric $-kq$ and an infinite series of reflected charges

$$q_n = (1 - k^2)k^{2n-1}q\ ; \quad n \geq 1. \qquad \text{(A20)}$$

Because

$$\sum_{n=0}^{\infty} p^n = \frac{1}{1-p} \qquad \forall 0 \leq p < 1\,,$$

and $k^2 < k < 1$,

$$\sum_{n=1}^{\infty} q_n = (1-k^2)kq\sum_{n=0}^{\infty}(k^2)^n = kq \,. \tag{A21}$$

Not only has the (standard) infinite sum a finite value, as required, also is the total amount of charge, original and virtual, unaltered: The dielectric mirror and the total of reflected charges add up to zero, as expected. Note that this is unlike mirroring a point charge in a conducting plane kept at a certain potential. In this case the total virtual charge is of the same magnitude and opposite sign as the stem charge, and equals the charge the connected battery has to add or remove in order to maintain the potential of the conductor in presence of the external point charge.
Less straightforward is how to encompass the requirements at the surface of the dielectric. The derivation of the proper boundary conditions [8, pp 16 -18] starts with Gauss's law and Faraday's law:

$$\nabla \cdot \mathbf{D} = \rho \,, \tag{A22}$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \,. \tag{A23}$$

It is stressed that ρ contains *only* the *free* charge, not bound charge that originates from polarization of matter. These equations can be conveniently cast into integral form by:

$$\oint_{\partial\Omega} \mathbf{D} \cdot \hat{n}_\Omega da_\Omega = \int_\Omega \rho d\Omega \,, \tag{A24}$$

$$\oint_C \mathbf{E} \cdot d\mathbf{l}_C = -\int_{S_C} \frac{\partial \mathbf{B}}{\partial t} \cdot \hat{n}_C da_C \,. \tag{A25}$$

The divergence theorem and Stokes' theorem are applied here respectively. In Eq. (A24), Ω is the volume of a Gauss surface (or 'Gaussian pillbox') encompassing thin slabs of two adjoining, not necessarily linear[40], media; see Fig. A13. The circular[41] sides are parallel to the interface. The box is assumed sufficiently small to uphold a notion of parallel even in case of a curved interface. $\hat{n}_\Omega$ denotes an outward pointing normal unit vector perpendicular to the Gaussian surface and the interface. The surface of this box is denoted with' $\partial\Omega$ '. The left-hand side integral is over surface elements da_Ω of which $\partial\Omega$ consists, whereas the right-hand side integrates the charge density ρ over volume elements $d\Omega$. In Eq. (A25), an Amperian contour C, straddling the boundary between the media, constitutes a closed path the electric field E is integrated over at the left-hand side. At the right-hand side, the time-derivative the flux of a possibly present magnetic field is integrated over the area S_C traced out by the Amperian loop, along the normal $\hat{n}_C$ of that area.

[40] In electrostatics, a medium is called 'linear' if the polarization field is proportional to the total electric field.
[41] 'Circular' was chosen for easy recognition in the corresponding figure. In fact, the actual shape is immaterial for the argument.

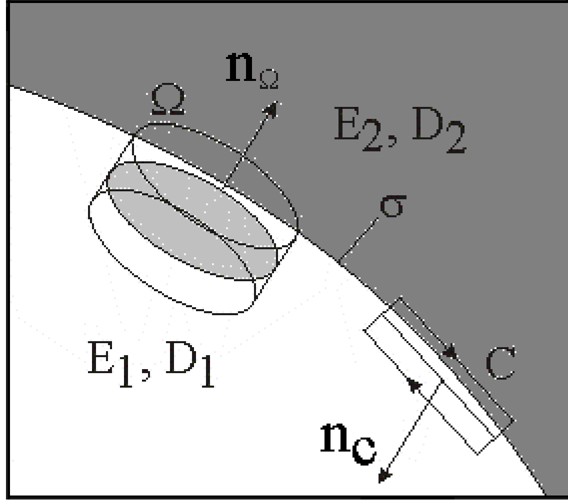


Fig. A13 Adjacent materials with Gaussian surface and Amperian loop

Concerning Eq. (A24), let us consider the limit of an infinitely shallow volume. For the left-hand side this means that the sides perpendicular to the material interface become infinitely thin, resulting in that this contribution to the integral vanishes. What is left is the contribution of the parallel parts. On the right-hand side, what remains is any sheet charge σ precisely at the boundary. Let us decompose the total enclosed charge into a proper volume charge ρ_Ω and a proper interface charge ρ_{int}.

$$\rho = \rho_\Omega + \rho_{\text{int}} , \qquad \rho_{\text{int}} = \sigma\delta(z - z_{\text{int}}) . \tag{A26}$$

The interface charge is singular with respect to a spatial coordinate z perpendicular to the material interface, right at this boundary z_{int}. Then Eq. (A24) becomes:

$$\int_{\partial\Omega //} \mathbf{D} \cdot \hat{n}_\Omega da_{\Omega //} = \int_{\partial\Omega //} \sigma da_{\Omega //} , \tag{A27}$$

Where the integration is now only over the parts of the area parallel to the interface, designated by //. Note that on the right-hand side, the volume integral is shrunk to a surface integral.
Concerning Eq. (A25), the limit of an infinitely tight Amperian loop is regarded. For the right-hand side, this means that the integration area is reduced to zero and hence the complete integral (as the integrand is finite), resulting in:

$$\oint_{C//} \mathbf{E} \cdot d\mathbf{l}_{C//} = 0 . \tag{A28}$$

On the left-hand side, only the parallel component of the contour is left as integration part. Consequently, the resulting boundary conditions for the perpendicular and parallel fields become:

$$D_{2\perp} - D_{1\perp} = \sigma , \tag{A29}$$

$$E_{2//} - E_{1//} = 0 . \tag{A30}$$

The material domains are indicated by numbers. The fields are evaluated at the interface, so meant is here $D_2 = D_2(z = z_{\text{int}})$. Equivalently, in terms of the potential:

$$\varepsilon_2 \left.\frac{dV_2}{dz}\right|_{z->z\,\text{int}} - \varepsilon_1 \left.\frac{dV_1}{dz}\right|_{z->z\,\text{int}} = \sigma\,, \tag{A31}$$

$$\left.\frac{dV_2}{dx}\right|_{z->z\,\text{int}} - \left.\frac{dV_1}{dx}\right|_{z->z\,\text{int}} = 0\,. \tag{A32}$$

Here x denotes the generalized coordinate parallel to the interface. In this context, there is no sheet of free charge present, so $\sigma = 0$ in Eq. (A30).

Charge above dielectric slab

To start with, the situation of Fig. A9 is considered, with a charge above a dielectric slab. This situation serves as the basis for a tip at a certain bias voltage relative to the sample. As has been shown in, this situation gives rise to an infinite series, or chain, of virtual charges. Their apperent strengths depend on whether the observer is located inside or outside the dielectric slab. On top, the dielectric observer sees charges in the vacuum region.
The stem charge, or original charge, is in the potential functions below assigned a location (0, 0, s).
In the present case, the total electric potential (in 3D) is given by:

Region II:

$$V_{II} = \frac{q}{2\pi\varepsilon_0(\varepsilon_r + 1)} \sum_{n=0}^{\infty} \left(\frac{\varepsilon_r - 1}{\varepsilon_r + 1}\right)^n \frac{1}{\sqrt{x^2 + y^2 + (z - (-1)^n (n \cdot 2d + s))^2}}\,. \tag{A33}$$

An observer in II sees the original charge ($n = 0$) plus all mirrored charges reflected back-and forth.

Region III:

$$V_{III} = \frac{q}{4\pi\varepsilon_0} \frac{1}{\sqrt{x^2 + y^2 + (z - s)^2}} - \frac{q}{4\pi\varepsilon_0}\left(\frac{\varepsilon_r - 1}{\varepsilon_r + 1}\right) \frac{1}{\sqrt{x^2 + y^2 + (z - 2d + s)^2}}$$
$$+ \frac{q\varepsilon_r}{\pi\varepsilon_0(\varepsilon_r + 1)^2} \sum_{n=1}^{\infty} \left(\frac{\varepsilon_r - 1}{\varepsilon_r + 1}\right)^{2n-1} \frac{1}{\sqrt{x^2 + y^2 + (z + (2n - 1) \cdot 2d + s)^2}}\,. \tag{A34}$$

An observer in III sees the original charge (the first term), its direct dielectric mirror (the second term) and all (and *only* those) reflections *below the* $z < 0$ *plane* (hence the ‘2n’ instead of ‘n’).

Metal mirror

The above potentials do not yet include the mirroring in the metal-dielectric interface. This latter effect is encompassed by mirroring the *complete situation*, both dielectric layer and charge, in the $z = 0$ plane. The potentials for the metal-mirror charge (Fig. A10) read:

Region II:

$$V_{II} = \frac{-q}{2\pi\varepsilon_0(\varepsilon_r + 1)} \sum_{n=0}^{\infty} \left(\frac{\varepsilon_r - 1}{\varepsilon_r + 1} \right)^n \frac{1}{\sqrt{x^2 + y^2 + (z + (-1)^n (n \cdot 2d + s))^2}} , \tag{A35}$$

Region III:

$$V_{III} = \frac{-q\varepsilon_r}{\pi\varepsilon_0(\varepsilon_r + 1)^2} \sum_{n=0}^{\infty} \left(\frac{\varepsilon_r - 1}{\varepsilon_r + 1} \right)^{2n} \frac{1}{\sqrt{x^2 + y^2 + (z + 2n \cdot 2d + s)^2}} . \tag{A36}$$

Charge in dielectric

For a description of this situation, see Fig. A11 and corresponding text. In this case there is no single, 'direct dielectric mirror' like in the case of a charge above the dielectric. However, a single trapped charge now gives rise to two infinite series of reflections. Again it applies that an observer in III does not 'see' any virtual charges for which $z > d$, as an observer in II does.

For the original trapped charge q the potentials are given by:

$$V_{II} = \frac{q}{4\pi\varepsilon_0\varepsilon_r} \times$$

$$\sum_{n=0}^{\infty} \left(\frac{\varepsilon_r - 1}{\varepsilon_r + 1} \right)^n \left\{ \frac{1}{\sqrt{x^2 + y^2 + (z - (-1)^n (n \cdot 2d + s))^2}} + \left(\frac{\varepsilon_r - 1}{\varepsilon_r + 1} \right) \frac{1}{\sqrt{x^2 + y^2 + (z + (-1)^{n+1} ((n+1) \cdot 2d - s))^2}} \right\} \tag{A37}$$

$$V_{III} = \frac{q}{2\pi\varepsilon_0(\varepsilon_r + 1)} \times$$

$$\sum_{n=0}^{\infty} \left(\frac{\varepsilon_r - 1}{\varepsilon_r + 1} \right)^{2n} \left\{ \frac{1}{\sqrt{x^2 + y^2 + (z + 2n \cdot 2d - s)^2}} + \left(\frac{\varepsilon_r - 1}{\varepsilon_r + 1} \right) \frac{1}{\sqrt{x^2 + y^2 + (z + (2n+1) \cdot 2d + s)^2}} \right\} \tag{A38}$$

And for the 'metal mirror' –q:

$$V_{II} = \frac{-q}{4\pi\varepsilon_0\varepsilon_r} \times$$

$$\sum_{n=0}^{\infty} \left(\frac{\varepsilon_r - 1}{\varepsilon_r + 1} \right)^n \left\{ \frac{1}{\sqrt{x^2 + y^2 + (z + (-1)^n (n \cdot 2d + s))^2}} + \left(\frac{\varepsilon_r - 1}{\varepsilon_r + 1} \right) \frac{1}{\sqrt{x^2 + y^2 + (z - (-1)^{n+1} ((n+1) \cdot 2d - s))^2}} \right\} \tag{A39}$$

$$V_{III} = \frac{-q}{2\pi\varepsilon_0(\varepsilon_r+1)}\times$$
$$\sum_{n=0}^{\infty}\left(\frac{\varepsilon_r-1}{\varepsilon_r+1}\right)^{2n}\left\{\frac{1}{\sqrt{x^2+y^2+(z+2n\cdot 2d+s)^2}}+\left(\frac{\varepsilon_r-1}{\varepsilon_r+1}\right)\frac{1}{\sqrt{x^2+y^2+(z+(2n+1)\cdot 2d-s)^2}}\right\}$$
(A40)

Now all that's left is to demonstrate that these potentials meet the boundary conditions Eqs. (A31) and (A32). This will be done for the situation of a charge above a dielectric slab only; for the other situations similar methods apply. For slightly easier writing, cylindrical coordinates are used: $r \equiv \sqrt{x^2+y^2}$, and work with a derivative d/dr. For the derivative to the vertical coordinate of the potential in region III, we have

$$\frac{4\pi\varepsilon_0}{q}\frac{\partial V_{III}}{\partial z}\bigg|_{z\downarrow d} = -\frac{d-s}{((d-s)^2+r^2)^{3/2}}-\left(\frac{\varepsilon-1}{\varepsilon+1}\right)\frac{d-s}{((d-s)^2+r^2)^{3/2}}$$
$$-\frac{4\varepsilon}{(1+\varepsilon)^2}\left(\left(\frac{\varepsilon-1}{\varepsilon+1}\right)\frac{3d+s}{((3d+s)^2+r^2)^{3/2}}+\left(\frac{\varepsilon-1}{\varepsilon+1}\right)^3\frac{7d+s}{((7d+s)^2+r^2)^{3/2}}+\ldots\right). \quad \text{(A41)}$$

And for region II:

$$\varepsilon\frac{4\pi\varepsilon_0}{q}\frac{\partial V_{II}}{\partial z}\bigg|_{z\uparrow d} =$$
$$-\frac{2\varepsilon}{\varepsilon+1}\left(-\frac{-d+s}{((d-s)^2+r^2)^{3/2}}+\left(\frac{\varepsilon-1}{\varepsilon+1}\right)\frac{3d+s}{((3d+s)^2+r^2)^{3/2}}-\left(\frac{\varepsilon-1}{\varepsilon+1}\right)^2\frac{3d+s}{((3d+s)^2+r^2)^{3/2}}\right)$$
$$-\frac{2\varepsilon}{\varepsilon+1}\left(\left(\frac{\varepsilon-1}{\varepsilon+1}\right)^3\frac{7d+s}{((7d+s)^2+r^2)^{3/2}}-\left(\frac{\varepsilon-1}{\varepsilon+1}\right)^4\frac{7d+s}{((7d+s)^2+r^2)^{3/2}}+\ldots\right). \quad \text{(A42)}$$

The first two terms of the derivative of V_{III} are from the original charge and its direct dielectric mirror. They exactly add up to the first term of the derivative of V_{II}. Higher order terms of $\partial V_{\mathrm{III}}/\partial z$ each correspond with two terms of $\partial V_{\mathrm{II}}/\partial z$, as the reader can verify for himself, for example for the $3d + s$ terms. The fact that both series extend to infinity guarantees that the boundary condition for the z-derivative is met. A comparable pattern is encountered for the boundary condition for the horizontal coordinate: $\partial V_{II}(z\uparrow d)/\partial r = \partial V_{III}(z\downarrow d)/\partial r$, which is not reproduced here.

APPENDIX 2

AFM ELECTROSTATICS: MISCELLANEOUS

A2.1 ELECTROSTATIC ENERGY OF A SYSTEM OF CONDUCTORS AND REFLECTION POTENTIAL

In Chapter 5, the principle of mirroring in the tip has been explained. The force on the tip, and from it the force gradient, is in principle calculated by summing all interactions that each charge in the tip (N in total), virtual and real, has with each of the 'outside-tip' charges. This calculation is quite cumbersome and longwinded, as the calculation time scales with N^2. More efficient would be to invoke the energy of the system, which scales only with N. For a particular charge configuration, this can be calculated. A small displacement of the tip Δz will change the total energy of the system ΔU, which leads one to determine the force as

$$F_z = -\frac{\Delta U}{\Delta z}. \tag{A43}$$

The force gradient is then again extracted by a differential of the force. This demands to determine the energy at three closely spaced points with high accuracy. This appendix will digress on the energy of the tip-charge system. To this end, started is with a system of n conductors at potentials V_i and charges Q_i and space charge distributions ρ_j spread between them, possibly embedded in a dielectric, see Fig. A14

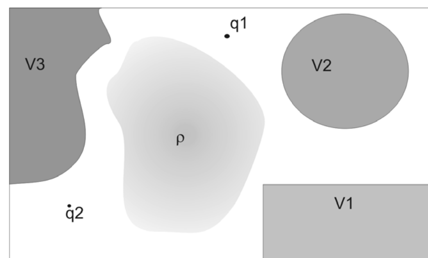


Fig. A14 System of conductors at potentials V_i and space charge densities ρ_i.

For this system, the energy of a volume Ω is given by:

$$U = \frac{1}{2}\int_{\Omega} V(\vec{x})\rho(\vec{x})d\Omega + \frac{1}{2}\sum_{i=1}^{m} V_i Q_i \, , \qquad \text{(A44)}$$

as derived in [4]. In our case, we consider the situation of Fig. A1 in Appendix 1, with one grounded electrode and the other electrode, the tip, at potential V. Here, the space charge ρ is a point charge q at point $S = (0, 0, s)$: $\rho = q\delta(0, 0, s)$. It is posed that the total charge Q on the tip has linear contributions from both the bias potential and the free charge q:

$$Q(V, q) = CV + aq \, , \qquad \text{(A45)}$$

In which C and a are constants, defined by geometry. The former is the (differential) capacitance of the system, defined by $C = \partial Q/\partial V$. Eq. (A45) is based on the superposition principle and of the fact that potential scales with charge, which is a consequence of the former. According to the same principles, we can unravel the potential at S:

$$V_S(V, q) = bV + q / c \, . \qquad \text{(A46)}$$

Again, b and c are again constants. The first term of Eq. (A46) is due to the bias potential only; the second term could be interpreted as “reflection potential”: the presence of the charge q at S induces charge on the electrodes, which contributes to the potential at S[42]. This shows that q is not a passive onlooker, or independent test probe, but contributes itself to the electrostatic landscape (potential) it experiences.[43] It is tempting to therefore entitle c a ‘reflection capacitance’. Now, we invoke the energy equation Eq. (A44) for our situation, which requires considering just one electrode with nonzero potential:

$$U(V, q) = \tfrac{1}{2} VQ(V, q) + \tfrac{1}{2} qV_S(V, q) \, . \qquad \text{(A47)}$$

In order to collect more information about the constants a and b, the differential of Eq. (A47) is taken and re-integrated. Proceeding thus and grouping terms in dq and dV provides:

$$dU = \left(\tfrac{1}{2} V \frac{\partial \tilde{Q}}{\partial V} + \tfrac{1}{2} \tilde{Q} + \tfrac{1}{2} q \frac{\partial \tilde{V}_S}{\partial V} \right) dV + \left(\tfrac{1}{2} V \frac{\partial \tilde{Q}}{\partial q} + \tfrac{1}{2} \tilde{V}_S + \tfrac{1}{2} q \frac{\partial \tilde{V}_S}{\partial q} \right) dq \, . \qquad \text{(A48)}$$

In Eq. (A48), “˜ ” denotes functional dependence on V and q. Integrating Eq. (A48) again yields:

$$U = \int dU = \frac{\partial \tilde{Q}}{\partial V} \int_0^V \tfrac{1}{2} V' dV' + \int_0^V \tfrac{1}{2} \tilde{Q} dV' + \frac{\partial \tilde{V}_s}{\partial V} \int_0^V \tfrac{1}{2} q' dV' + \frac{\partial \tilde{Q}}{\partial q} \int_0^q \tfrac{1}{2} V dq' + \int_0^q \tfrac{1}{2} \tilde{V}_s dq' + \frac{\partial \tilde{V}_s}{\partial q} \int_0^q \tfrac{1}{2} q' dq' \, . \qquad \text{(A49)}$$

[42] That is, disregarding the infinite (singular) potential at S due to q directly.

[43] This reminds of the widely spread philosophical notion that a person is not a passive figure, just perceiving (receiving?) a reality independent of him, but in fact himself influences (some would even say *creates*) the world he experiences.

According to Eqs. (A45) and (A46), the partial derivatives are constants and can be taken outside the integrals.
Inserting Eqs. (A45) and (A46) into Eq. (A49) of course returns Eq. (A47). In an operational sense, assume we start with no charge and no bias voltage. To arrive at the final situation, we could first raise the bias voltage to its final value, which gives the 1st and the 2nd term, and then increase the free charge q to its final value, building up the last three terms of Eq. (A49). During the whole process, the 3rd term remains zero. Alternatively, we could first increase the free charge q, which creates the 5th and 6th terms. After this, the bias potential is raised: terms 1-3. Now the 4th term stays zero. This implies that the contributions of the 3rd and 4th terms must be the same. Consequently, the partial derivatives heading the integral signs, and hence a and b, are equal. On the basis of this argument, we can now write for the energy:

$$U(V,q) = U_{cap,V^2} + U_{charge-bias,qV} + U_{refl,q^2} = \tfrac{1}{2}CV^2 + qV_{S,bias} + \tfrac{1}{2}qV_{S,refl} \,. \tag{A50}$$

The energy hence consists of three contributions. The first one is from the pure capacitor. The second term is the interaction between the trapped charge and the bias voltage. Here $V_{\mathrm{S,bias}} = bV$ is the potential in S due to the bias voltage *only*. The third term finally denotes the energy it requires to build up a charge q in the presence of grounded conductors. The reflection potential V_{refl} is given by q/c, see Eq. (A46). The validity of (A50) is easily checked for the force on a point charge in the presence a grounded sphere or infinite metal or dielectric, in which cases only the reflection potential applies. A biased sphere is then the next topic.
One little trap is worth mentioning however. Consider the textbook example of Fig. A2, with a charge above a grounded plane. The reflection potential at the charge's location here is given by $4\pi\varepsilon_0 V_{refl} = q/(2s)$. If we now calculate the force on q by $F = -\frac{d}{ds}qV_{refl} = -(8\pi\varepsilon_0)^{-1}q^2/s^2$, we overestimate the force by a factor of 2. The mistake made is that in this case, moving the charge over a distance ds moves also the mirror at this distance. In fact, one should keep the mirror at a fixed position $-s$ and the charge at a variable position z. Now $F = -\frac{d}{dz}qV_{refl} = -(4\pi\varepsilon_0)^{-1}q^2\frac{d}{dz}(s+z)^{-1}$ and replacing $z \to s$ afterwards yields the correct result. Systems involving more conductors are not so trivial.
The numerical task lies now in establishing the values and positions of 'all' charges and their reflections. And derive from that the capacitance C, the bias-related potential $V_{\mathrm{S,bias}}$ at the location S of the trapped charge and finally the reflected potential $V_{\mathrm{S,refl}}$ After the intermezzo, an attempt is made to approximate these quantities by closed-from expressions.

INTERMEZZO: INFINITE FRACTION AND RECURRENCE RELATION

Calculations in the next part of this appendix make use of the following recurrence relation:

$$r_{i+1} = \frac{1}{p - r_i}\,, \tag{A51}$$

so that

$$r_1 = \frac{1}{p - r_0}, \qquad r_2 = \frac{1}{p - r_1} = \frac{1}{p - \frac{1}{p - r_0}}, \qquad r_3 = \frac{1}{p - \frac{1}{p - \frac{1}{p - r_0}}} \tag{A52}$$

etcetera. In general, r_i is an ith order continued fraction. For physical reasons, we want the recurrence relation Eq. (A51) to converge, i.e. the successive terms must stabilize at some fixed value r. This happens when $r_{i+1} = r_i$, which has the solutions

$$r_\pm = \tfrac{1}{2} p \pm \tfrac{1}{2}\sqrt{p^2 - 4}\,, \tag{A53}$$

with the properties that $r_+r_- = 1$ and $r_+ + r_- = p$. The limit is independent of the first 'seed' term, r_0. A convergence limit exists only when $|p| \geq 2$. We see that $r_{i+1} < 1$ for $r_i < p - 1$. $r_{i+1} > 1$ will be in cases of $r_i > p - 1$. In this case however, r_{i+2} will be < 1, so that all subsequent elements (for $i > 3$) will be smaller than 1. Consequently, the convergence limit is $r_- < 1$, shortly written r. Now, we may define a series δ_i, which denotes the difference of each next element r_{i+1} with the convergence limit. By this definition:

$$r_{i+1} = r - \delta_i. \tag{A54}$$

Combining this with the recurrence relation Eq. (A51):

$$r_{i+1} = \frac{1}{r_+ + \delta_{i-1}} = r - \delta_i\,, \tag{A55}$$

in which definition Eq. (A54) has again been invoked. Hence

$$\delta_i = \frac{\delta_{i-1}}{r_+(r_+ + \delta_{i-1})}\,. \tag{A56}$$

We can write a similar equation for δ_{i-1}, so that

$$\delta_i = \frac{1}{r_+\left(\frac{r_+^3}{\delta_{i-2}} + r_+^2 + 1\right)}\,, \tag{A57}$$

and more in general

$$\frac{1}{\delta_i} = r_+\left(\frac{r_+^{2i-1}}{\delta_0} + \underbrace{r_+^{2i-2} + r_+^{2i-4} + \ldots + r_+^2 + 1}_{i}\right) = r_+\left(\frac{r_+^{2i-1}}{\delta_0} + \frac{1 - r_+^{2i}}{1 - r_+^2}\right). \tag{A58}$$

Here, according to (A55),

$$\delta_0 = r - r_1 = r - \frac{1}{p - r_0}\,. \tag{A59}$$

Now the ith element of the δ_i is known, we have consequently a closed form for the ith element of the r series. Combining Eq. (A58) with Eq. (A54) and rearranging yields:

$$r_i(\lambda) = \frac{1}{r}\frac{\beta_\lambda r^{2-2i} - 1}{\beta_\lambda r^{-2i} - 1}, \qquad \beta_\lambda \equiv \frac{1-\lambda r}{(p-\lambda)r - 1}, \qquad \lambda \equiv r_0 . \tag{A60}$$

Here, the λ - symbol has been introduced to slightly facilitate notation and to highlight the independent significance of the first element r_0 for the r_i series. A special case we encounter for $\lambda = 0$, when β simplifies to

$$\beta_0 = \frac{1}{pr-1} = \frac{1}{r^2}, \tag{A61}$$

so that

$$r_i(0) = \frac{1}{r}\frac{r^{-2i} - 1}{r^{-2-2i} - 1}. \tag{A62}$$

A2.2 APPROXIMATIONS FOR ELECTROSTATIC ENERGY IN THE AFM SYSTEM

The three relevant interactions are drawn in Fig. A15.

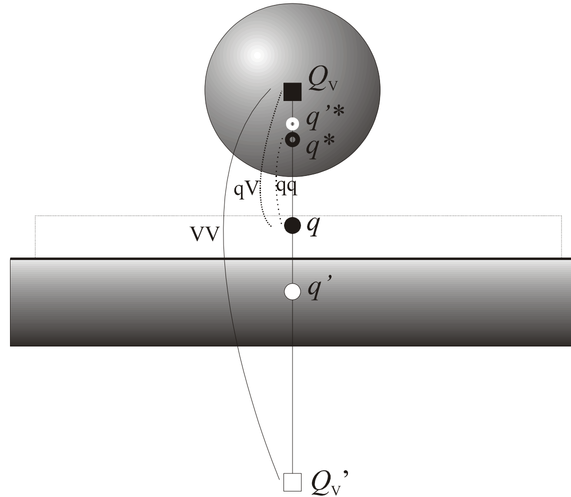


Fig. A15. Tip with bias charge Q_V, mirrored bias charge Q_V'. The dashed area is a dielectric layer. In the coming discussion, its influence is not yet treated. Its incorporation is covered later. The layer contains a trapped charge q, which is mirrored in the metal q'. Both have images in the tip q and q'* respectively. The basic interactions are indicated with dashed and solid curved lines. These are VV, qV and qq: the bias interaction of the tip with the sample (charge-charge), the interaction between the trapped charge and the bias charge (dipole-charge) and the interaction of the trapped charge with its image in the tip (dipole-dipole).*

A2.2.1 Capacitance (VV)

In order to calculate the first term of the energy U, Eq. (A50), we need to know the capacitance of the system. In principle this is done by all bias charges in the tip, divided by the potential of the tip (radius R). The latter can be represented by a suitable charge $Q_V = 4\pi\varepsilon_0 RV$. This bias charge is mirrored in the $z = 0$ plane. The ith charge in the tip Q_i at position z_i results in a mirror Q_i in the $z = 0$ plane, at position z_i':

$$Q_n' = -Q_n , \qquad z_n' = -z_n . \tag{A63}$$

This mirror is imaged again by the tip, resulting in the i+1st charge:

$$Q_{i+1} = -\frac{R}{z_t - z_i'} Q_i' = \frac{R}{z_t + z_i} Q_i , \qquad z_{i+1} = z_t - \frac{R^2}{z_t - z_i'} = z_t - \frac{R^2}{z_t + z_i} \tag{A64}$$

Here z_t is the position of the centre of the spherical tip. The first few tip charges are

$$\begin{aligned} Q_0 &= Q_V & z_0 &= z_t \\ Q_1 &= \frac{R}{z_t + z_0} Q_0 = \frac{R}{2z_t} Q_0 & z_1 &= z_t - \frac{R^2}{z_t + z_0} = z_t - \frac{R^2}{2z_t} \\ Q_2 &= \frac{R}{z_t + z_1} Q_1 = \frac{R}{2z_t - \frac{R^2}{2z_t}} \frac{R}{2z_t} Q_0 & z_2 &= z_t - \frac{R^2}{z_t + z_1} = z_t - \frac{R^2}{2z_t - \frac{R^2}{2z_t}} . \end{aligned} \tag{A65}$$

We define

$$p \equiv \frac{2z_t}{R} , \tag{A66}$$

so that $p > 2$. Then we can write

$$Q_3 = \frac{1}{p - \frac{1}{p - \frac{1}{p}}} \underbrace{\left(\frac{1}{p - \frac{1}{p}} \underbrace{\left(\frac{1}{p} \underbrace{Q_V}_{Q_0} \right)}_{Q_1} \right)}_{Q_2} , \tag{A67}$$

in which one immediately recognizes a pattern of a continued fraction. The next step is to define

$$r_n \equiv \frac{Q_n}{Q_{n-1}} < 1 \tag{A68}$$

as the ratio between two successive charges. If we apply this to Eq. (A67), we see that these ratios obey the recurrence relation

$$r_{n+1} = \frac{1}{p + r_n},$$

which is the one treated in the intermezzo, see Eq. (A51). As we have $r_1 = 1/p$, $r_0 = \lambda = 0$ applies, which is consistent with (Eq. (A51), formula for r in intermezzo). For the nth charge we have:

$$Q_n = r_n r_{n-1} ... r_1 Q_V = Q_V \prod_{i=1}^{n} r_i \,. \tag{A69}$$

This works also in case of $n = 0$: for $n < i$, we obtain the so called “empty product”. This means “no multiplication takes place”, and returns the “multiplicative identity”, which is 1. We obtain $Q_0 = Q_V$, as desired. Meanwhile,

$$\prod_{i=1}^{n} r_i = \left(\frac{1}{r}\right)^n \left(\frac{r^{-2}-1}{r^{-4}-1}\right) \times \left(\frac{r^{-4}-1}{r^{-6}-1}\right) \times ... \times \left(\frac{r^{-n+4}-1}{r^{-n+2}-1}\right) \times \left(\frac{r^{-n+2}-1}{r^{-n}-1}\right).$$

All numerators and denominators cancel, save one of each, so that the product Π_n evaluates explicitly:

$$\Pi_n \equiv \prod_{i=1}^{n} r_i = \left(\frac{1}{r^2} - 1\right) \frac{r^n}{r^{-2} - r^{2n}} = \frac{Q_n}{Q_V} \,. \tag{A70}$$

For general λ this result reads:

$$\Pi_n(\lambda) \equiv \prod_{i=1}^{n} r_i(\lambda) = (\beta_\lambda - 1) \frac{r^n}{\beta_\lambda - r^{2n}} \,. \tag{A71}$$

For the capacitance, it is required to sum the Q_n as defined by Eqs. (A69) and (A71). The summation cannot be carried out exactly, but we can estimate it by comparing it to the geometric series. Because $0 < r < 1$, for the denominator we have $0 < r^{2n} < 1$, while in the numerator r^n can be summed explicitly. Therefore

$$1 + r < \sum_{n=0}^{\infty} \frac{Q_n}{Q_V} < \frac{1}{1-r} \,. \tag{A72}$$

The general series (A71) converge approximately with powers in r, but can be rewritten in a form that converges faster, by expanding it in powers of y, where $y \equiv r^n$:

$$\frac{y}{\beta_\lambda - y^2} = \frac{y}{\beta_\lambda} + \frac{y^3}{\beta_\lambda^2} + \frac{y^5}{\beta_\lambda^3} + ... = \sum_{j=1}^{\infty} \frac{y^{2j-1}}{\beta_\lambda^j} \,,$$

So

$$Q \equiv \sum_{n=0}^{\infty} Q_n = Q_V (\beta_0 - 1) \sum_{j=1}^{\infty} \frac{1}{\beta_0^j} \sum_{n=0}^{\infty} \left(r^{2j-1}\right)^n = Q_V (\beta_0 - 1) \sum_{j=1}^{\infty} \frac{1}{\beta_0^j} \frac{1}{(1 - r^{2j-1})} \,. \tag{A73}$$

This series converges roughly with powers of β_0^{-1}, which means a reduction of each term with a factor r^2 instead of r.
In many cases, roughly when tip-distance ratio $p = 2z_t/R > 2.3$[44], the capacitance of the system can be approximated by taking only the first term of Eq. (A73). We have then explicitly:

$$C = \frac{dQ}{dV} = 4\pi\varepsilon_0 R(1+r) = 2\pi\varepsilon_0 R\left(2 + p - \sqrt{p^2 - 4}\right). \qquad \text{(A74)}$$

For a very small radius ($R \downarrow 0$) or a very large distance ($z_t \rightarrow \infty$), we see that $p = 2z_t/R$ extends to infinity (and hence $r \downarrow 0$). The capacitance (A74) reduces to the spherical capacitance $4\pi\varepsilon_0 R$. For closer spacings, larger radii or just enhanced accuracy, a second term from Eq. (A73) can be added:

$$C = \frac{dQ}{dV} = 4\pi\varepsilon_0 R\left(1 + 2r - \frac{r}{1+r+r^2}\right) = 4\pi\varepsilon_0 R\left(1 + p - \sqrt{p^2 - 4} - \frac{1}{1+p}\right). \qquad \text{(A75)}$$

The capacitance and hence the capacitive energy $U_{cap} = ½CV^2$ can therefore well be approximated by a closed-form expression.
It is however stressed that for close approximations ($2 < p < 3$), these formulas are inadequate for deriving the force. The force is the spatial derivative of the capacitance. The provided capacitance approximations however underestimate the relatively strong upwards curvature of the capacitance close to the surface. This will be demonstrated later in the context of dielectrics. For a force calculation, it is therefore advised to (numerically) carry out the full sum in Eq. (A73).

INTERMEZZO: PASCAL'S TRIANGLE, FIBONACCI, GOLDEN RATIO

The Q_n can alternatively be described in n^{th} order polynomials P_n in the variable p: The reader can verify that applies:

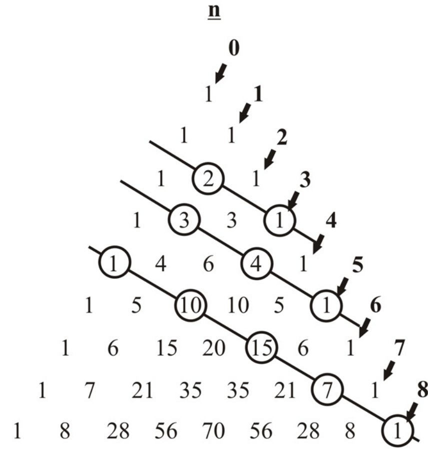
$$\frac{Q_V}{Q_0} = 1, \qquad \frac{Q_V}{Q_1} = p, \qquad \frac{Q_V}{Q_2} = p^2 - 1, \qquad \frac{Q_V}{Q_3} = p^3 - 2p,$$

etcetera.

[44] This number is admittedly arbitrary. It applies for example for a tip of 100 nm radius hovering with a spacing of 30 nm (z_{tip} = 130 nm) above a metal surface.

The first few polynoms $P_n(p)$ are:

$$\begin{gathered} 1 \\ p \\ p^2-1 \\ p^3-2p \\ p^4-3p^2+1 \\ p^5-4p^3+3p \\ p^6-5p^4+6p^2-1 \\ p^7-6p^5+10p^3-4p \\ p^8-7p^6+15p^4-10p^2+1\,. \end{gathered} \tag{A76}$$



Fig, A16 Pascal's triangle supplies coefficients for capacitive polynomials $P_n(p)$.

According to Eq. (A68), we can write:

$$r_n = \frac{P_{n-1}(p)}{P_n(p)}\,. \tag{A77}$$

The polynomials $P_n(p)$ are alternating even or odd and the successive terms in each polynomial have alternating signs. The coefficients are recognizable as numbers in Pascal's triangle. The nth diagonal represents the nth power of p, showing the absolute value of the corresponding coefficient. The polynomials are built up of 'flat diagonals' in the triangle. In Fig. A16, the 3rd, 5th and 8th degree capacitance polynomials are shown. Note the alternating signs for the coefficients. The coefficient for the leading power term is always +1. Because Pascal's triangle is made up of binomial coefficients an alternative formulation for Q_n is:

$$\frac{Q_V}{Q_n} = \sum_{m=0}^{n/2} a_{n,n-2m}\, p^{n-2m} = \sum_{m=0}^{n/2} (-1)^m \binom{n-m}{m} p^{n-2m} = \sum_{i=0}^{n/2} (-1)^m \frac{(n-m)!}{m!(n-2m)!} p^{n-2m}\,. \tag{A78}$$

It is known that the sums of the numbers on the highlighted diagonals in Fig. A16 form members of the Fibonacci sequence 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, … which is defined by that every entry is the sum of the previous two. Indeed, one easily verifies for sample cases:

$$\sum_{m=0}^{n/2} \left| a_{n,n-2m} \right| = F_{n+1}\,, \tag{A79}$$

in which F_j is the jth Fibonacci number ($F_0 = 0$). This perhaps remarkable property of the tip charges is no coincidence. Recall that the Fibonacci sequence can be generated by the *golden ratio* $\varphi = \frac{1}{2} + \frac{1}{2}\sqrt{5} \approx 1.618$:

$$F_n = \frac{\varphi^n - (1-\varphi)^n}{\sqrt{5}}\,.$$

One defining description of the golden ratio is the *continued fraction*

$$\varphi = \cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{1+\dots}}}},$$

which reminds us of the relative magnitudes of the successive tip charges (Eqs. (A52) and (A67)).

A2.2.2 Charge – bias interaction

For the second energy-term, describing the interaction of the trapped charge with the applied bias voltage, many of the so far derived results can be used. For the bias potential in S, the required ingredients are firstly the bias charges in the tip and their antipolar mirrors in the flat electrode: the $z = 0$ plane. These are known. Secondly, the positions of these charges must be determined or, more precisely, the distances between S and these charges. The locations of the tip charges up to second order have already been given in Eq. (A65). In general, it can be seen that the positions of the bias charges in the tip obey

$$z_n = z_t - Rr_n(0)\,, \tag{A80}$$

where the $r_n(0)$ obey the recurrence relation (A51). The $r_n(0)$ are explicitly given by Eq. (A62). In the limit of $n \to \infty$, the z_n stabilize on $z_\infty = \sqrt{(z^2 - R^2)}$. Now, for the potential in S we write

$$V_{S,bias} = \frac{Q_V}{4\pi\varepsilon_0}\sum_{n=0}^{\infty}\Pi_n\left(\frac{1}{z_t - Rr_n - s} - \frac{1}{z_t - Rr_n + s}\right) = \frac{2sV}{R}\sum_{n=0}^{\infty}\frac{R^2\Pi_n}{(z_t - Rr_n)^2 - s^2} \equiv \frac{2sV}{R}\sum_{n=0}^{\infty}\Phi_n(s) \tag{A81}$$

in which $\Phi_n(s)$ are dimensionless potential functions. Now consider the case $s = z_t - R$, which is located at the edge of the tip. Then

$$\Phi_{t,n} \equiv \Phi_n(z_t - R) = \frac{\Pi_n}{p - 1 - pr_n + r_n^2}\,, \tag{A82}$$

So that

$$\Phi_{t,0} = \frac{1}{p-1}, \qquad \Phi_{t,1} = \frac{p+1}{p^2 - p - 1}, \qquad \Phi_{t,2} = \frac{p^2 + p}{p^3 - p^2 - 2p + 1},$$

etc. and

$$V_{S,t} = (p-2)V\sum_{n=0}^{\infty}\Phi_{t,n}\,. \tag{A83}$$

On the tip, obviously $V_{S,t} = V$; consequently the infinite sum evaluates to $(p - 2)^{-1}$. This result can be expressed by a very simple function of the first term:

$$\frac{1}{p-2} = \frac{\Phi_{t,0}}{1 - \Phi_{t,0}}\,. \tag{A84}$$

Thus at the tip, the infinite sum can be exactly expressed. For the trivial case $s = 0$, (where $V_S = 0$), the same approach applies. At both endpoints of the range $0 < s < z_t - R$, working with only the first term of the potential gives an exact result. It is therefore proposed that for the *whole* range of s,

$$V_S = \frac{2sV}{R} V \sum_{n=0}^{\infty} \Phi_n(s) \approx (p-2)V \frac{\Phi_0}{1-\Phi_0} = \frac{2sRV}{z_t^2 - s^2 - R^2}. \tag{A85}$$

This simple looking expression has desirable properties. One can verify that at the points $s = 0$ and $s = z_0 - R$ the correct potential is returned. The potential vanishes if the tip is very remote ($z_t \to \infty$). If we write $z = R + a$, with a the distance between the edge of the tip and the metal, we see that for infinite tip radius $R \to \infty$ we arrive at $V_S \to sV/a$, the potential at S in a parallel plate configuration. Last but not least, if we consider the force on the tip due to this interaction, $F = -\frac{d}{dz_t} qV_S$, we observe a combination qs originates, reminding us of the fact that observing a trapped charge is in fact a measurement of a *dipole*. This character is stronger for small s. This explains also the difficulty of resolving charge magnitude and depth, as signaled in Chapter 5.
The accuracy of this approximation is plotted in Fig. A17.

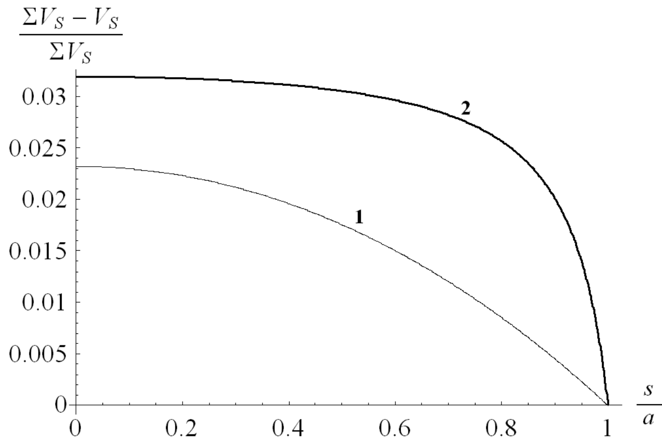


Fig. A17 The scaled difference between the numerical infinite sum of the potential (denoted ΣV_S) and the approximated potential V_S against relative s scaled to the spacing $a = z_{tip} - R$., for 1) a small p (p =2.3) and 2) a large p (p = 28).

The approximation is better for small p and in fact very accurate over the whole range of s, as is demonstrated for $p = 2.3$ (for example a tip with $R = 100$ nm, $z_{tip} = 140$ nm, so a spacing a of 15 nm). But even for a much larger p of 28 (e.g. $R = 10$ nm, $z_{tip} = 140$ nm, so $a = 130$ nm), the accuracy is not much worse. The ‘inaccurate’ part is only less favorably distributed; it improves only little for small s, close to the metal surface.

A2.2.3 Reflection potential

It remains to treat the reflection potential of the charge, which is the potential that a charge induces at its own location as a result of surrounding grounded conductors (also dielectrics) reacting on the charge. Also in this case, earlier derived results can be consulted. The situation is a bit more complicated in the present case. Fig. A18 shows a schematic aimed at clarifying the reflection interaction.

The first contribution to the reflection potential V_{refl} in S is given by the mirror of the charge q' in the metal surface. This term amounts to $-q/8\pi\varepsilon_0 s$. Then, the tip (upper, big circle) sees the charge q and its mirror in the flat electrode q'. The tip itself is mirrored in the metal (lower big circle), which is imaged in the tip again, etc. Hence there are five regions that supply a contribution to the reflection potential.

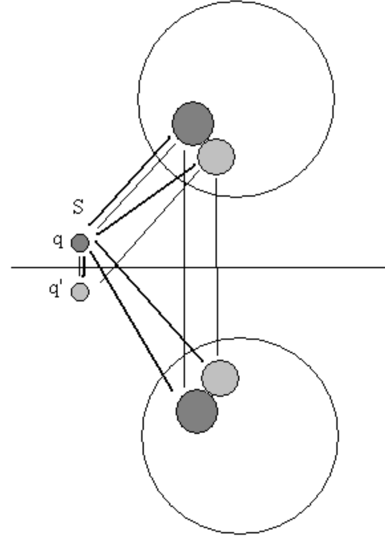


Fig. A18 reflection interaction; The large grey circles are 'clouds' of image charges in the tip. Thin lines represent important "imaging routes", thick lines indicate the contributions to the reflected potential in S. The tip has been moved a little away from the charge in order to better distinguish the relevant regions

The first order images in the tip due to the trapped charge q at S read:

$$Q_{f,0} = -\lambda q\,, \qquad Q_{f,1} = -\frac{1}{p-\lambda}\lambda q\,, \qquad Q_{f,2} = -\frac{1}{p-\frac{1}{p-\lambda}}\frac{1}{p-\lambda}\lambda q\,,$$

etcetera.

In the present case, $\lambda = R/(z_t - s)$. The charges in the tip are generally given by:

$$Q_{f,n} = -\lambda q \Pi_n(\lambda)\,, \tag{A86}$$

with corresponding positions:

$$z_n = z_t - Rr_n(\lambda)\,. \tag{A87}$$

Then there is a series of images in the tip resulting from $q' = -q$. For this we define $\lambda' \equiv R/(z_t + s)$, to arrive at

$$Q'_{f,n} = -\lambda' q' \Pi_n(\lambda') = \lambda' q \Pi_n(\lambda')\,, \tag{A88}$$

$$z'_n = z_t - Rr_n(\lambda')\,. \tag{A89}$$

The tip contribution to the voltage in S is then retrieved by these tip charges $Q_{f,n}$ and $Q'_{f,n}$ and their corresponding positions z_n and z'_n. The final share of the reflection potential comes from the tip *mirrored* in the $z = 0$ plane. This results in charges $-Q_{f,0}$ and $-Q'_{f,0}$ at positions $-z_n$ and $-z'_n$. The total reflection potential thus reads:

$$V_{refl}(S) = \underbrace{\frac{1}{4\pi\varepsilon_0} \times \frac{-q}{2s}}_{\text{mirrored charge}} + \frac{1}{4\pi\varepsilon_0} \sum_{n=0}^{\infty} \left(\underbrace{\frac{Q_{f,n}}{z_n - s} + \frac{Q'_{f,n}}{z'_n - s}}_{\text{tip}} \underbrace{- \frac{Q_{f,n}}{z_n + s} - \frac{Q'_{f,n}}{z'_n + s}}_{\text{mirrored tip}} \right)$$

$$= \frac{-q}{8\pi\varepsilon_0 s} + \frac{2qs}{4\pi\varepsilon_0} \sum_{n=0}^{\infty} \left(-\frac{\lambda \Pi_n(\lambda)}{z_n^2 - s^2} + \frac{\lambda' \Pi_n(\lambda')}{z'^2_n - s^2} \right). \quad \text{(A90)}$$

The height-dependence of the reflection potential in s of a single electron is shown in Fig. A19. The tip radius R is 100 nm. The center z_t is at 140 nm, so that the spacing between tip and sample $z_t - R$ is 40 nm. The charge is then placed at 0-40 nm from the bottom. The end points of the graph have infinite values, because the charge comes very close to its most direct image then. Finally, the reflection potential is proportional to the amount of trapped charge q, hence the graph can be vertically scaled according to this amount.

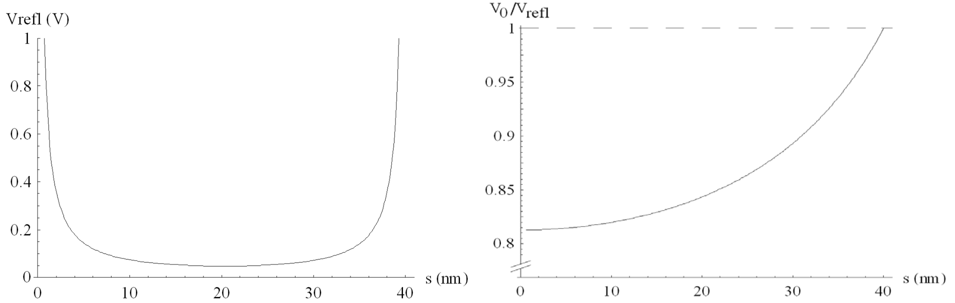


Fig. A19 (left) Reflection potential vs. position of charg, including the mirrored charge. Fig. A20 (right) Quality of approximation of reflection potential. Here V_{refl} is <u>*without*</u> *mirror charge term*

The charge-charge force F_{refl} on the tip can be calculated by evaluating the z-derivative of the reflection energy

$$F_{refl} = -\frac{dU_{refl}}{dz_t} = -\tfrac{1}{2} q \frac{d}{dz_t} V_{refl} \quad \text{(A91)}$$

In the calculation of the force *on the tip,* the first term in the reflection potential, which stems from the mirrored charge, plays no role. It is independent of the tip coordinate z_t. This term is only important if one is to calculate the force *on the charge*.
In the contexts of the bias-bias interaction ("capacitance") and the charge-bias interaction, expressions have been provided as approximations to the infinite sum. The simplest procedure is to take just the $n = 0$ - term of the reflection potential, cf. Eq. (A90). Omitting the contribution from the mirrored charge would then give:

$$V_0 = \frac{q}{\pi\varepsilon_0} \left[\frac{s^2 R(z_t^2 + R^2 - s^2)}{(R^2 + s^2 - z_t^2)((s - z_t)^2 - R^2)((s + z_t)^2 - R^2)} \right]. \quad \text{(A92)}$$

This approximation is compared with the summed reflection potential Eq. (A90) without the mirrored charge. This term is omitted, in order to highlight only the terms relevant for the force. This calculation is shown in Fig. A20, in which the same parameters are used for Fig. A19. The approximation is always reasonably good. It is 'worst' close to the sample (metal interface) and it is perfect close to the tip, at s = 40 nm. For higher values of $p = 2z_t/R$ (small tip, large spacing), the approximation Eq. (A92) improves. For p = 10, the 'worst end' at s = 0 amounts to $V_0/V_{\text{refl}} = 0.91$.
Finally we remark that the first energy term $U_{\text{refl},0} = ½qV_0$ gives a factor $(qs)^2$, which highlights that this contribution is essentially a *dipole-dipole* interaction. In this way it is seen that the mirror charge q' is still important for the interaction, because it forms a dipole with the original trapped charge q.

A2.3 DIELECTRIC EXTENSIONS

A2.3.1 Dielectric layer

Until now, the three interactions have been calculated for a charge floating above a metallic substrate. In practice, this charge is embedded in a dielectric layer. Not only does this modify the apparent strength of the trapped charge, but also there are infinitely many reflections introduced.
For charges above the dielectric layer, these are the bias charge and all image charges in the tip, there are also infinitely many reflections because of the dielectric layer. Apart from that, these tip charges will have direct mirrors in the dielectric layer, just as if it were a metal surface, only with a weaker mirror.
It has been attempted to provide approximate expressions for the three interactions VV, qV and qq in the case of no dielectric layer, or equivalently ε = 1. There is no convenient way to explicitly sum the interaction of the tip with the chain of infinite reflections caused by the dielectric layer. There is however a different handy approximation available, based on a coordinate transformation. This transformation entails:

$$\begin{aligned} z &\rightarrow z / \varepsilon & & 0 < z \leq d \\ z &\rightarrow z - d + d / \varepsilon & & z \geq d \end{aligned} \tag{A93}$$

In fact it comes down to squeezing the dielectric layer with a factor ε and solving the vacuum problem afterwards. The transformation affects the geometrical parameters the oxide thickness d, the location of the charge s and the position of the centre of the tip z_t accordingly:

$$d \rightarrow d / \varepsilon, \qquad s \rightarrow s / \varepsilon, \qquad z_t \rightarrow z_t - d + d / \varepsilon \tag{A94}$$

The radius of the tip R and the spacing $a = z_t - R - d$ are invariant. The latter is a derived parameter, being the distance from the tip apex to the surface of the dielectric, which after the transformation has dielectric constant ε = 1.

In a true vacuum problem, one would prefer to associate the tip spacing towards the metal surface. In this interpretation, the derived spacing $z_t - R$ *has* changed, but indeed because of a reinterpretation. Finally, the important p parameter changes accordingly:

$$p = 2z_t/R \;\rightarrow\; p - d(1 - \varepsilon^{-1})/R. \tag{A95}$$

For weak dielectrics, with a low dielectric constant, the change in force on the tip is limited, because the chain charges and the direct mirror are weak anyway. This motivates the simplification in these cases.
For strong dielectrics with a high dielectric constant, the interaction forces *do* change. However, the additional force induced by the coordinate transformation is less height-dependent for stronger dielectrics. The force gradient therefore remains relatively unaffected after the coordinate transformation: $dF(z)/dz \rightarrow d(F(z) + F_0)/dz \sim dF(z)/dz$ (note: d is the differential symbol here), in which the additional force F_0 depends only very weakly on z. The latter point is sought to illustrate in Fig. A21.

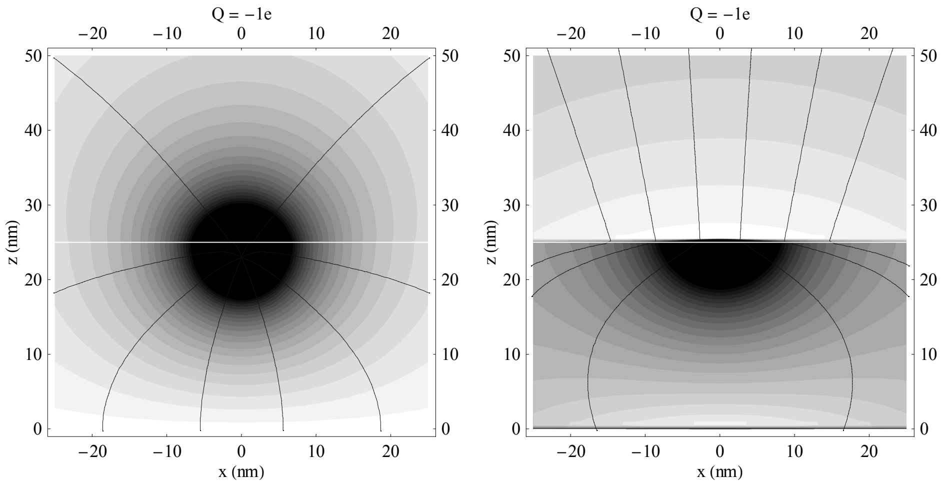


Fig. A21 Left: Electric potential and field of charge + sample; Right: Reflection (or response) potential and field of the sample alone.

On the left, there is a negative unit charge in a dielectric with $\varepsilon = 10$ and $d = 25$ nm. This dielectric fills the lower half of the plot and is on top of a metallic substrate ($z < 0$). The top half of the plot is vacuum. The potential contours and electric field lines are drawn. They form the electrostatic "map" of the charge, with which a biased tip is going to interact. The field lines in the vacuum part ($z > 25$ nm) run apart. The total field is hence clearly dependent on the vertical coordinate. On the right, the contribution of the charge *itself* to the potential and field is subtracted. Shown here is the potential and the field lines from the dielectric layer and the metallic substrate, *in response* to the charge. It is seen that this contribution of the sample to the electrostatics does not have such a large spread in the vacuum part; the sample field is rather homogeneous. Although the *force* on the tip may change by the dielectric layer, the *force gradient* changes only very little. Hence in the calculation of force *gradients*, the coordinate transformation is permissible as a simplification of the calculations for accounting for the dielectric later.

As we have seen, there are various expressions conceivable to estimate the capacitance between a closely spaced tip and a sample with a dielectric layer. Before comparing them, it is helpful to define

$$r_\varepsilon \equiv \tfrac{1}{2} p_\varepsilon \pm \tfrac{1}{2}\sqrt{p_\varepsilon^2 - 4} \qquad\qquad p_\varepsilon \equiv \frac{2(z_t - d + d/\varepsilon)}{R} \tag{A96}$$

as the coordinate-transformed versions of r and p respectively. The total capacitance (from Eq. (A73)) is then:

$$C_\varepsilon = 4\pi\varepsilon_0 R\left(\frac{1}{r_\varepsilon^2} - 1\right)\sum_{n=1}^{\infty}\frac{r_\varepsilon^n}{1 - r_\varepsilon^{2n-1}}\,. \tag{A97}$$

Its boundaries indicated by Eq. (A72) are:

$$C_{\min} = 4\pi\varepsilon_0 R(1 + r_\varepsilon) \tag{A98}$$

$$C_{\max} = 4\pi\varepsilon_0 R\frac{1}{1 - r_\varepsilon} \tag{A99}$$

Their average is:

$$C_{av} = 4\pi\varepsilon_0 R\left(\frac{1 - \frac{1}{2}r_\varepsilon^2}{1 - r_\varepsilon}\right) \tag{A100}$$

Another version is to take the first term of the infinite sum Eq. (A97) and then add the minimal value of the remaining terms of the sum. The “between” result is

$$C_{bet} = 4\pi\varepsilon_0 R\left(\frac{1}{1 - r_\varepsilon}\right)\left(\frac{1}{1 + r_\varepsilon^2}\right) \tag{A101}$$

Every converging infinite sum can be approximated by an integral. For a monotonously decreasing function $f(\nu)$ we have in general that

$$\int_0^\infty f(\nu)d\nu < \sum_{n=1}^{\infty} f(n) < \int_0^\infty f(\nu - 1)d\nu\,.$$

The latter ‘<’ can become ‘=’ if the argument of the integrand is shifted by a special amount less than 1. There exist a γ such that the integral of $f(\nu - \gamma)$ equals the required sum. If we apply this to the capacitance sum, we retain a closed-from expression, but with an a priori unknown parameter γ:

$$C_{\text{int}} = 4\pi\varepsilon_0 R\left(\frac{1}{r_\varepsilon^2} - 1\right)\int_0^\infty \frac{r_\varepsilon^{\nu-\gamma}}{r_\varepsilon^{-2} - r_\varepsilon^{2(\nu-\gamma)}}d\nu = 4\pi\varepsilon_0 R\left(\frac{1}{r_\varepsilon^2} - 1\right)\frac{r_\varepsilon \operatorname{arctanh}(r_\varepsilon^{1-\gamma})}{\log(1/r_\varepsilon)}\,. \tag{A102}$$

For a suitable γ the integrated formula approximates the capacitance very well over a significant range of tip spacings a.

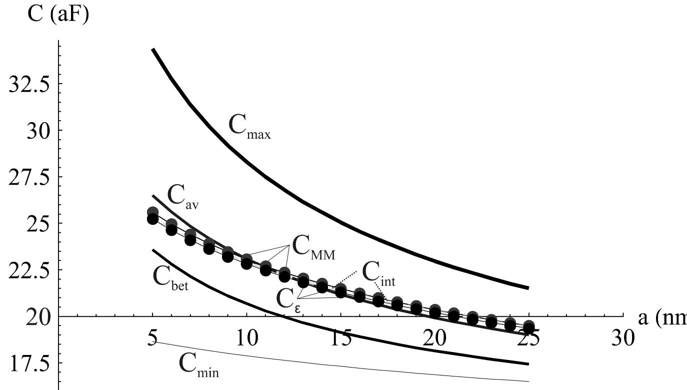


Fig. A22 Capacitance approximations, compared to a pure multi-mirror expression C_{MM}, retaining numerous charge reflections. The integral approximation C_{int} has been assigned $\gamma = 0.44$ and lies between C_{MM} and C_ε over the whole indicated range of tip spacing a. Tip radius R = 100 nm, thickness dielectric = 25 nm, $\varepsilon = 9.1$

In Fig. A22 all capacitive expressions are compared for close tip spacings *a*.
Most approximations do a reasonable to good job in approximating the true multimirror capacitance C_{MM}. It is seen that the coordinate transformation C_ε (round black dots) approximates C_{MM} very well, provided the full (numerical) sum is taken. Compare for example C_{min}, which represents only the first term. Although even in this case the capacitance value is not badly predicted by the latter, the force (spatial derivative of *C*) is. It can be seen from the graph that the slopes at *a* = 5 nm are quite dissimilar. This is illustrated even better in Table A1, in order of quality of approximation.

ΔC_{MM}	ΔC_ε	ΔC_{int} (ν=0.44)	ΔC_{bet}	ΔC_{av}	ΔC_{max}	ΔC_{min}
-0.629	-0.601	-0.599	-0.782	-0.905	-1.631	-0.179

Table A1, Capacitance differences between a = 5 nm and a = 6 nm

Thus the point stated underneath Eq. (A79) is understood.

A2.3.3 Dielectric substrate

The multimirror model has been developed for a dielectric layer on a *metallic* substrate. It can be readily extended however to encompass a *dielectri*c substrate as well, provided that this substrate is conducting, i.e. its potential relative to the tip should be definable by connecting it to a voltage source. The immediate example is a silicon substrate, which has a dielectric constant of $\varepsilon = 11.9$. The so far encountered dielectric layers (SiO_2, Al_2O_3) have lower dielectric constants, but it is entirely possible that the dielectric layer has a dielectric constant that is *higher* than that of the substrate.
To accomplish this, the apparent strength of the virtual charges changes, their positions remain unaltered.

Firstly, the dielectric mirror Eq. (A9) is unaffected, because no mirroring in the substrate is involved. Furthermore, the expressions (A10) and (A13) can be conveniently combined in a single expression. The interpretation then comes down to counting the number of internal reflections between the layer surface and the substrate/layer – interface. The analogue of the combination of Eqs. (A10) and (A13) then becomes

$$Q_n = (-k_s k)^n\, k_s(1 - k^2)Q_V \qquad (n \geq 0). \qquad \text{(bias charge)} \qquad \text{(A103a)}$$

And the analogues of the trapped charge q chains Eqs.(A15ab) and Eqs. (A17ab) become:

$$q_{n,1} = (-k_s k)^n(1 - k)q \qquad (n \geq 0), \qquad \text{(trapped charge, series 1)} \qquad \text{(A103b)}$$

$$q_{n,2} = -(-k_s k)^n\, k_s(1 - k)q \qquad (n \geq 0). \qquad \text{(trapped charge, series 2)} \qquad \text{(A103c)}$$

Here

$$k_s \equiv \frac{\varepsilon_s - \varepsilon}{\varepsilon_s + \varepsilon}$$

is the dielectric coupling between the substrate and the layer. In here, ε_s and ε are the dielectric constants of the substrate and the layer respectively. For a metallic substrate, $\varepsilon_s \rightarrow \infty$, when $k_s \rightarrow 1$. The original charges Eqs. (A10), (A15), (A15ab) and (A17ab) are returned then.

A2.4 CALCULATION METHODS

The Multimirror model relies on large amounts of image charges, as explained on the basis of the Appendix equations through to Eq. (A18) and in Chapter 5. More mirror charges obviously yield a more precise answer for the force and force gradient on the tip. Of course the computer's resources and the human operator's patience are limited, so below there are a few pros and cons of methods to calculate the force from these charges.

1) Direct point charge force

The force on the tip is calculated by the interaction of every single image charge and bias charge with all charges outside the tip. There are infinitely many; the infinite series have to be truncated for a practical calculation. The gradient is then determined by calculating the force again, after a small vertical displacement of the tip.
The advantage is that this direct force method converges relatively quickly, i.e. the cyclic process steps of imaging charge in the tip, mirroring it in the sample, image that mirror, etc. Also, the force between the charges that are taken into account is calculated exactly. The disadvantage is that the computing time scales with N^2, with N the number of charges. For a not-too-complicated case, N is in the order of several thousand.

2) Maxwell stress tensor

A convenient method to compute the force on a collection of charges, such as an AMF tip, uses Maxwell's electrostatic stress tensor T_M. In absence of magnetic fields, this tensor is given by

$$T_{M,ij} \equiv \varepsilon_0 E_i E_j - \tfrac{1}{2}\varepsilon_0 \delta_{ij} E^2 \tag{A104}$$

Here i and j label the Cartesian coordinates x, y, z. The Kronecker delta δ_{ij} is defined as $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ otherwise. E_i are the components of the total electric field and E is its norm. The force on a volume Ω can then be calculated by integrating the stress tensor over its boundary $\partial\Omega$:

$$\vec{F} = \oint_{\partial\Omega} \vec{T}_M \cdot d\vec{a}\ . \tag{A105}$$

This method involves the square of the electric field. The electric field depends only linearly on the number of charges. This is an obvious advantage. The computing time scales then as $N \times M$, with M the number of integration elements over the tip surface. This partly undoes the gain in time, because for good precision M should not be taken too small. This is the price to pay, because the force between the charges
This method is most advantageous when the tip is directly above the charge. The problem is then axially symmetric. The integration involves only the zenith angle θ. Discrete steps in this angle divide the surface of the spherical tip in ring-shaped integration surface elements of width $Rd\theta$. It is found that between 50 and 200 of these rings are usually sufficient, while still quite good first approximations are retrieved by less. When the tip is not in the zenith of the trapped charge, there is only planar symmetry. Even then, the stress tensor calculation is often less time-consuming then the direct force method. Obviously, because the integration is done numerically[45], the forces between the involved charges are not calculated exactly, unlike the direct force method.

3) Energy

In the foregoing text, it has been considered how the composition of the charges builds up electrostatic energy. Also here computation time scales only with N. There are no integration elements involved, so the energy is calculated exactly and very quickly. But also here there are effects that partially cancel the time gains. First, to derive the force and the force gradient, an extra gradient step has to be taken with respect to methods 1) and 2). This multiplies the calculation time by 1½, but more importantly, for each step the energy has to be calculated relatively accurately. This requires involving more charges than needed for the force methods. Furthermore, the energy converges considerably slower than the forces: the cycle of imaging in the tip and mirroring in the sample has to be repeated more often before a stable solution is reached. For extremely close spacings (roughly $a < 0.1R$), up to 10^8 charges or more might be needed. On an ordinary PC (2.4 GHz dual core), calculation times become limiting factors again, and even physical memory, because of the large arrays containing the charges. By chopping

[45] A symbolic expression of the force integral does exist in principle, but its monstrous complexity puts this road drastically out of question.

up the arrays, the latter factor can be suitably addressed in the programming code, but the time factor is harder. Still, in general this energy method is clearly the fastest and accurate.

4) Approximation models

As briefly discussed in this Appendix, it is possible in many cases to approximate the energy in 3) by ready-made, closed-from expressions or simple numerical summations. Of course, the time consumption is negligible compared to the aforementioned methods. However, one has to check that the approximations are valid for a particular geometry or problem. One such case has already been mentioned: The capacitance can be conveniently determined by some expression, but in order to derive from it a reliable figure for the force on the tip and even more its gradient, the capacitance must be calculated quite accurately, typically down to a few per cent or better.

REFERENCES

[1] D. J. Griffiths, "Introduction to electrodynamics 2nd edition", Prentice Hall London (1989)
[2] E. Durand, "Electrostatique tome II Problèmes généraux conducteurs", Masson et Cie, Paris (1966)
[3] J. D. Jackson, "Classical Electrodynamics 3rd edition", John Wiley & Sons, New York, USA (1998)
[4] E. Philippow, "Grundlagen der Elektrotechnik", Hüthig Heidelberg (1989)

SUMMARY

Brief overview

In this thesis, the phenomenon of charge trapping in dielectric layers has been addressed via two approaches. First and foremost, the subject is treated on device level: the integrated, temperature dependent effects of charge trapping in capacitive MEMS-based sensors. Chapters devoted to the physics of MEMS and more specifically charge trapping are Chapters 2, 3, 4 and 6. Chapter 2 covers introductory theory of MEMS, specific topics relevant for the 'empirical' chapters 3 and 4 and finally discusses parasitic electrostatic forces. Chapter 3 presents temperature dependent capacitive-voltage measurements on available real devices (RF power sensors). Chapter 4 presents the design, fabrication and mechanical characterization of capacitive MEMS structures according to a novel fabrication concept, followed by capacitive-voltage measurements serving to extract the temperature dependent stability of the "built-in voltage", signaling (moving) parasitic charges.
Chapter 6 discusses when these trapped charges pose a limit to the performance of real devices: a gravity gradiometer and the RF power sensor of Chapter 3.
The second approach is aimed at focusing on trapped charges at a differentiated, fundamental level: Studying localized trapped charges by conducting AFM. This is subsumed by Chapter 5 and the Appendices. Chapter 5 explains the principles of Force modulated, non-contact, conducting AFM and connects directly surface interactions to AFM read-out quantities. Experimental work by Marko Sturm is shortly presented. Further, it contains the results of a new model covering the electrostatic interaction between the conducting tip of an AFM and a metallic substrate with a dielectric layer containing a localized charge. This model is compared to an existing model from literature, showing strikingly different predictions. This new "multi mirror model" is supported on theoretical grounds, by finite element simulations and by empirical considerations.
The appendices digress on the foundations of the multi mirror model and provide the theoretical justification. Approximation schemes are discussed that provide under many circumstances a viable shortcut to the sometimes involved proper multi mirror calculations.

Summary with conclusions

Chapter 2 starts with introductory theory of capacitive MEMS including statics, pull-in voltage and an account of how dielectric layers alter the basic expressions (section 2.1). Section 2.2 is dedicated to theory of a clamped-clamped beam deflected by an electrostatic load. Many accounts assume a uniform load, which however underestimates the force after attraction. The electrostatic force depends on the actual capacitive gap and is thus dependent on the deflection profile. This mutual influence is nonlinear and the very reason for an unstable point (pull-in), as is demonstrated. Important in this section is the influence of an axial load on the beam stiffness. Specifically, thermal stress is covered, which is important for analysis performed in section 3.3. Section 2.3 introduces the well-known concept of "built-in voltage", which is defined as the shift of the minimum of a capacitance-voltage curve away from the vertical axis: a nonzero bias

voltage is needed in order to compensate for some remnant voltage and nullify the electrostatic force. This section provides an overview of possible causes: dissimilar work functions of the electrodes present in a device are in for a static contribution. Very important are however parasitic trapped charges, which influence the device's characteristics in several ways. *Static uniform* charges introduce additional forces by interacting with their counterparts on the opposite electrode, their interaction with a bias voltage and by inducing image charges on the opposite electrode or dielectric and interacting with it. Furthermore, a *static, non-uniform* distribution also introduces forces, even if the net total charge is zero. Then, during operation of a MEMS device, charges move along the varying bias voltage. It is demonstrated that this can result in a built-in voltage. Finally, an overview of types of trapped charges is presented, as well as relevant transport mechanisms. Section 2.4 discusses the dynamics of a MEMS bridge under influence of an oscillating voltage through the use of equivalent circuits. This coverage is used for the mechanical characterization of newly designed and made devices in Chapter 4. Finally, section 2.5 touches a few situations that might influence proper capacitive MEMS behavior. Firstly, if the electrodes of a plate capacitor are non-parallel, characteristics derived from a capacitance-voltage curve can be misinterpreted. Secondly, the Casimir force is briefly mentioned. It is calculated that this contribution is negligible, although experiments are designed aimed at employing very sensitive capacitive MEMS-based sensors in the very measurement of this force.

Chapter 3 first shortly describes the deployed measurement set-up and the principle of capacitance-voltage measurements. In section 3.2 measurements performed on a type of a MEMS RF power sensor are presented. Its mechanical characteristics show clear dependence on the temperature. Meanwhile, the presence of thick dielectric layers (1 µm of silicon nitride) offers plenty of opportunity for parasitic charges to accumulate and disturb the device's proper operation. The RF power sensor studied in section 3.3 shows even worse (though well repeatable) mechanical response toward thermal variations. Thermal stress is involved in this discussion. Analysis according to clamped-clamped beam theory (section 2.2) however considerably overpredicts this influence. An explanation to this might be that the suspension points are flexible, rather than absolutely immovable. The thin dielectric layer of this device (2.5 nm native Al_2O_3) still hosts detectable parasitic charges. The thermal instability of this sensor however prevents an account of the temperature dependence of the instability of these charges. To demonstrate these relatively small effects in a widely employed electrode like aluminum with native oxide, structures had to be designed with considerably enhanced thermal immunity.

Chapter 4 reviews in section 4.1 in quite detail the design and the fabrication process of a thermally stable MEMS test structure. Special attention is devoted to the etching process, which is a novel approach in bulk micromachining. Though the KOH-etching process itself is everything but new, it has for the first time been employed to fabricate the beam springs that maintain parallelity of the capacitive electrodes upon actuation. The process is well-controlled. The first offspring of this process had been tested and proven proper capacitance behavior, cf. section 4.2 A discrepancy between capacitance and sensitivity could be explained upon closer inspection of the beam springs, which were effectively shortened because of the presence of still unremoved silicon pillars. Extraction of the spring constant, through resonant, bias dependent behavior, is the subject of section 4.3. Resonant behavior is only observable in concurrence with a bias voltage, which however decreases the effective stiffness. By measuring at various voltages, the intrinsic spring constant could still be distilled. It was found in good agreement with expectation. The translation between mechanical quantities (mass, spring constant, damping) and their

corresponding equivalent circuit elements (induction, capacitance and resistance respectively) however showed a uniform, persistent mismatch, which could not be resolved. Finally, this section mentions a single measurement series with varying pressure, with indications of squeezed film damping. The structures researched so far had however not sufficient sensitivity for temperature-varied measurements concerning trapped charge stability. Section 4.4 deals with structures that did have sufficient sensitivity, while showing excellent thermal stability, herewith approving the concepts implanted in the design. Series of capacitance voltage curves at various, stabilized temperatures (pressure: 10^{-3} mbar) were taken in order to consider two aspects of built-in voltage. The instantaneous value depends first and foremost on the direction of the voltage sweep. Alternating sweeps thus show oscillating values for this. Below 150 to 200 K, this ‘two-valuedness’ quite suddenly collapses to a single value. This has been observed repeatedly for both the RF power sensor discussed in section 3.3 and the structures considered in the current section, 4.4. The presence of aluminum electrodes is common to these structures, which are for the rest completely different. This is strongly indicative of an effect attributable of the electrode material, in particular trapped charges. Apart from this, there is the long-term behavior of the built-in voltage. Also the gradual change in this clearly diminishes at lower temperature. Section 4.5 finally contains measurements performed at high vacuum ($5 \cdot 10^{-7}$ mbar) and room temperature. Especially built-in splitting does not disappear. Ambient conditions can therefore not be fully accounted for the observed “built-in splitting”.

Chapter 5 enters the path of AFM-based research of trapped charges. In section 5.1 principles of AFM are described, gradually but quickly concentrating on Force Modulated non-contact conducting AFM. How tip-surface interactions are translated in an observable quantity (resonance frequency shift of the AFM cantilever) is given. FM conducting AFM has been used for experimental work performed by Marko Sturm (reference chapter 5) on imaging local oxide charges (section 5.2). The remainder of this chapter contains calculations and simulations concerning the electrostatic interaction between the tip of a conducting AFM and a metal electrode with dielectric layer and a localized trapped charge. The particular electrostatic problem is introduced in section 5.3, together with how it is approached by a literature model (Ludeke - Cartier and Lambert - Saint-Jean). Section 5.4 shows some results and predictions of the newly presented Multi Mirror model (for its foundations, the reader is referred to the Appendix). These include conditions for electrostatic tip-sample *repulsion* (where normally the force is attractive) and critical behavior at short distances. Special focus is on the resolvability of the magnitude and depth of trapped charge: because of the proximity of Hence the tip-charge interaction essentially assumes a dipole character. Because of its form $\mathbf{p} = q\mathbf{s}$, (infinitely) many combinations of charge q and distance s would build the same dipole p. Performing AFM measurements at various bias voltages is not expected to reveal any resolution. Measurements at varying tip height helps in some cases. A horizontal scan over the charge does not provide any solution if the radius of the tip is much larger than the relevant dimensions (thickness of the dielectric, position of the charge relative to the back electrode). However, very sharp tips (< 10 nm) at very close spacings (< 5 nm) should perform well. Section 5.5 covers finite element modeling and confirms the applicability of approximating the tip by a conducting sphere. The three approaches (Lambert & Saint-Jean, FEM and Multi Mirror) are compared in section 5.6 for a thin (2.5 nm) and a medium (25 nm) dielectric of $\varepsilon = 9.1$ (aluminum oxide) and a medium dielectric of $\varepsilon = 3.9$ (silicon dioxide). The models are compared for the quantities force, force gradient, peak height and V_{min}, with which the complete relevant electrostatic phenomenology is

covered. For force, force gradient and peak height, Multi Mirror predicts drastically higher values than Lambert & Saint Jean (sometimes more than an order of magnitude). In doing so, it feels supported by FEM. An additional important feature of the Multi Mirror model is that it predicts that Peak height and V_{min}, quantities characterizing the interaction with the trapped charge *sec*, subside to zero when the charge is sunk to the interface of the metal and the dielectric, where it recombines with its mirror charge and should escape detection. On the contrary, according to Lambert & Saint-Jean, a considerable interaction remains alive in this case. Finally, in section 5.7 it is shown that Multi Mirror performs excellent in predicting the shift of resonance frequency of an AFM cantilever. The literature model could not be matched with the observed frequency shift in a comfortable way.
At last, it is reasoned that the multi mirror model is consistent with profile measurements on charges, if it is assumed that trapped charges (charge clusters) consist of more than one unit charge (~50).
Chapter 6 discusses when charge trapping impairs the operation of real capacitive MEMS devices. The larger part of this chapter is devoted to a MEMS gravity gradiometer, which is in its design stage. It is shown that its targeted sensitivity of 0.1 Eötvös/√Hz cannot be reached if charge trapping is not given due attention. This is calculated for the static case, while also a start is made with a dynamic approach. A shorter discussion has been set up for the RF power sensor, which has been object of experimental study already in section 3.3. The same conclusion applies. This sensor has natural thin dielectric layers, which nevertheless provide opportunity for charge trapping to the extent that its target sensitivity of 1 nW/√Hz remains (far) out of reach.
Appendix I finally provide the fundamental building blocks and a theoretical footing of the Multimirror model. The finiteness of the model is demonstrated.
In Appendix II the quadratic form of the electrostatic force is derived. Additional miscellaneous results are derived, that help to accelerate the sometimes time-consuming calculations of the normal application of the multi-mirror model, by seeking suitable approximations. These results include:
A) A discussion of the capacitance between a spherical conducting tip at a certain potential relative to a conducting infinite plane. This capacitance can be calculated via an infinite series of mirror charges. Though an expression for the total capacitance, involving an infinite sum, is known from literature, three different expressions are derived here: 1) It is achieved to cast the nth mirror charge in a closed-form expression, without needing to first calculate all preceding $n - 1$ charges. 2). The sum over these charges can be rewritten in a form that converges faster. 3) It was recognized that these charges could be written in polynomials of finite degree in terms of the ratio of the tip radius R and the tip position z_t. The coefficients of these polynomials are given by certain diagonals in Pascal's triangle.
B) Suitable approximations are found for the qV and qq – interactions in absence of a dielectric layer. The dielectric layer can often approximately be taken into account by a coordinate transformation.

The multimirror model has been worked out for a metallic substrate. It is shown to be easily generalizable to the case of a dielectric substrate, i.e. with a finite relative dielectric constant. At last, a brief discussion on the pros and cons of methods that calculate the multimirror model finishes this appendix.

SAMENVATTING (IN DUTCH)

Kort overzicht

In dit proefschrift wordt het fenomeen van parasitaire lading in diëlectrische lagen bestudeerd volgens twee benaderingen. Eerst en vooral wordt dit bestudeerd op het niveau van een complete sensor: de geïntegreerde, temperatuursafhankelijke effecten van parasitaire lading in capacitieve MEMS sensoren. Hoofdstukken gewijd aan de fysica van MEMS en meer specifiek parasitaire lading zijn 2, 3, 4 en 6.
Hoofdstuk 2 beslaat basistheory van MEMS, specifieke onderwerpen relevant voor de 'experimentele' hoofdstukken 3 en 4 en behandelt tenslotte parasitaire electrostatische krachten. Hoofdstuk 3 behandelt temperatuursafhankelijke metingen aan beschikbare, werkelijke sensoren (RF-vermogen sensoren). In hoofdstuk 4 worden ontwerp, fabricage en mechanische karakterisatie van capacitieve MEMS structuren gepresenteerd, gebaseerd op een nieuw fabricageconcept. Dit wordt gevolgd door metingen van de capaciteit tegen aangelegde spanning, ten dienste van het extraheren van de temperatuurafhankelijke stabiliteit van de "built-in spanning", welke een (mobiele) parasitaire ladingen. Hoofdstuk 6 beschouwt wanneer deze parasitaire ladingen een limiet betekenen voor de prestaties van werkelijke sensoren: een gravitatiegradiometer en de RF-vermogen sensor van Hoofdstuk 3.
De tweede benadering is gericht op het bestuderen van parasitaire ladingen op een locaal, fundamenteel niveau, door middel van geleidende AFM. Dit is opgenomen in Hoofdstuk 5 en de Appendices. In Hoofdstuk 5 worden de principes van Force modulated, non-contact, geleidende AFM uitgelegd en worden direct oppervlak-interacties gerelateerd aan AFM uitleesgrootheden. Experimenteel werk door Marko Sturm wordt kort besproken. Verder bevat het de resultaten van een nieuw model dat de electrostatische interactie beschrijft tussen een geleidende AFM-tip en een metallisch substraat met een dielectrische laag waarin zich een locale lading bevindt. Dit model wordt vergeleken met een ouder model bekend uit de literatuur. De voorspellingen zijn zeer verschillend. Het nieuwe "multi mirror model" wordt ondersteund door theoretische argumenten, finite element simulaties en empirische overwegingen. De appendices wijden uit over de fundamenten van het multi mirror model en verschaffen de theoretische rechtvaardiging. Er worden benaderingsschema's voorgesteld die voor veel situaties de soms tijdrovende berekeningen van het multi mirror model kunnen versnellen.

Samenvatting en conclusies

Hoofdstuk 2 begint met het introduceren van capacitieve MEMS, incusief statica, pull-in spanning en hoe dielectrische lagen de basisgrootheden beïnvloeden (sectie 2.1).
Sectie 2.2 is gewijd aan de theorie van een tweezijdig vastgeklemde balk die buigt onder een electrostatische last. Vaak wordt een uniforme lastverdeling aangenomen, hetgeen echter de kracht onderschat na buiging. De electrostatische kracht hangt af van de momentane capacitieve verwijdering en is dus afhamkelijk van het buigingsprofiel. Deze wederzijdse invloed is niet-lineair en precies de oorzaak van een instabiliteit (pull-in), zoals wordt gedemonstreerd. Belangrijk in deze sectie is de invloed van een axiale trekspanning op de stijfheid van de balk. Specifiek wordt ingegaan op thermische spanning, hetgeen van belang is voor een analyse die in Sectie 3.3. gemaakt wordt.

Sectie 2.3 introduceert het bekende concept “built-in spanning”, welke is gedefinieerd is als de verschuiving langs de spannings-as van het minimum van een capaciteit-tegen-spanning grafiek: een biasspanning ongelijk aan nul moet worden aangelegd om te compenseren voor een remanente spanning en aldus de electrostatische kracht per saldo te nullificeren. Deze sectie geeft een overzicht van de mogelijke oorzaken: ongelijke uittreepotentialen van de electrodes in een sensor dragen bij aan een statische component. Zeer belangrijk zijn echter parasitaire ingevangen ladingen, welke de karakteristieken van de sensor op verschillende manieren beïnvloeden. *Statische uniforme* ladingen introduceren additionele krachten door te interageren met hun tegenhangers op de tegenovergeplaatste electrode, door hun interactie met een biasspanning en door interactie met geïnduceerde spiegellading op de andere electrode of de dielectrische laag. Verder veroorzaakt een *statische, niet-uniforme* lading ook krachten, zelfs wanneer de netto lading nul is. Bovendien bewegen tijdens de werking van een MEMS ladingen mee met een variërende biasspanning. Het wordt gedemonstreerd dat dit kan resulteren in een built-in spanning. Tenslotte wordt een overzicht gegeven van de typen van parasitaire lading en relevante transportmechanismen.
In sectie 2.4 wordt de dynamica besproken van een MEMS-brug onder invloed van een oscillerende spanning via equivalente circuits. Dit is relevant voor de mechanische karakterisatie van nieuw ontworpen en vervaardigde MEMS in Hoofdstuk 4.
Tenslotte stipt Sectie 2.5 enkele situaties en effecten aan die afwijken van het standaard gedrag van een MEMS. Ten eerste, als de electrodes van een plaatcondensator MEMS niet parallel zijn, kunnen de karakteristieken afgeleid van een capaciteit-tegen-spanning grafiek onjuist worden geïnterpreteerd. Ten tweede, de Casimirkracht wordt kort genoemd. Er wordt berekend dat deze bijdrage verwaarloosbaar is, hoewel er experimenten worden ontworpen die juist capacitieve MEMS sensors gebruiken om deze kracht te meten.

Hoofdstuk 3 beschrijft eerst kort de meetopstelling en het principe van capaciteit-tegen-spanning metingen.
In Sectie 3.2 worden metingen gepresenteerd uitgevoerd aan een type RF MEMS sensor. De mechanische karakteristieken blijken duidelijk afhankelijk van de temperatuur. De aanwezigheid van dikke dielectrische lagen (1 μm silicium nitride) biedt ruim gelegenheid voor accumulatie van parasitaire lading die de werking van het MEMS verstoort.
De RF-vermogen sensor bestudeerd in Sectie 3.3 heeft zelfs een ongunstigere (hoewel reproduceerbare) mechanische respons op temperatuurfluctuaties. Thermische stress wordt in deze analyse betrokken. Analyse volgens de theorie van een tweezijdig vastgeklemde balk (Sectie 2.2) overschat deze invloed echter. Een verklaring hiervoor zou kunnen zijn dat de ophangpunten flexibel zijn in plaats van absoluut onbeweeglijk. Betreffende parasitaire lading: zelfs in de dunne dielectrische laag van dit MEMS (2.5 nm natuurlijk Al_2O_3) is die meetbaar aanwezig. De thermische instabiliteit van deze sensor verhindert echter het temperatuurafhankelijke gedrag van deze ladingen vast te stellen. Om deze relatief kleine effecten te demonstreren in een veelgebruikte electrode zals aluminium met natuurlijk oxide, moesten MEMS structuren worden ontworpen met aanzienlijk verbeterde thermische stabiliteit.

Hoofdstuk 4 beschouwt in sectie 4.1 in detail het ontwerp en het fabricageproces van een thermisch stabiele MEMS teststructuur. Speciale aandacht heeft het etsproces, welke een nieuwe benadering is in “bulk micromachining”. Hoewel het KOH-etsproces zelf beslist

niet nieuw ism is het voor het eerst gebruikt om bladveren te fabriceren die evenwijdigheid van de capacitieve electrodes behouden wanneer een electrostatische kracht aanwezig is. Het proces is goed gecontroleerd.
De eerste structuren uit dit proces zijn getest en het capaciteitt-tegen-spanning gedrag is correct bevonden, zie Sectie 4.2. Een discrepantie tissen capaciteit en gevoeligheid kon worden verklaard na nauwkeuriger inspectie van de bladveren, die effectief korter waren vanwege de aanwezigheid van onbedoelde dwarsverbindingen tussen de veren.
Onderwerp van sectie 4.3 is het bepalen van de veerconstante via meting van resonant, biasspanning-afhankelijk gedrag. Resonant gedrag is alleen zichtbaar wanneer een biasspanning wordt aangelegd. Dit vermindert echter de effectieve stijfheid. Door bij verschillende spanningen te meten kan de intrinsieke veerconstante niettemin worden gedestilleerd. Het bleek in overeenstemming met wat verwacht werd. De vertaalslag tussen mechanische grootheden (massa, veerconstante, demping) en hun corresponderende elementen van een equivalent circuit (respectievelijk inductie, capaciteit en weerstand) lieten echter een uniforme en consequente discrepantie zien, welke niet opgelost kon worden. Tenslotte wordt in deze sectie een enkele meting genoemd waarbij de omgevingsdruk werd gevarieerd, die het optreden van squeezed film damping doet vermoeden. Echter, de structuren waren nog onvoldoende gevoelig om de temperatuurafhankelijkheid van parasitaire ladingseffecten te meten.
Sectie 4.4 beschrijft structuren die voldoende gevoelig zijn en waarvan de thermische stabiliteit aangetoond is. Hiermee is een essentieel doeleinde gehaald. Bij verscheidende, gestabiliseerde temperaturen (druk: 10^{-3} mbar) zijn series van C(V)-curven gemeten teneinde twee aspecten van built-in spanning te demonstreren. De instantane waarde hangt ten eerste af van de richting van de spannings-sweep. Afwisselende sweep-richtingen laten oscillerende waarden zien voor de built-in spanning. Beneden 150 à 200 K valt deze “tweewaardigheid” plotseling terug naar een enkele waarde. Dit is herhaaldelijk gezien voor zowel de RF-vermogen sensor van sectie 3.3 als de structuren in deze sectie, 4.4. Deze structuren hebben gemeen dat aluminium het electrodemateriaal is; voor de rest zijn ze totaal verschillend. Dit is een sterke aanwijzing voor dat het electrodemateriaal debet is aan dit effect, in het bijzonder het invangen van ladingen. Verder is er het lange-termijngedrag van de built-in spanning. Ook de geleidelijke verandering hierin wordt duidelijk langzamer bij lagere temperatuur.
In sectie 4.5 tenslotte zijn metingen opgenomen uitgevoerd bij hoog vacuüm ($5 \cdot 10^{-7}$ mbar) en kamertemperatuur. Met name built-in splitsing verdwijnt niet. Dit effect kan dus niet volledig aan de omgeving worden toegeschreven.

Chapter 5 enters the path of AFM-based research of trapped charges. In section 5.1 principles of AFM are described, gradually but quickly concentrating on Force Modulated non-contact conducting AFM. How tip-surface interactions are translated in an observable quantity (resonance frequency shift of the AFM cantilever) is given. FM conducting AFM has been used for experimental work performed by Marko Sturm (reference chapter 5) on imaging local oxide charges (section 5.2).
Hoofdstuk 5 beschrijft onderzoek aan (locale) parasitaire lading gebaseerd op AFM. In sectie 5.1 worden de principes van AFM beschreven. Dit wordt geleidelijk maar vlot afgebakend richting Force modulated non-contact geleidende AFM. Er wordt aangegeven hoe tip-oppervlak interacties worden vertaald in een observabele grootheid (verschuiving van de resonantiefrequentie van de AFM cantilever). FM geleidende AFM is aangewend voor experimenteel werk uitgevoerd door Marko Sturm (zie Hoofdstuk 5 voor referenties) aan het afbeelden van locale oxideladingen.

De rest van dit hoofdstuk bevat berekeningen en simulaties betreffende de electrostatische interactie tussen de tip van een geleidende AFM en een metaalelectrode met een dielectrische laag en een locale ingevangen lading. Het electrostatische probleem wordt geïntroduceerd in sectie 5.3, alsmede hoe het is benaderd door een model uit de literatuur (Ludeke-Cartier en Lambert-Saint-Jean).
Sectie 5.4 laat enkele resultaten en voorspellingen zien van het hier voorgestelde multi mirror model (voor de opbouw van dit model zij de lezer verwezen naar de Appendices). Hieronder vallen de voorwaarden voor electrostatische *afstoting* van de tip door het sample (waar de kracht normaal gesproken aantrekkend is) en kritisch gedrag op korte afstanden. Speciale aandacht wordt besteed aan de oplosbaarheid van de grootte en de diepte van de ingevangen lading: vanwege de nabijheid van een metallisch (derhalve equipotentiaal) oppervlak, gevormd door de substraatelectrode, wordt een spiegellading gevormd. In feite heeft de interactie tussen tip en lading een dipool-karakter. Vanwege haar vorm $\mathbf{p} = q\mathbf{s}$, kunnen (oneindig) veel combinaties van lading q en afstand s dezelde dipool p opbouwen. Deze kunnen uit AFM-metingen bij verschillende biasspanningen naar verwachting niet opgelost worden. Metingen bij variërende tiphoogte kan in enkele gevallen uitkomst bieden. Een horizontale scan over de lading geeft geen verbetering als de tip veel groter is dan de relevante afmetingen (dikte van het dielectricum, positie van de lading ten opzichte van de substraatelectrode). Echter, zeer scherpe tips (<10 nm) op korte afstand (< 5 nm) moeten grootte en diepte van de lading kunnen oplossen.
Sectie 5.5 beschouwt finite element modeling en bevestigt de toepasbaarheid van het benaderen van de tip door een geleidende bol.
De drie benaderingen (Lambert & Saint-Jean, FEM en multi mirror) worden vergeleken in sectie 5.6 voor een dun (2.5 nm) en een modaal (25 nm) diëlectricum met $\varepsilon = 9.1$ (aluminium oxide) en een dielectricum met $\varepsilon = 3.9$ (silicon dioxide). De modellen worden vergeleken voor de grootheden kracht, kracht gradient, piek hoogte en V_{min}, waarmee de complete relevante electrostatische fenomenologie is afgedekt. Voor kracht, kracht gradient en piekhoogte voorspelt multi mirror drastisch hogere waarden dan Lambert & Saint-Jean (soms meer dan een orde van grootte). Multi mirror wordt hierin ondersteund door finite element. Een ander belangrijk aspect van multi mirror is dat het voorspelt dat piekhoogte en V_{min}, grootheden die de interactie met de ingevangen lading *sec* karakteriseren, naar nul afzwakken wanneer de lading wordt afgezonken naar het interface tussen het metaal en het diëlectricum, alwaar zij samenvalt met haar spiegellading en dus detectie zou moeten ontduiken. Echter, volgens Lambert & Saint-Jean wordt de interactie maar weinig zwakker.
Tenslotte wordt in sectie 5.7 gedemonstreerd dat multi mirror uitstekend presteert bij het voorspellen van de verschuiving van resonantiefrequentie van een AFM cantilever. Het literatuurmodel kan geen rekenschap geven hiervan. Tenslotte wordt beredeneerd dat het multi mirror model consistent is met profielmetingen van ladingen, als wordt aangenomen dat ingevangen ladingen voorkomen in clusters van meer dan één eenheidslading (~50).

Hoofdstuk 6 beschrijft wanneer ladingsvangst de werking van realistische MEMS schaadt. Het grotere deel van dit hoofdstuk is gewijd aan een MEMS gravitatiegradiometer, welke in zijn ontwerpfase verkeert. Het blijkt dat de beoogde gevoeligheid van 0.1 Eötvös/√Hz niet gehaald kan worden als aan het fenomeen van ladingsvangst geen aandacht wordt geschonken. Dit wordt berekend voor het statische geval, terwijl ook een begin wordt gemaakt met een dynamische benadering. De resterende deel behandelt het statische geval voor de RF-vermogen sensor, welke in

sectie 3.3 onderwerp is van experimenteel onderzoek. Dezelfde conclusie als voor de gravitatie gradiometer geldt hier ook. Deze sensor heeft natuurlijke, dunne diëlectrische lagen, waarin niettemin zodanig ladingsvangst kan optreden dat de streefgevoeligheod van 1 nW/√Hz (ver) buiten bereik blijft.

Appendix I verschaft uiteindelijk de fundamentele bouwstenen van het multimirror model en een theoretische rechtvaardiging ervoor. De eindigheid van het model wordt aangetoond.
In Appendix II wordt de kwadratische vorm van de electrostatische kracht afgeleid. Extra resultaten worden afgeleid, die de soms tijdrovende berekeningen van de normale toepassing van het multi mirror model kunnen helpen versnellen, door geschikte benaderingen na te streven. Deze resultaten zijn onder andere:
A) Een uiteenzetting betreffende de capaciteit tussen een bolvorminge geleidende tip op een bepaalde potentiaal ten opzichte van een geleidend oneindig vlak. Deze capaciteit kan worden uitgerekend via een oneindige serie spiegelladingen. Hoewel een uitdrukking voor de totale capaciteit, gebaseerd op een oneindige som, bekend is uit de literatuur, worden hier drie verschillende uitdrukkingen afgeleid: 1) een gesloten uitdrukking voor de n^{de} spiegellading kan worden opgeschreven. 2) De som over deze spiegelladingen kan worden omgeschreven in een vorm die sneller convergeert. 3) Deze ladingen kunnen worden geschreven in polynomen van eindige graad in termen van de ratio van de tipstraal R en de tip positie z_{t}. De coefficiënten van deze polynomen worden gegeven door bepaalde diagonalen in de driehoek van Pascal.
B) Geschikte benaderingen zijn gevoinden voor de qV- en qq- interacties zonder diëlectrische laag. De diëlectrische laag kan vaak bij benadering in rekening worden gebracht door een coördinatentransformatie.

Het multi mirror model is uitgewerkt voor een metallisch substraat. Het kan gemakkelijk worden uitgebreid voor het geval van een dielectrisch substraat, i.e. met een eindige relatieve diëlectrische constante. Tenslotte wordt de Appendix besloten met een korte discussie over de voors en tegens van enkele methoden om het multi mirror model door te rekenen.

DANKWOORD

Het vervult mij met voldoening dat u deze woorden leest. De verheugende aanleiding is de totstandkoming van het proefschrift dat u nu voor u heeft, waarmee een jarenlange periode van onderzoek wordt afgesloten. Velen hebben direct of indirect bijgedragen hieraan. Met dit dankwoord wil ik in de eerste plaats graag een aantal hiervan noemen. Maar tegelijkertijd noodt het schrijven hiervan tot terugblikken op deze periode. De lezer vergeve mij dat ik mij niet uitsluitend beperk tot een obligate opsomming van namen.

Mijn dank is in de eerste plaats gericht aan mijn promotor Horst Rogalla, die mij de kans gaf dit promotie-onderzoek te doen en er vervolgens ter verslaglegging enige zinnen aan te wijden. Op beslissende momenten stond hij mij bij om de eindstreep te halen. Zonder hem had ik bovendien een mooie tijd in zijn vakgroep Lage Temperaturen moeten missen.
Direct in zijn kielzog gaat mijn dank uit naar Jaap Flokstra, aan wie de dagelijkse begeleiding was toevertrouwd. Hij slaagde erin om op regelmatige basis zijn promovendi op gesprek te ontvangen in bovendien soms herculische sessies. Nauwgezetheid is zijn handelsmerk. Met het binnen zijn “sensors”-groepje notoire werkwoord “*japen*” pleegt men het onderwerpen van een tekst aan zijn niets ontziende adelaarsogen en rode pen aan te duiden. Het heeft dit proefschrift geen kwaad gedaan.

De Stichting FOM ben ik erkentelijkheid verschuldigd voor de “uitbesteding” van het onderzoek aan de UT en voor mijn plaasting aldaar, en voor alle overige steun die ik van haar heb gekregen om mijn werk in Enschede te doen en mijn promotie tot een goed einde te brengen.
Erkentelijkheid is ook op zijn plaats jegens Philips, die het project in eerste instantie financieel mede mogelijk heeft gemaakt. Theo Rijks en Joost van Beek dank ik ook voor hun wetenschappelijke input.

Voor de inhoud van dit proefschrift heb ik veel gehad aan de samenwerking met de leerstoel Transducers, Science and Technology (TST) van de faculteit Electrotechniek, Wiskunde en Informatica (EWI). Dit betreft in het bijzonder Remco Wiegerink. De ideeën die ten grondslag liggen aan de thermisch stabiele MEMS beschreven in Hoofdstuk 4 van dit proefschrift, waaronder de gebulkmicromachineerde dubbele veren, werden voor een belangrijk deel door hem voorgesteld tijdens de tête-à-têtes op Jaaps kamer. Zijn rustige, weloverwogen en kristalheldere uitleg en zijn snelle denken heb ik als stimulerend en bijzonder aangenaam ervaren. Een andere gewaardeerde suggestie van hem is Karmeliet.

Samenwerking was er verder met de leerstoel Vaste Stof Fysica. Aanvankelijk Marko Sturm en later vooral Herbert Wormeester hadden het wellicht twijfelachtige genoegen samen met mij bijna een jaar lang met de conducting AFM van het clustersysteem te stoeien. Olie dweilen en gevallen samples uit de vacuumkamers hengelen is niemands hobby. Des te meer bedank ik hen voor de geïnvesteerde tijd. Verder heeft Herbert kritisch meegedacht over het multi mirror model.

Ik denk met warme gevoelens terug aan Lage Temperaturen, de leerstoel waar ik vijf jaar lang deel van heb mogen uitmaken.

De secretaresses Ans en Inke zijn in velerlei opzicht de spin in het web van deze omvangrijke groep. Ik zoek nog een elegantere beeldspraak dan “smeerolie”. Aan hen is het bovendien te danken dat in dit mannenbolwerk planten en kerstversiering het werkklimaat veraangenamen.
Anders dan Frank Roesthuis’ achternaam zou kunnen doen vermoeden, prefereert hij in het lab elegante en kwalitatief hoogwaardige oplossingen boven “quick & dirty” gefröbel. Ook in de koffiekamer mengt hij zich graag onder ons mekkerende promovendi en houdt zijn humor en humeur (vooral vrolijk) niet verborgen. Eveneens aan Dick Veldhuis mijn dank, die ondanks een geleidelijke verschuiving van feitelijke werkplek van lab naar bureau mij niettemin van technische assistentie heeft voorzien.
“LT” heeft zijn naam feitelijk te danken aan Harry “Helium” Steffens, de onverwoestbare “kou-boy”. Gewapend met een heliumpistool in zijn holster is hij de schrik van ieder gaslekje. Ook Jan Talman zij genoemd.

Ik heb mij altijd meer dan thuis gevoeld in Jaaps sensors-groepje, hierboven al genoemd. Mede hierdoor bood de zondagavond soms zelfs nog plezierigere perspectieven dan de vrijdagavond ☺.
Aanvankelijk was daar Martin Podt. Met fotolithografie, de Nordiko en 4-op-een-rij heb ik dankzij mijn naamgenoot kennis gemaakt. Zijn kalme persoonlijkheid doet geen voorliefde vermoeden voor ontembare afstandsbestuurbare auto’s.
Concerning Javier (“Ja4”) Sesé, I envy his completely de-stressed composure, never in a hurry, and nevertheless (consequently?) getting a whole lot of work done in a short time. Also his clear thinking and explanation are exemplary. He introducing me to MEMS and the flow cryostat served a solid starting point for my PhD work. Formally I should also say “Gracias” for showing me the “máquina de microcontactos”. It is a reliable tool for training anger management.
Luis Fernández was “shared” between LT and TST. There are of course the fruitful scientific discussions and the (too) sensitive RF MEMS devices that he generously supplied me with. There is the teamwork in the MESA+ cleanroom; Luis, me and the robot mask aligner were an unbeatable trio. Pero mis recuerdos los muchos bonitos son de Tarazona. Era inolvidable.
Johan Reitsema, je hebt na je afstuderen een meer spirituele kant gekozen, veel geluk op dit pad!
Voor vier jaar hebben Johannes Pleikies en ik aangrenzende bureaus gehad en men kan zich niet beter wensen. Met bewondering heb ik zijn gietijzeren onverstoorbaarheid gadegeslagen. Onder de meest luidruchtige omstandigheden (verstrooiende praat, LA woman, Jantje Smit) bleef hij onbewogen de parameters van zijn SQUID-simulaties afstellen. Voor wèl relevante discussies, over wetenschap en politiek, was hij volop beschikbaar. Als Kollege wirst Du mir fehlen. Sehr vielen Erfolg mit dem letzten Stadium deines Promotions!
Reinder Cuperus, de onverzettelijke Fries. Op basis van zijn IT-kwaliteiten was het opzetten van de “sensor-server” hem wel toevertrouwd. Ik denk dat we aardig aan elkaar gewaagd waren wat betreft het maken van (taal)grappen van een soms deplorabele flauwheid. Reinder, aan jou de schone taak de eer van MEMS hoog te houden binnen LT! Veel succes met de gravitatie gradiometer!
Kris is er het voorbeeld van dat binnen canonieke werktijden (08.00u. 1e depositie, 16.59u. uitklokken) ook een promotie-onderzoek gedaan kan worden. Vanwege zijn lengte (en trouwe presentie) was hij altijd hèt herkenningspunt wanneer LT zich

verzamelde op de Oude Markt om de kroegen te bestormen. Succes met de magnetometers en het dresseren van vortices!
Kees (“Qäß”) Verwijs was voor mij onmisbaar tijdens de dagelijkse lunch in de mensa. Onbevangen en gretig stortte hij zich in welke discussie dan ook, zonder gêne voor een ongebruikelijke mening of een niet-sociaal afgestemde smaak. Bulkend van de posterprijzen smeedt hij nu zijn proefschrift tot een grafisch meesterwerk. Succes!!
Zoals al eerder aangegeven kon met LT-collega’s ook buiten het werk om gecongregeerd worden, een niet te onderschatten factor in een promotietijd. Buiten de reeds genoemde personen waren de kleurrijke Aico (de koffietafel heeft je nodig!), Pieter, Maarten, Jeroen “Nymus”, Mark, Aleksandar, Alexander B, Joska, Hen3, Harald, Vedran, Menno, Tim (al IM??), ... dan in enigerlei samenstelling hierbij present.

I am very happy that I have been given the chance to guide and work with many students from abroad. I appreciate warmly their contribution to my PhD work and especially the good company with them.
A Cédric Pronchéry, mon premier étudiant de Grenoble, merci de m'avoir aidé pour les expériences avec le "flow cryostat".
Antoine Becherel et Xavier Fix, vous avez très vite compris comment allier la vie estudiantine sur le campus et une présence matinale à labo, tout en y étant un binôme très efficace. Xavier, il est maintenant inutile de t’expliquer que l’alcool est dangereux au guidon et que pédaler après avoir (a)bu(sé) de la bière est dangereux pour le visage.
Javier Galán, muchas gracias para tu trabajo valioso en el laboratorio y tu presencia amable y impasible. Yo quiero repetir mía gratitud para la hospitalidad de Elena Crespo y Javier durante un week-end muy agradable en Nijmegen con Johannes.
Grégory Sale et Claire Patel, votre passage dans le groupe reste inoubliable. Claire, merci beaucoup aussi pour tes présentations de PowerPoint très créatives, au nom de tout le groupe LT!!!!!!!
In dit verband wil ik nog Sybolt Harkema en Paul te Riele bedanken voor de röntgendiffractometrie die jullie samen met haar hebben willen doen.
I wish every one of you good luck and good health!

From outside the LT division I would like to mention Aneta & Lukasz, Isabel, Ana and Andreea. I enjoy a lot to have met you and to have talked with you on several occasions!

Ik wil graag mijn paranimfen Johannes en Bart bedanken.
Bart, alle steun in serieuze en vrolijke momenten zijn voor mij van grote waarde geweest. Als je vastzit in het schrijven van jouw proefschrift, denk eraan: in Amersfoort heb je die lange wandeling ook tot aan de streep gehaald. Ik zal mijn backhand weer eens oefenen voor ons volgende spelletje poolbiljart! Heel veel geluk met alle grote veranderingen.

Speciaal wil ik mijn waardering uitspreken voor mijn ouders en naaste familie. Met hun onvoorwaardelijke, niet-aflatende steun en tomeloze belangstelling kun je je niet rijker wensen.
Séverine, nous n’avons que fait connaissance avec nous durant les derniers phases de ma promotion, mais je veux te remercier très fort pour soulager et ragaillardir cette temps et pour ton soutien patient et très affectueux. C’est de grand valeur pour moi.

Ik wens iedereen geluk toe.

Martin

STELLINGEN

bij het proefschrift

Capacitive MEMS-based sensors:
Thermo-mechanical stability and charge trapping

1) Thermo-mechanische stabiliteit is een noodzakelijke voorwaarde voor het bestuderen van de effecten van parasitaire ladingen binnen een breed, cryogeen temperatuurbereik.

2) Ingevangen ladingen in natuurlijk aluminiumoxide vertonen een differentiatie in beweeglijkheid. De beweeglijkheid neemt altijd af bij lage temperaturen.

3) In capacitieve MEMS sensoren kunnen zelfs ultradunne diëlectrische lagen zodanig parasitaire lading herbergen dat de werking van de sensoren meetbaar en intolerabel verstoord.

4) Wanneer in artikelen de tip-samplecapaciteit wordt berekend, wordt het spiegelen van de lading vaak slechts éénmaal uitgevoerd. Toch dient *herhaaldelijk* te worden worden gespiegeld, teneinde een aanzienlijke onderschatting van de capaciteit te vermijden. Dit geldt in toenemende mate voor de kracht en de krachtgradiënt.

5) Een locale, parasitaire lading in een ultradun diëlectricum op een metallisch substraat vormt in feite een dipool met zijn spiegelbeeld. Om deze reden kunnen grote AFM-tips grootte en positie van de lading vrijwel niet oplossen.

6a) Virtuele ladingen, de basis van Hoofdstuk 5, bestaan niet, maar er wordt mee gerekend alsof ze wel bestaan. Virtuele ladingen zijn dus leugens.
6b) Hoe herhaaldelijker wordt gelogen, des te dichter de waarheid wordt benaderd.

7) De meest constante factor in een promotie-onderzoek is onwetendheid. De meest gevaarlijke factor is onwetendheid dienaangaande.

8) Het stellen van betrekkingen helpt de wetenschap vooruit; dit in tegenstelling tot het stellig betrekken van stellingen, aangezien stellingen slechts betrekkelijk zijn.

9) Een tegenstelling tussen schaken en wetenschap is de wenselijkheid van een "bord voor de kop".

11) De noodzakelijkheid van het onderwerpen van stellingen aan de censuur van de promotor is enerzijds beperkend voor de artistieke expressie, anderzijds een stimulans voor het creatieve proces.

Martin van Essen
Enschede 14 januari 2009