D I S S E R T A T I O N

Optimization for Enhanced Thermal Technology CAD Purposes

ausgeführt zum Zwecke der Erlangung des akademischen Grades
eines Doktors der technischen Wissenschaften

eingereicht an der Technischen Universität Wien
Fakultät für Elektrotechnik und Informationstechnik

von

Stefan Holzer

[redacted]

Matrikelnummer 9525914
geboren am 18. Januar 1976 in Wien

Wien, im Juni 2007

Abstract

In the present phase of development, microelectronics has reached a state in which, because of the diminutive size of structures, quantum mechanical effects on the one hand side, and thermal and mechanical effects on the other hand side, gain increasing significance. As a result of the ever-shrinking size of the semiconductor devices, more and more components per given chip area can be integrated for the purpose of handling highly complex tasks more efficiently. This high integration density of semiconductor components entails new challenges for their design, operation, and reliability.

While — on the average — over a period of eighteen months the density of integration of components per chip doubles, over the same span of time the resulting power-loss density keeps growing exponentially — and, concurrently, resulting a dramatic increase in the thermal load on the components. Consequently, it becomes essential to provide considerably better cooling for these high-performance semiconductor devices than for traditional components of the same type.

In order to deal with these effects, a deepened insight into thermal effects and developments is required, while in addition, these effects and their consequences should be rigorously considered by an effective simulation software tool already during the development phase. This way, by the use of suitable optimizing strategies, the components involved can be optimally designed to meet specific operating conditions. As a result, effectively combining the simulation of electrical and thermal effects will be a task of ever increasing importance. Furthermore, for components of such diminutive size, it will become essential to also include mechanical aspects into relevant investigations and research, since particularly these effects are of considerable relevance for reliability.

Since both thermal and mechanical loads have significant bearing on the electrical properties as well as on the reliability of semiconductor components, it becomes a must to optimize them for their intended specific use. An absolute requirement for this is determining material properties with utmost precision, so that characteristics of critical components can be optimized for their specific purpose.

Furthermore, this dissertation is to demonstrate how — by means of simple transient electrical measurements and sophisticated optimizing strategies — important electrical and thermal material parameters can be identified without taking recourse to costly and time-consuming caloric measuring procedures. Subsequently, material parameters thus identified find use in complex compound structures in order to provide the even more precise results that are required for describing the transient temperature developments, and thus being able to calculate the resulting mechanical loads.

Kurzfassung

Die Mikroelektronik hat mittlerweile einen Stand erreicht, in dem aufgrund der Strukturgröße quanatenmechanische Effekte auf der einen Seite und thermische und mechanische Effekte auf der anderen Seite, immer mehr an Bedeutung gewinnen. Aufgrund der immer kleiner werdenden Halbleiterbauelemente können immer mehr Bauelemente pro Flächeneinheit auf dem Halbleiterchip intergriert werden, um hoch komplexe Aufgaben noch effizienter bewältigen zu können. Die hohe Intergrationsdichte der Halbleiterbauelemente birgt aber auch neue Herausforderungen in sich, für das Design, den Betrieb und die Zuverlässigkeit dieser Bauelemente.

Während sich die Integrationsdichte der Bauelemente pro Chip in achtzehn Monate im Durchschnitt verdoppelt, wächst die auftretende Leistungsdichte exponentiell im selben Zeitintervall und damit auch die thermische Belastung der Bauteile. Darum ist es notwendig, diese hochperformanten Halbleiterbauelemente noch stärker als bisher zu kühlen.

Um diesen Effekten vorbeugen zu können ist es notwendig, die thermischen Effekte einerseits besser zu verstehen und andererseits diese Effekte und deren Auswirkungen durch Simulation schon in der Entwicklungsphase aufzuzeigen. Dadurch ist es möglich mittels Optimierungsverfahren die Bauteile für bestimmte Betriebsbedingungen optimal zu entwickeln. Eine Kopplung der Simulation von elektrischen und thermischen Effekten gewinnt daher zunehmend an Bedeutung. Für sehr kleine Strukturgrößen der Bauelemente ist es von außerordentlicher Wichtigkeit, mechanische Aspekte in den Untersuchungen mitzuberücksichtigen, da diese Effekte erheblichen Einfluß auf die Zuverlässigkeit haben.

Thermische und mechanische Belastungen haben erheblichen Einfluß sowohl auf die elektrischen Eigenschaften als auch auf die Zuverlässigkeit der Halbleiterbauelemente. Daher ist es unumgänglich, die Halbleiterbauelemente für ihren Einsatz zu optimieren. Dafür müssen die Materialeigenschaften auf das genaueste bestimmt werden, um eine optimumale Bauteilcharakteristik zu erreichen.

Diese Arbeit zeigt wie anhand einfacher transienten elektrischen Messungen und ausgeklügelten Optimierungemethoden wichtige elektrische und thermische Materialparameter identifiziert werden können ohne teure und zeitintensive chalorische Meßmethoden anwenden zu müssen. Im Anschluß daran werden die identifizierten Materialparameter in komplizierteren Verbindungsstrukturen Anwendung finden um genauere Resultate zu liefern. Diese Daten sind erforderlich um die transiente Temperaturentwicklung zu beschreiben und um damit die mechanischen Belastungen berechnen und im Vorfeld abschätzen zu können.

Acknowledgments

I want to express my gratitude to my supervisors Tibor Grasser, Erasmus Langer, and Siegfried Selberherr, who gave me the possibility to enroll my PhD program at the Institute for Microelectronics. Moreover, I have to say big thanks to Siegfried for teaching me how to teach and supervise my students and some of my colleagues through several national and international projects. Special thanks go Hermann Schichl, who was always available at short call providing me with the appropriate help and valuable discussions regarding optimization issues.

I am indebted to my collegues Christian Hollauer, Alireza Sheikholesami, and Gregor Meller and their excellent knowledge of mechanics and thermodynamics which brought me a sympathy to the mechanics and thermodynamics side of microelectronics.

Furthermore, I want to thank Stephan Wagner and Wilfried Wessner for their valuable contributions and discussions on compilers, solvers, and meshes over a relaxing beer after work.

Special thanks go to Markus Karner, Andreas Gehring, and Oskar Baumgartner, the very helpful and understanding room mates. They always had good ideas for lunch and showed much imagination in interpreting some of my written works.

To my colleagues Robert Entner, Stephan-Enzo Ungersböck, and Martin Wagner, I want to say thanks for their valuable contributions and discussions, even concerning private matters. In addition to our research work, we had great times doing sports together. Moreover, I was allowed to learn a lot from Paul Jürgen Wagner. He excels in discussions on topics related to electrodynamics, mathematics, and LaTeX. I am also indebted to Rainer Minixhofer and Helmut Puchner for their valuable discussions on challenging and emerging technology and device issues from an industrial point of view.

Je remercie tous ceux qui m'ont aidé avec les corrections linguistiques de ce thèse et par leur soutien personnel. Mais particulièrement, je voudrais exprimer mes meilleurs sentiments de reconnaissance envers Ursula & Thomas Heumesser, Eleonora Luisa, Penelope & Karl-Erwin Lichtenecker, et biensûr ma professeur d'anglais et français Sissi Schiemel.

Zu guter Letzt möchte ich meiner Familie und meinen Freunden dafür danken, daß sie mich in diesem schwierigen Lebensabschnitt tatkräftig und unermüdlich unterstützt und mir stets so viel Verständnis entgegen gebracht haben.

En poussant le cri triomphal je remercie tous: "TÖÖÖÖRRRRÖÖÖÖ" .

Contents

List of Physical Constants

Symbol	Description	Relative Uncertainty
c_0	Vacuum speed of light $c_0 = 299\,792\,458\,\mathrm{m\,s^{-1}}$	per definitionem [1]
μ_0	Vacuum permeability constant $\mu_0 = 4\pi \times 10^{-7}\,\mathrm{H\,m^{-1}} = 12.56637... \times 10^{-7}\,\mathrm{H\,m^{-1}}$	per definitionem [1]
ε_0	Vacuum permittivity constant $\varepsilon_0 = \frac{1}{\mu_0\, {c_0}^2} = 8.854187817... \times 10^{-12}\,\mathrm{F\,m^{-1}}$	per definitionem [1]
h	Planck's constant $h = 6.6260693 \times 10^{-34}\,\mathrm{J\,s}$	1.7×10^{-7} [1]
k_B	Boltzmann's constant $k_\mathrm{B} = 1.3806503 \times 10^{-23}\,\mathrm{J/K}$	1.8×10^{-6} [1]
$\mathrm{N_L}$	Loschmidt's number or Avogadro's constant $\mathrm{N_L} = 6.0221415 \times 10^{23}\,\mathrm{mol^{-1}}$	1.7×10^{-7} [1]
q	Elementary charge of an electron $q = 1.60217653 \times 10^{-19} C$	8.5×10^{-8} [1]
R	Universal gas constant $\mathrm{R} = \mathrm{N_L}\, k_\mathrm{B} = 8.314472\,\mathrm{J\,K^{-1}\,mol^{-1}}$	1.7×10^{-6} [1]
σ_SB	Stefan-Boltzmann's constant $\sigma_\mathrm{SB} = \frac{2\pi^5 {k_\mathrm{B}}^4}{15 h^3 c_0} = 5.6704 \times 10^{-8}\,\mathrm{W\,m^{-2}\,K^{-4}}$	7.0×10^{-6} [1]

Notations

x	...	Scalar-valued quantity
$\mathbf{x}$	...	Vector-valued quantity
$\tilde{x}$	...	Tensor-valued quantity
$<x>$	...	Mean value of the quantity x
$f_1 \circ f_2$	...	Nested Function: equivalent to $f_1(f_2)$
$(\cdot)^{\mathcal{T}}$	...	Transposition of a vector or matix
∂_t	...	Time derivative operator: $\frac{\partial}{\partial t}$
∇	...	Nabla operator
$\nabla(\cdot)$	...	Gradient of $(\cdot)$
$\nabla \cdot (\cdot)$	...	Divergence of $(\cdot)$
$\nabla \times (\cdot)$	...	Curl (or Rotation) of $(\cdot)$
$\Delta(\cdot)$	...	LAPLACEan operator: $\nabla \cdot \nabla(\cdot)$
$\wedge$	...	Logical conjunction
$\Longrightarrow$	...	Logical implication

List of Acronyms

ALD	...	Atomic layer deposition
CAD	...	Computer aided design
CNT	...	Carbo-Nanotube
CORBA	...	Common Object Request Broker Architecture
CPU	...	Central processing unit
CVD	...	Chemical vapour deposition
CoO	...	Cost of ownership
DD	...	Drift Diffusion
DNA	...	Deoxyribonucleic acid
ELK	...	Extreme low-κ
ELSA	...	Enhanced Level-Set Applications
EOT	...	Effective oxide thickness
FET	...	Field effect transistor
Gb	...	Gigabit
GB	...	Gigabyte
GHz	...	Gigahertz
GUI	...	Graphical user interface
ILD	...	Interlayer dielectric
IPD	...	Inputdeck language
IT	...	Information Technoloy
ITRS	...	International Technology Roadmap for Semiconductors
LPCVD	...	Low pressure chemical vapour deposition
MC	...	Monte Carlo
MTTF	...	Mean time to failure
MILP	...	Mixed integer linear programming
n-Si	...	n-doped Silicon
NFS	...	Network file system
NTC	...	Negative temperature coefficient
OSG	...	Organo-silcate glass
p-Si	...	p-doped Silicon
poly	...	polycrystalline
polySi	...	polycrystalline Silicon
PTC	...	Positive temperature coefficient
PTFE	...	Polytetrafluoroethylene
PZT	...	Lead Zirkonium Titanate $Pb(Ti, Zr)O_3$
RE	...	Rare earth elements, member of the lanthanoide group {La, Ce, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu}
RTD	...	Resonant tunneling diode

SeiLib	...	Simulation Environment Interaction Library
SEM	...	Scanning electron microscopy
Siesta	...	Simulation Environment for Semiconductor Technology Analysis
TCAD	...	Technology computer aided design
TEM	...	Transmission electron microscopy
TEOS	...	Tetra-ethoxy-silane, Tetra-ethyl-ortho-silicate: $Si(C_2H_5O)_4$
VLSI	...	Very large scale integration
WSS	...	Wafer-State-Server: File format for distributed quantities
XML	...	Extensible markup language

List of Chemical Symbols

Ac	...	Actinium	Gd	...	Gadolium	Po	...	Polonium
Ag	...	Silver	H	...	Hydrogenium	Pr	...	Praseodymium
Al	...	Aluminum	He	...	Helium	Pt	...	Platinum
Am	...	Americum	Hf	...	Hafnium	Pu	...	Plutonium
Ar	...	Argon	Hg	...	Mercury	Ra	...	Radium
As	...	Arsenic	Ho	...	Holmium	Rb	...	Rubidium
At	...	Astatine	In	...	Indium	RE	...	Rare earth elements
Au	...	Gold	Ir	...	Iridium	Re	...	Rhenium
B	...	Bor	J	...	Iodine	Rh	...	Rhodium
Ba	...	Barium	K	...	Pottassium	Rn	...	Radon
Be	...	Beryllium	Kr	...	Krypton	Ru	...	Ruthetium
Bi	...	Bismuth	La	...	Lanthanum	S	...	Sulphurous
Bk	...	Berkelium	Li	...	Lithium	Sb	...	Antimony
Br	...	Bromine	Lr	...	Lawrencium	Sc	...	Scandium
C	...	Carbon	Lu	...	Lutetium	Se	...	Selenium
Ca	...	Calcium	Md	...	Mendelevium	Si	...	Silicon
Cd	...	Cadmium	Mo	...	Molybdenum	Sm	...	Samarium
Ce	...	Cerium	Mg	...	Magnesium	Sn	...	Tin
Cf	...	Californium	Mn	...	Manganesium	Sr	...	Stronitium
Cl	...	Chlorine	Mo	...	Molybdenum	Ta	...	Tantalum
Cm	...	Curium	N	...	Nitrogenium	Tb	...	Terbium
Co	...	Cobalt	Na	...	Sodium	Tc	...	Technetium
Cs	...	Cesium	Nb	...	Niobium	Te	...	Tellurium
Cu	...	Copper	Ne	...	Neon	Th	...	Thorium
Dy	...	Dysprosium	Nd	...	Neodymium	Ti	...	Titanium
Er	...	Erbium	Ni	...	Nickel	Tl	...	Thallium
Es	...	Einsteinium	No	...	Nobelium	Tm	...	Thulium
Eu	...	Europium	Np	...	Neptunium	W	...	Tungsten
F	...	Fluorine	O	...	Oxygen	U	...	Uranium
Fe	...	Iron	Os	...	Osmium	Xe	...	Xenium
Fm	...	Fermium	P	...	Phosphorus	Y	...	Yttrium
Fr	...	Francium	Pa	...	Protactinium	Yb	...	Ytterbium
Ga	...	Gallium	Pb	...	Lead	Zn	...	Zinc
Ge	...	Germanium	Pd	...	Palladium	Zr	...	Zirconium

Chapter 1

Introduction

In the early 1970s, at the beginning of the age of microelectronics, the integration density was approximately two transistors per square millimeter [2]. This density is very low compared to integration densities of today's microelectronic devices [2, 3]. With the continuous miniaturization of semiconductor devices and a nearly constant power dissipation per transistor, the power densities on microelectronic chips have increased in much the same way as the devices have shrunk. Therefore, thermal issues have become more and more important for the design of state-of-the-art devices determined by the maximum thermal budget for the device fabrication process and for the device operation.

The thermal budget can be influenced by the choice of the materials and chemical reactions that are used to form the microelectronic device structure. The increase of temperature causes changes in chemical reactions and growth rates as well as an increase of the electrical resistance and also broadening of doping profiles. Obviously, finding the appropriate temperature is crucial for all parts in device fabrication and device operation. Therefore it is important to investigate and predict the influence of the temperature to determine the process window for fabrication and the operation window to run the device.

The main goal of today's and future microelectronic designs is to determine the produced heat and heat flow in the device as well as to estimate the temperature and its impact on the device characteristics and its consequences for surrounding materials. In addition to the thermal requirements for fabrication and operation, the occurring self-heating can be critical if the heat flow through the device from the heat source to the heat sink cannot be controlled appropriately. This phenomenon can often be observed in devices with very high power densities.

Hence, thermal effects are becoming the dominant factor which determine the maximum performance of integrated circuits due to limited heat transport to the heat sink. The temperature dependence has been neglected for a long time, but today for certain characteristic parameters of microelectronic devices and technologies, these formerly neglected of self-heating effects in the simplified models have to be adapted. Hence, these parameters are becoming functions of temperature and their maximum heat transport capabilities. For example the maximum clock frequency is mostly determined by the temperature increase induced by the increased switching currents and the enhancement of integration density, which produce a higher power dissipation. Hence, the surrounding devices are also heated by this additional power density. Due to the temperature increase, also the volume expansions of the materials have to be considered which result in changes of the crystal structure and therefore also in the band structure of the semiconductor material. This effect can increase or decrease the carrier mobility depending on the direction of

the mechanical stress. Thus, the complexity and the maximum integration density of integrated circuits is limited by electrical, thermal, and mechanical constraints of material properties.

To achieve the present and future goals proposed by the International Technology Roadmap for Semiconductors (ITRS) [3], faster devices have to be created which need less chip area and less power to operate. Therefore, imminent thermal effects [4, 5] have to be considered carefully in order to keep right on track. Since the physics behind these effects is very complex and is only rudimentary implemented, more rigorous physics-based models have to be applied which increase the computational effort tremendously. Rigorous investigations of problems which are of industrial interest are therefore limited by the finite resources both in time and CPU power.

However, despite of these limitations, the time-to-market has to be reduced in order to position new microelectronic products at an early stage for an advantageous market share. Therefore, experiments have to be carried out either with measurements or simulations to achieve the required quality criteria. The faster the developments can be achieved and verified, the faster the market entry and therefore, the higher the earnings are for the first development stage. Hence, rapid results can be obtained with a combination of simulations and experiments, where the simulations give an overview on the intrinsic material parameters and quantity distributions, while the measurements ensure that the overall functionality remains with in the specifications. Measurements of material parameters of real devices cost a lot of money and time because it takes the whole time of a complete fabrication process before the measurements can be performed. With simulation tools, the same experiments can be performed within a very small fraction of the time a new device fabrication process would take. Moreover, the simulation results can be calibrated in order to perfectionize the prediction from the simulation. Again, these new and better results can be used to develop better simulation models which have to be calibrated as well for each new technology. The main benefit from this procedure is that once the simulation has been calibrated to the technology, the prediction gained from the simulation results supports the engineers during the design and the fabrication process. Therefore, physics-based models have to be developed, that allow the calculation of sufficiently accurate simulation results within reasonable time for the evaluation. For such purposes, but especially for thermal problems, new approaches or enhancements of existing models are needed in order to detect hot spots due to heat accumulation, heat conduction paths, and heat fluxes as well as other phenomena that coincide with the increase of the temperature. In order to obtain accurate simulation results within reasonable time, improved numerical methods are required as well.

1.1 Challenges Discussed in this Thesis

This chapter presents the motivation of this thesis and points out the resulting benefits for semiconductor devices, their enhancements, and their applications.

In Chapter 2, the impact of thermal effects on important parameters of fabrication processes and operation conditions of microelectronic devices is discussed. Moreover, this chapter shows how coupling of different simulation regimes is possible and consequently provides a better insight into today's critical thermal issues for semiconductor devices.

Chapter 4 gives an introduction to available optimization techniques that can be used within a certain range of applications for industrial requirements. Furthermore, important industrial requirements for optimization in general are presented as well as an approach which allows to apply an inverse modeling technique for parameter extraction and identification purposes in an efficient way.

Examples are presented in Chapter 5 which show the wide range of applications for optimizations and for the optimization environment SIESTA (Simulation Environment for Semiconductor Technology Analysis). In order to improve device characteristics as well as to reduce thermally induced parasitic phenomena, some special sub-problems are discussed that are crucial for certain device fabrication processes.

To conclude this thesis, Chapter 6 discusses future trends in microelectronics with a focus on exploiting thermally induced phenomena and presents alternatives for the semiconductor devices and materials in the micro and nano regime.

1.2 The ITRS and MOORE's Law

Since the beginning of the microelectronics age, the increase in sophistication of semiconductor devices seems to grow exponentially. This behavior was described by MOORE[1] [6] in 1965 and remains valid until today [7]. The transistor integration densities as well as the complexity of microelectronic devices have increased in the same speed. On average, they have doubled every eighteen months [3, 8–11]. This observation has become known as MOORE's law.

It seems that despite some fluctuations of the average speed of the enhancements, the growth rate remains exponential and keeps constant for microprocessors and memory and seems to remain valid for future devices [7]. At present, the upcoming challenges are the absolute physical constraints of the atomic material structure [3] in terms of the electrical behavior and mechanical properties due to the high stress gradients in thin film materials which appear during fabrication and may be enhanced during device operation.

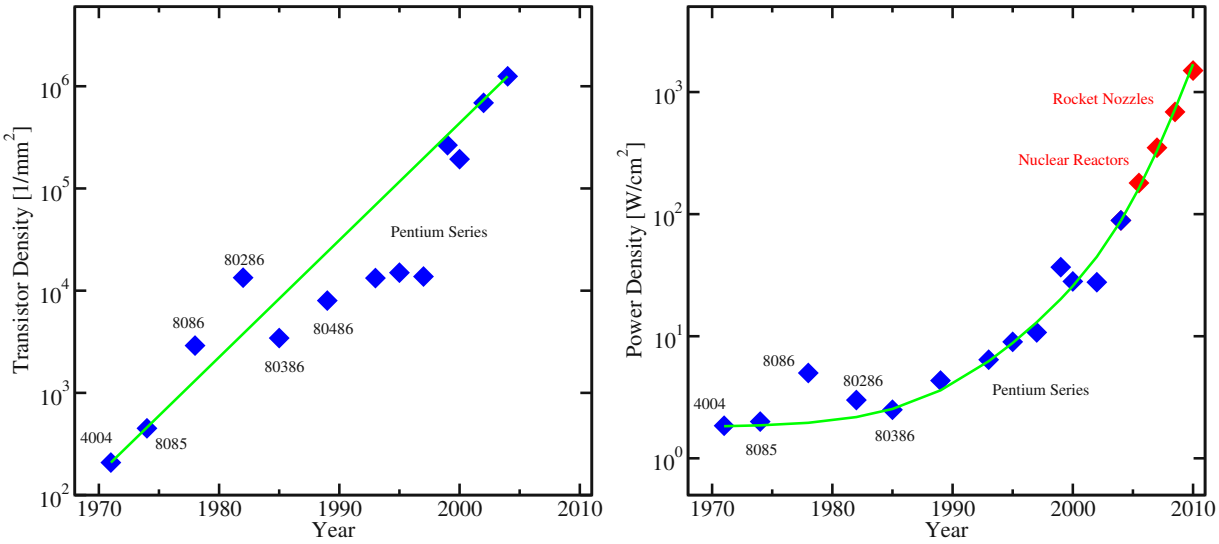
The enhancement of integration and the shrinking of the device feature size leads to smaller and more sophisticated devices. Thus, more devices can be integrated within the same area as before, leading to a higher package density. At a certain package density, however, new effects become important, e.g. self-heating of the semiconductor devices and the interconnect lines [11, 12]. To overcome these problems, the device engineers have to develop smart device arrangements and new technologies for fabrication to reduce parasitic (thermal) effects or to exploit them by using these effects for compensation purposes.

In 1959, the first integrated circuit was presented by KILBY [13, 14] which was followed by the first microprocessor on a single chip, the 4004 from INTEL in 1971 [2]. This processor series was improved and continued by the 8008 in 1974 and other processors. In 1978, the 8086 [2] was announced and newer and more powerful processors have followed.

The outlook of MOORE on the future of semiconductors has motivated and forced design and research centers to fulfill the predictions of MOORE. A viewgraph of the achieved enhancements for microprocessors is outlined in Figure 1.1. The current version of the International Technology Roadmap for Semiconductors (ITRS) summarizes the requirements which are mandatory for the future enhancements to obtain the predicted sophistication of semiconductor devices.

Figure 1.1 shows the achievements for microprocessors over the last decades since 1972 [2]. In Figure 1.1a a logarithmic plot shows the transistor density of selected microprocessors versus the year of their announcements. Obviously, the transistor density on die shows an exponential growth. It doubles every 18 months. Figure 1.1b depicts the corresponding power loss densities on these selected microprocessor chips. As it is clearly depicted in Figure 1.1b, the power loss density increases exponentially even on the logarithmic scale. Extrapolated by the current growth

[1] Gordon Earl Moore (born January 3, 1929)



(a) Transistor integration density per die

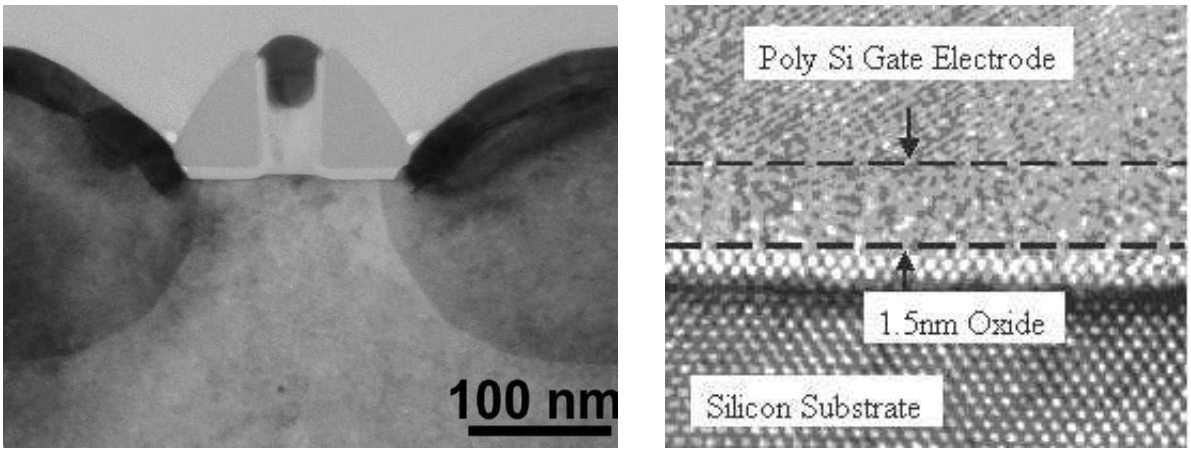
(a) Power loss density per die

Figure 1.1: The chronological evolution of selected microprocessors in terms of transistor integration density per die (a) and the observed power loss density per die (b).

rates and assuming that the past trends in development will continue, we would soon reach power loss densities of nuclear power plants or rocket nozzles in a few years.

With the fast development of the early semiconductor technology in terms of purity of the materials, the integration density, and the complexity of the device structures, new processor types have become possible, e.g. the 80286, 80386, 80486, and today Core2Duo and Core2Quad. In 1975, MOORE's forecast of the electrical operation conditions as well as the package density of the processors from the 1980s were assumed to be highly futuristic and not realizable in terms of former device fabrication processes [8]. However, this vision has come true and the processors were realized in spite of general disbelief and are being constantly improved.

Today, a similar development of the enhancement process can be observed. According to the future predictions of the current ITRS [3], device structures have been proposed which reach or have already gone beyond of the absolute physical constraints for classical devices. For instance, the thicknesses of gate oxides have been successfully reduced to approximately three to four atomic layers of SiO_2 [15, 16] (cf. Figure 1.2), which raises questions about the efficiency due to high leakage currents through these thin layers [17] and long-life reliability [18]. However, for switching transistors used in digital logic devices, the provided functionality is sufficient for reliable operation. This point of view is completely different from ten to twenty years ago, but changing the view point has enabled engineers to find new solutions within these new constraints. From our current understanding, shrinking this type of devices by one more step would result in oxide thicknesses of two atomic layers and the energy barriers in this structure would be far too small for a proper transistor operation. As an alternative to these limits, new materials with a higher dielectric constant are introduced which provide the same capacitance as SiO_2 but with a larger gate dielectric thickness. In order to compare the effective thickness of the gate dielectric, the physical thickness of the gate dielectrics is mapped to the thickness of a SiO_2 layer with the same electrical properties. This thickness is called the effective oxide thickness (EOT).



(a) Structure of a 35 nm MOS transistor (b) Gate dielectric layer

Figure 1.2: Transmission electron microscopy (TEM) picture of a typical 35 nm MOS transistor [15] (a) and a typical gate dielectric layer structures [16] (b) with a thickness of 1.5 nm.

This example shows the capabilities of how new technologies can exploit certain material properties to enhance the transistor performance and to reduce the area per transistor on a die. In addition, this simple example has shown that the future will bring new paradigms for building and operating electronic devices.

1.3 Electrical and Physical Properties

With the same speed as integration density of semiconductor devices increases, the requirements of material properties have to be adjusted accordingly which also implies improvements of fabrication technology in terms of purity of the material sources as well as the uniform applications on reactor-scale. Hence, the fluctuation of the electrical and thermal conductivity and of the permittivity has been dramatically reduced to obtain devices with uniform characteristics and with improved reliability. Many enhancements have been proposed by former editions of the ITRS [10] and most of them have been achieved in advance. Only a few forecasted technologies are missing. Most of the missing parts are related to too high power densities or too high current densities [3]. This lack of achievements leads to more research activities and the rapid development of new types of devices and has pushed the effort to find alternative materials for device interconnect structures [19, 20], dielectrics [21, 22], and for the semiconductor device parts of microelectronic devices [15, 23].

To overcome these types of problems, the current technique is to use other materials which have a better performance than SiO_2. The requirements for alternatives are to provide a higher dielectric constant, under the constraint that the life-time and the leakage current meet the specifications of the design.

By the introduction of materials with smaller lattice constants, the number of atomic layers slightly increases, but the possible enhancement is less than approximately 20% due to their similar lattice constants compared to that of SiO_2 [24, 25]. To achieve better results for downscaling, it is necessary to introduce materials which perform better at large thicknesses than

SiO_2 but behave similar to SiO_2 in terms of insulation and interface behavior. To choose the appropriate materials for a proper operation of a FET, the value of the control capacitance between the gate and the channel must not fall under a certain threshold value for the capacitance.

The control gate capacitance C_{GB} can be roughly estimated by using the formula of a capacitor for parallel plates

$$C_{\mathrm{GB}} = \varepsilon_0 \varepsilon_{\mathrm{r}} \frac{A}{d}, \tag{1.1}$$

where ε_0 is the dielectric constant for the matter-free space, ε_{r} the material-specific dielectric constant, A the effective area of the gate electrode, and d the average distance between the gate contact and the channel of the FET. For real devices, the capacitance of the control device is a function of the work function of the gate contact material, the channel doping, and the applied voltage as presented in Figure 1.3. However, to improve the prediction of the electrical behavior of state-of-the-art devices the range of validity has to be considered in advance. For this particular example of a MOS capacitance, a more accurate approach is required to obtain rigorous models for the depletion zones in the gate material for polySi and the voltage dependence of the MOS capacitance structure.

In order to increase the value of the gate capacitance, the area A and the dielectric constant ε_{r} have to be increased or the distance of the gate dielectric layer has to be decreased. Due to technology and cost-efficient reasons, the area has to be kept constant. The thickness of the gate dielectric layer cannot be reduced any more due to the hard limit of the lattice constants. The only remaining variable part in (1.1) is the material-specific relative dielectric constant ε_{r}. If there exist materials which behave like SiO_2 but have a higher ε_{r}, the thickness of the gate dielectric can be increased while the overall gate capacitance remains constant or increases.

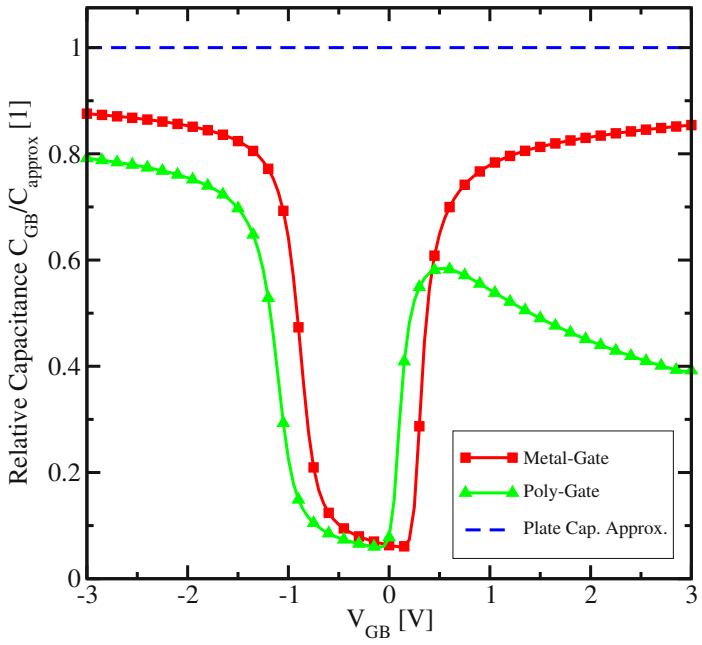


Figure 1.3: Capacitance-voltage characteristic of a typical MOS transistor with a gate dielectric thickness of 1.2 nm.

By using materials with lower dielectric constants than SiO_2 the capacitance from (1.1) can be reduced as well. This can be applied for dielectric layers in between of interconnect lines to reduce for instance the cross-talk of these lines due to the reduced dielectrics which are currently widely used in VLSI devices.

Another considerable performance improvement for VLSI devices was the replacement of Al by Cu for the reduction of the specific line resistivity for high speed applications. This material change has brought many improvements, e.g. less power consumption due to the reduced resistivity of the interconnect lines, but also a lot of new challenges for the device fabrication processes. Hence, due to the higher diffusivity of Cu, new barrier layers had to be introduced [26]. Moreover, for special applications, where the performance of Si-based materials is not sufficiently enough, new materials have been investigated in order to improve the overall performance and in addition also the reliability of these devices.

Some these special applications are memory cells, which require high dielectric constants in order to provide capacitances at very high package densities, so SiO_2 has been partially replaced by Ta compounds [27] or by Perovskite materials[2] [29] for capacitor materials.

All these measures require a considerable effort in terms of development, time, and money. Hence, the industrial manufactures try to exploit the available technology node as long as possible to save money. This sometimes leads to quite unconventional, but very effective results in device designs [30], to new improvements in process technology [31] and material science [32].

The use of alternative materials has posed new challenges, which have to be identified and considered as well. For instance, the substitution of Al by Cu has increased the conductivity and decreased bulk electromigration and therefore the long-life reliability, but due to technology reasons, parasitic side effects occur at the Cu interfaces [33]. First, Cu easily diffuses into Si and SiO_2 [26, 34, 35]. Therefore, a new barrier layer has to be introduced into the device structure. Moreover, the adhesion between Cu and the barrier layer is quite weak, which causes a high disorder of the crystal lattice structure at the side walls of the metal lines. Hence, high-speed diffusion paths have been established and the desired enhancement of the long-life reliability is not as high as estimated at the introduction of this technology [36].

1.4 Thermal Properties and Power Density

The decreasing feature sizes and the increasing integration densities go hand in hand with the increase of power dissipation per unit area. Since the output specifications of electrical circuit designs are kept within the same range of magnitude for backward compatibility, the local power density increases more than exponentially as the device sizes shrink (cf. Figure 1.1 and ITRS [11]).

In the early stages of the microelectronic age, the performance was mainly determined by the achievable electrical properties. Today's microelectronic devices are determined also by thermal issues. The heat produced on the chip has to be transported to the heat sink at the surface of the chip. Since modern chip designs consist of multiple metalization layers, there is the Si substrate on the bottom which is normally thermally connected to a heat sink but at the top side there are only thermally insulating layers which include also the ceramic or plastic package on top of the chip. Hence, the heat is transported through the metal layers because their thermal conductivity is much higher than the thermal conductivity of the insulation layer on the lower side. Due to the heat transport through the metal layers, the metal lines heat up. This leads to

[2] The members of the Perovskite group are oxides which follow the general chemical formula AXO_3, where A is substituted by {Ce, Ca, Na, Sr, Pb, RE} and the X site can be occupied by {Ti, Nb, Fe} [28].

a global warming of the chip which means in the worst case that a single hot spot at one side could determine the temperature of the whole chip, if the heat transport is unfavorably arranged or designed.

To overcome this particular type of design problem, stationary and transient thermal effects have to be included foremost into the device and process simulation models [37–39]. Since the major goal of TCAD applications in the industrial design flow is to obtain sufficiently accurate simulation results in reasonable time, simulation models have to cover the most dominant effects. Many of these effects have been described earlier in a fundamental way. If the physical models are still not accurate enough or are computationally too expensive, parameterized compact models have to be introduced and calibrated to obtain sufficiently accurate results in reasonable time.

The ideal thermal requirements for today's microelectronic devices can be briefly summarized as follows. First, the thermal conductivity of dielectrics or their heat capacitance should be adjustable by choosing different materials which are compatible with other involved metals in terms of fabrication and material interactions. The optimum for the designer engineers would be to choose the materials for interconnects and contacts according to their needs, for instance in terms of electrical or thermal conductivity or according to the material interaction properties. However, since the materials within a certain technology node cannot be arbitrarily chosen, the design engineers have to decide on the best materials according to several technology-dependent constraints, e.g. the thermal budget for the fabrication process, the electrical and thermal load capacity, and other limits which occur at device fabrication processes or during device operation.

Once the materials are chosen, the device can be finally designed. However, after process and device simulations, the results proof whether the designs of the device structures and its chosen materials yield the desired performance and characteristics. Without simulation, the procedure of producing a test wafer to measure the characteristics takes several weeks instead of few minutes or hours. At this stage of development, thermal effects like self-heating, heat conduction, or heat accumulation are commonly neglected or have mostly not been considered correctly using standard design tools. Hence, without electro-thermal simulation or early measures of test devices, unintentional problems might occur within the prototype phase.

1.5 Challenges

Due to the high power loss densities in the interconnect structures, the materials for these structures are required to have a high electrical conductivity to provide a good electrical connection between two or more contacts. The heat transfer of the connected contacts has to be reduced or enhanced according to design requirements of the particular device. This requested feature is only possible by changing the material or the material composition in the case of binary or ternary materials. Since the exchange of materials often requires several additional process steps, such design decisions have to be considered carefully due to additional cost. Hence, the achievable optimum device characteristics are often not realizable in terms of given economical constraints. Therefore, other solutions have to be found to implement them in the fabrication process. The first step for the implementation is the determination of the electrical and thermal behavior for non-optimal device structures. Together with given constraints of fabrication and the design requirements, the best solution can be obtained using adaptive optimization strategies provided from state-of-the-art optimization frameworks [40–44].

For changing electrical or economical requirements, a similar optimization procedure can be applied. Because some of those requirements demand alternative materials, various numbers

of new materials have been introduced and are gaining more and more importance for future applications. However, their application is very expensive and can be expressed as the "Cost of Ownership" (CoO), which includes the costs for fabrication and for additional precautions like the maximum allowed electrical or thermal burden.

Moreover, many new constraints appear if new materials are considered for optimization. For instance the use of Cu interconnect lines demands additional barrier layers to avoid the diffusion of Cu into the surrounding dielectrics [26]. Therefore, the application of alternative materials like Perovskites [28, 29, 45] or other high-κ materials [22, 27, 45] or low-κ materials [21, 46] is very limited with respect to the benefits per costs ratio.

Approaches

Materials in small device structures underly certain parameter fluctuations as much as bulk materials, but due to the small dimensions of the material regions, the impact of the fluctuations is much higher as compared to bulk material. Hence, fluctuations have to be considered from the beginning of the design. Even if an optimal device has been designed, the characteristic after its fabrication might be completely different. In order to minimize these discrepancies, certain technology-specific constraints have to be introduced which have to be considered within the optimization frameworks to improve the characteristics. To perform the optimization tasks the state-of-the-art simulation and optimization framework *Simulation Environment for Semiconductor Technology Analysis* (Siesta) [44] is used and refined in this thesis which provides an open interface that allows to easily add new software tools. Siesta can be used with several optimization strategies for specific optimization tasks. The optimizer varies and proposes values for the unknown or uncertain parameters. The framework sends the parameters from the optimizer in an appropriate format to the simulators. The simulator may be arranged in a simulation tool flow where the output of one simulator is submitted as the input to another simulation tool. At the end of the simulation flow, the quality of the final simulation result is determined by an objective function which returns a score value which is a quantified representation of the quality of the simulation result. The following presents typical applications in which optimization is used.

Parameter extraction can be used to identify model parameters which are not accurately known [47]. The required input data for this task includes the simulation software with the appropriate models, as well as measurements or reference data to which the simulation result can be compared, and a score function (or objective function) that determines the quality of the simulation result. This extraction mechanism uses the inverse modeling technique [48–50], which is often performed to characterize novel device structures and new materials as well as material compositions in order to develop compact models at a specified scope.

Calibration is a special case of the parameter extraction [49]. The range of the uncertain parameters can be further constrained which enables in general faster convergence to complete the calibration task. The main difference between parameter extraction and calibration is that calibration needs a much higher accuracy because the initial guess is normally very close to the optimum, but should be further improved, if for instance a sample has to be calibrated to a certain set of measurements to minimize the model error. Due to the higher quality demands, the determination of the quality of the simulation result is a very critical issue for calibration. These quality criteria (objective or score functions)

have to be specified by the user for each particular problem class and tuned for each individual problem. This function can include comparisons of absolute and relative values to calculate a significant metric to determine the quality of the simulation result with respect to reference data.

General optimization is the most general approach and can be used for arbitrary purposes. The optimization is performed until a certain quality criterion has been reached. There exists a wide range of applications for the optimization related to TCAD or electronic devices [51, 52]. More general electronic design purposes have been discussed in [53, 54], and specific optimizers and application for other regimes for instance in economics have been discussed in [55, 56].

With a rigorous implementation of the major aspects occurring in a particular setup problem the optimization framework is able to minimize or maximize certain figures of merit within user-defined specifications. Hence, many trade-offs can be optimized together to obtain a reasonable solutions for the specified problem.

Chapter 2

Thermal Effects in Semiconductor Devices

"Alles Vergängliche ist nur ein Gleichnis."

Johann Wolfgang von Goethe[1]

MANY CHALLENGES faced in modern semiconductor devices are related to heating phenomena. Since shrinking the device feature size causes higher power loss densities and therefore higher and faster temperature evolution inside the device structure, many additional problems occur due to material-related constraints because the produced heat cannot be transported to the heat sink fast enough. Hence the surrounding device structure heats up and the global microelectronic chip heats up globally.

Since the absolute temperature is not zero, matter is in steady motion at least in terms of BROWN's[2] molecular movements. Therefore, the most probable consequence is that the number of possible states of a closed system, e.g the quantum states, is increasing until a temporary state of thermal equilibrium has been reached. Hence, also the entropy which represents the information about the reachable states in a system is not decreasing spontaneously. This fundamental theorem of thermodynamics and its derivations challenge today's electronic devices including the decrease of the device feature size on the wafer while the ITRS request that the operational current density remains the same. Correspondingly, the power loss density increases quadratically with the reduction of the feature size. Two possible alternatives to overcome these problems are to reduce the supply voltage or the use of alternative materials which produce and inherit less parasitic effects. To describe the general behavior of the electro-magnetic system, fundamental electromagnetic field equations are given by MAXWELL[3] [58–60] as

$$\nabla \times \mathbf{E} = -\partial_t \mathbf{B}, \tag{2.1}$$
$$\nabla \times \mathbf{H} = \mathbf{J} + \partial_t \mathbf{D}, \tag{2.2}$$
$$\nabla \cdot \mathbf{D} = \rho, \tag{2.3}$$
$$\nabla \cdot \mathbf{B} = 0, \tag{2.4}$$

[1] Johann Wolfgang von Goethe (August 28, 1749 – March 22, 1832)
"All transitory things are only symbols or reflections." [57]
[2] Robert Brown (December 21, 1773 – June 10, 1858)
[3] James Clerk Maxwell (June 13, 1831 – November 5, 1879)

which are independent of the temperature and independent of the type of matter. The quantities **E** and **D** are the electric field and the displacement vector while **H** and **B** are the magnetic field and the magnetic flux density vector, respectively. **J** represents the current density and ρ the electric space charge density. Information about the material properties is included via the material equations for each region of matter by

$$\mathbf{D} = \tilde{\varepsilon} \cdot \mathbf{E}, \tag{2.5}$$

$$\mathbf{J} = \tilde{\sigma} \cdot \mathbf{E}, \tag{2.6}$$

$$\mathbf{B} = \tilde{\mu} \cdot \mathbf{H}. \tag{2.7}$$

With these equations, the field quantities and the flux quantities are coupled via the material-dependent parameters $\tilde{\varepsilon}$, $\tilde{\sigma}$, and $\tilde{\mu}$. Here, $\tilde{\varepsilon}$ denotes the permittivity tensor, $\tilde{\sigma}$ the conductivity tensor, and $\tilde{\mu}$ the permeability tensor. While the MAXWELL equations describe only field properties and field physics, the introduced material parameter tensors $\tilde{\varepsilon}$, $\tilde{\sigma}$, and $\tilde{\mu}$ reflect the material physics and describe the observed material-related phenomena within the involved materials. To complete the equations for the material parameters the heat conduction equation can be derived from FOURIER's[4] law [61]

$$\mathbf{q}_{\mathrm{th}} = -\tilde{\lambda} \cdot \nabla T, \tag{2.8}$$

in which $\mathbf{q}_{\mathrm{th}}$ represents the local heat flux density. This quantity is proportional to the temperature gradient ∇T, where the proportionality factor $\tilde{\lambda}$ is the thermal conductivity tensor. In conjunction with (2.8), the energy conservation law can be expressed by

$$\nabla \cdot \left(\tilde{\lambda} \cdot \nabla T\right) = \rho_{\mathrm{m}}\, c_{\mathrm{p}}\, \partial_t T - H_{\mathrm{th}}. \tag{2.9}$$

Here, ρ_{m} denotes the mass density, c_{p} the specific heat, H_{th} is the heat generation term, and T the local temperature.

The heat conduction equation (2.9) has some critical quantities: the heat generation H_{th} term, the thermal conductivity $\tilde{\lambda}$, and the specific heat capacitance c_{p}. The heat generation term H_{th} is mostly represented by JOULE's[5] power loss (cf. Section 2.2.3). For the thermal conductivity $\tilde{\lambda}$ and the specific heat capacitance c_{p} several tables of material parameters exist, which have also different ranges of validity (cf. Section 2.2.1).

In order to describe the behavior of semiconductor devices more specifically, the necessary equations can be derived from MAXWELL's equations [62]. Equation (2.3) and (2.2) are used to derive the continuity equation between the charge carrier current for space charge

$$\nabla \cdot \mathbf{J} + \partial_t \rho = 0, \tag{2.10}$$

where the space charge density ρ can be described by the different contributing charge carriers and the net doping

$$\rho = -q\,(n - p - C_{\mathrm{net}}). \tag{2.11}$$

Here, q denotes the elementary charge and n, p, and C are the carrier densities for electrons and holes and the net doping density, respectively. The net doping itself can be split into acceptor and donor doping concentrations to account for different dopant species. With a carrier generation

[4] Jean Baptiste Joseph Fourier (March 21, 1768 – May 16, 1830)
[5] James Prescott Joule (December 24, 1818 – October 11, 1889)

rate G, equation (2.10) can be finally separated into equations for current densities for electrons and holes $\mathbf{J}_n$ and $\mathbf{J}_p$, respectively:

$$\nabla \cdot \mathbf{J}_n = +q\,\partial_t n + q\,G, \tag{2.12}$$
$$\nabla \cdot \mathbf{J}_p = -q\,\partial_t p - q\,G, \tag{2.13}$$
$$\mathbf{J}_n = q\,n\,\tilde{\mu}_n^{\mathrm{mob}} \cdot \mathbf{E} + q\tilde{D}_n \cdot \nabla n, \tag{2.14}$$
$$\mathbf{J}_p = q\,p\,\tilde{\mu}_p^{\mathrm{mob}} \cdot \mathbf{E} - q\tilde{D}_p \cdot \nabla p. \tag{2.15}$$

In these equations, n and p denote the carrier concentration for electrons and holes, $\tilde{\mu}^{\mathrm{mob}}$ is the carrier mobility tensor, $\mathbf{E}$ the electric field vector, and $\tilde{D}_n$ and $\tilde{D}_p$ are the diffusion coefficient tensors for electrons and holes, respectively. The equation set (2.12)-(2.15) denotes the drift diffusion model as a solution of the Boltzmann[6] transport equation using the first two moments of the distribution function, a parabolic dispersion relation, and the macroscopic relaxation time approximation [62, 63].

By introducing an electrical potential φ as

$$\mathbf{E} = -\nabla\varphi, \tag{2.16}$$

where the electric field $\mathbf{E}$ is expressed as the spatial gradient of the electrostatic potential, (2.3) reads

$$\nabla \cdot \mathbf{D} = -\nabla \cdot \left(\tilde{\varepsilon} \cdot (\nabla\varphi)\right) = \rho. \tag{2.17}$$

For homogeneous materials, the material tensor $\tilde{\varepsilon}$ can be approximated by a scalar-valued quantity ε, equation (2.17) degenerates to the well known Poisson [7] equation

$$\Delta\varphi = -\frac{\rho}{\varepsilon}, \tag{2.18}$$

where the electrical potential is determined by the space charge concentration and the dielectric constant in the matter within the specified simulation domain.

More problem-specific models have to be introduced instead of the generally used ones to described the discrepancies between reality and the observed model behavior. For instance, if the transient behavior of a clock frequency shift of an oscillator has to be considered during different operation conditions, many additional thermal and transient phenomena occur and influence the device behavior significantly [64]. Therefore, the applied simulation models have to be adapted for each particular case appropriately to achieve an accurate problem description.

The following part of this chapter gives an overview of the most important parts of the thermodynamics in semiconductor devices with respect of their application to industrial-relevant examples [62, 65–69].

2.1 Temperature

To fully describe a complex system on a microstate level, an enormous number of different microstates has to be known, and their interactions have to be determined in order to obtain the future behavior from the past states. Because is not possible to store the bulk of data

[6] Ludwig Eduard Boltzmann (February 20, 1844 – September 5, 1906)
[7] Siméon-Denis Poisson (June 21, 1781 – April 25, 1840)

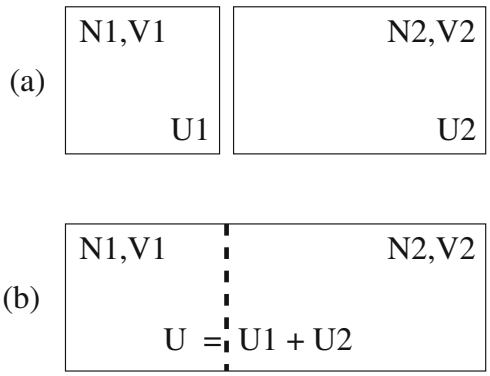


Figure 2.1: Two separately isolated subsystems (a) are brought into thermal contact (b). The number of particles and the volume remains constant for each system but the energy can be exchanged through the thermal contact.

that would be necessary to calculate all the effects correctly, a statistics-based description has to be used [70]. A possible way to obtain a representative quantity is to count the number of occupied or unoccupied microstates. Historically, the maximum number of possible states which can be theoretical occupied was chosen to determine the disorder of a system. This maximum value of disorder correlates with the energy of the system. Because the number of microstates is always a positive integer and is normally enormously large, the corresponding information content of a given system $\mathcal{S}$ can be logarithmically counted according to information theory. This introduced logarithmic quantity is the so-called entropy $\sigma(\mathcal{S})$ which represents the level of maximum disorder.

Hence, the historical definition of the entropy of a given system is a measure for the number of all possible quantum states which can be achieved following a uniform probability distribution. If the number of reachable microstates for a system $\mathcal{S}$ is determined by $N_\sigma \in \mathbb{N}$, the corresponding entropy σ of this system is defined as its natural logarithm

$$\sigma(\mathcal{S}) = \ln(N_\sigma). \tag{2.19}$$

This entropy $\sigma(\mathcal{S})$ is a function of the energy U, the number of particles N, and the volume V of the system because N_σ depends on these parameters itself.

If two systems are considered as spatially and thermally insulated systems $\mathcal{S}_1$ and $\mathcal{S}_2$, where each of them has a certain internal energy U_1 and U_2 as shown in Figure 2.1a. and if they are brought into thermal contact (cf. Figure 2.1b), the number of particles and the volumes remain constant, but the individual energies U_1 and U_2 are no longer spatially confined [69]. Therefore, an energy transmission can be observed. In this case, the total energy $U = U_1 + U_2$ remains constant if no other energy fluxes are observed. So the energy flows in the most probable case from one side to the other under the constraint that the product of the single entropies $N_{\sigma 1} N_{\sigma 2}$ maximizes. That is again a measure for the total number of states of the global system and therefore, also the sum

$$\sigma(\mathcal{S}) = \ln(N_{\sigma 1}\, N_{\sigma 2}) = \sigma_1 + \sigma_2 \tag{2.20}$$

increases. After a certain time, the energy fluctuation from one side to the other becomes zero in

average. Hence, these two systems $\mathcal{S}_1$ and $\mathcal{S}_2$ are in a state of thermal equilibrium if the equation

$$\left(\frac{\partial \sigma_1}{\partial U_1}\right)_{N_1,V_1} = \left(\frac{\partial \sigma_2}{\partial U_2}\right)_{N_2,V_2} \tag{2.21}$$

holds for the whole system $\mathcal{S}(U)$. Here, U_i and σ_i are the energy and the entropy of the system i. This property of equivalence in the thermal equilibrium is exactly what we expect to be the temperature. Therefore, the fundamental temperature τ is thermodynamically defined as

$$\frac{1}{\tau} = \frac{\partial \sigma}{\partial U} \quad \Longrightarrow \quad \tau = \frac{\partial U}{\partial \sigma}, \tag{2.22}$$

where the fundamental temperature τ has the unity of an energy. Determining τ to be the reciprocal of $\partial\sigma/\partial U$ guarantees that energy flows from the system with higher τ to the system with lower τ. The temperature T is measured in Kelvin and is proportional to τ by the equation

$$\tau = k_\mathrm{B}\, T, \tag{2.23}$$

where k_B is Boltzmann's constant and therefore, the conventional entropy S is defined as

$$S = k_\mathrm{B}\, \sigma. \tag{2.24}$$

Hence the conventional temperature can be expressed as

$$T = \frac{\partial U}{\partial S}. \tag{2.25}$$

An interesting corollary to definition (2.22) is the fact that the value zero for the fundamental temperature cannot be reached under the constraint of finite energy resources because the energy gradient would become infinity which has been proven to be impossible [71].

2.2 Heating Phenomena

In the age of the industrial revolution where the machines ran with water vapor, fundamental theorems of thermodynamics have been established to deepen the knowledge and to pursue the development of new more sophisticated machines. Research progress in electronics and material science nowadays have already enhanced knowledge in the micro and nanoelectronic regime but still, fundamental knowledge is missing for certain problems where material interactions occur, especially if coupled heating phenomena appear with chemical material interactions or volume expansions where also material phases may change. This section starts with the fundamental postulates in thermodynamics which are followed by the most important phenomena required to accurately describe modern semiconductor devices.

Zeroth Law of Thermodynamics

The zeroth law of thermodynamics describes the fundamental behavior of the temperature of systems in thermal equilibrium [72]. If two given systems $\mathcal{S}_1$ and $\mathcal{S}_2$ are in thermal equilibrium with a third system $\mathcal{S}_3$, then, the system $\mathcal{S}_1$ has also to be in thermal equilibrium with the system $\mathcal{S}_2$:

$$\left(\frac{\partial \sigma_1}{\partial U_1}\right)_{\substack{N_1\\V_1}} = \left(\frac{\partial \sigma_3}{\partial U_3}\right)_{\substack{N_3\\V_3}} \wedge \left(\frac{\partial \sigma_2}{\partial U_2}\right)_{\substack{N_2\\V_2}} = \left(\frac{\partial \sigma_3}{\partial U_3}\right)_{\substack{N_3\\V_3}} \Rightarrow \left(\frac{\partial \sigma_1}{\partial U_1}\right)_{\substack{N_1\\V_1}} = \left(\frac{\partial \sigma_2}{\partial U_2}\right)_{\substack{N_2\\V_2}}. \tag{2.26}$$

Thus, the zeroth law of thermodynamics describes the transitivity and the symmetry [60] of the equilibrium relationship [73]. According to the definition of the temperature τ from (2.22), this law can be also expressed by their fundamental temperatures τ_i

$$\tau_1 = \tau_3 \quad \wedge \quad \tau_2 = \tau_3 \quad \Rightarrow \quad \tau_1 = \tau_2, \tag{2.27}$$

where τ_i is proportional to T_i according to (2.23).

First Law of Thermodynamics

The first law of thermodynamics has been first proposed by Mayer[8] in 1841 [74]: "Heat is a kind of energy and can therefore neither be created nor destroyed." Hence, heat has been defined as the transfered thermal energy between two systems if they are brought into thermal contact [75]. The infinitesimal change of the internal energy U of a system $\mathcal{S}$ can be expressed as

$$\mathrm{d}U \;=\; \partial Q - \partial W, \tag{2.28}$$

where ∂Q is the infinitesimal heat added to the system and ∂W represents the infinitesimal work performed by the system $\mathcal{S}$. This equation follows directly from the total differential of the entropy $\mathrm{d}S$

$$\begin{aligned} \mathrm{d}S = \mathrm{d}S(U,V,N) \;&= \\ &= \frac{\partial S}{\partial U}\,\mathrm{d}U + \frac{\partial S}{\partial V}\,\mathrm{d}V + \frac{\partial S}{\partial N}\,\mathrm{d}N \;= \\ &= \frac{\partial S}{\partial U}\,\mathrm{d}U + \frac{\partial S}{\partial U}\,\frac{\partial U}{\partial V}\,\mathrm{d}V + \frac{\partial S}{\partial U}\,\frac{\partial U}{\partial N}\,\mathrm{d}N. \end{aligned} \tag{2.29}$$

With the equations (2.22)–(2.24), the total differential of the entropy becomes

$$\mathrm{d}S \;=\; \frac{1}{T}\,\mathrm{d}U + \frac{1}{T}\,p\,\mathrm{d}V + \frac{1}{T}\,(-\mu)\,\mathrm{d}N, \tag{2.30}$$

where the pressure p^{mech} can be found as

$$p^{\mathrm{mech}} \;=\; T\,\frac{\partial S}{\partial V} \;=\; \frac{\partial U}{\partial V} \tag{2.31}$$

and the chemical potential μ^{chem} as

$$\mu^{\mathrm{chem}} \;=\; -\,T\,\frac{\partial S}{\partial N} \;=\; -\,\frac{\partial U}{\partial N}. \tag{2.32}$$

By comparison of the coefficients from (2.28) and (2.30)–(2.32), the net heat flow Q and the work W done by the system $\mathcal{S}$ can be found as

$$\partial Q \;=\; T\,\mathrm{d}S \tag{2.33}$$

$$\partial W \;=\; p^{\mathrm{mech}}\,\mathrm{d}V - \mu^{\mathrm{chem}}\,\mathrm{d}N. \tag{2.34}$$

[8] Julius Robert von Mayer (November 25, 1814 – March 20, 1878)

Second Law of Thermodynamics

For a closed system which is not in thermal equilibrium, the most probable consequence is that the entropy of the system is monotonically increasing until a state of thermal equilibrium has been reached [73]. The traditional version of this theorem was given by THOMSON[9] and PLANCK[10]: A thermo-mechanical circle process which performs only a conversion of heat from a heat capacitor to its equivalent work is impossible [61]. Machines which violate this theorem are called perpetuum mobiles of second order. This can be expressed by

$$T\,\mathrm{d}S \;=\; \mathrm{d}U + p^{\mathrm{mech}}\,\mathrm{d}V - \mu^{\mathrm{chem}}\,\mathrm{d}N \quad \geq \quad 0, \tag{2.35}$$

where the equivalent of the increase of the disorder of the system $\mathcal{S}$ is greater then the maximum of the performed work of the system.

Many thermodynamic problems in microelectronics deal among others with systems in which chemical reactions are often neglected. Hence, the number of particles in the system can be assumed to be constant. So, for a given system $\mathcal{S}$ where the number of particles remains constant ($\mathrm{d}N = 0$), the second law of thermodynamics can be equivalently formulated as the so called thermodynamic identity (2.36) and its conversions (2.37):

$$T\,\mathrm{d}S = \mathrm{d}U + p^{\mathrm{mech}}\,\mathrm{d}V \tag{2.36}$$

$$\mathrm{d}U = T\,\mathrm{d}S - p^{\mathrm{mech}}\,\mathrm{d}V. \tag{2.37}$$

Hence, the energy increase $\mathrm{d}U$ of a system can be expressed by the sum of the heat transfer $T\,\mathrm{d}S$ to the system and the mechanical work $-p\,\mathrm{d}V$ done on the system.

Since every thermodynamic system can be assumed to be a dissipative system, the energy can be split into reversible and irreversible state variable

$$U \;=\; U_{\mathrm{rev}} + Q_{\mathrm{irr}}. \tag{2.38}$$

For instance, if the dissipated heat is collected and again converted to energy, the conversion process has also a finite power loss, and the originally dissipated heat cannot be completely restored, otherwise a perpetuum mobile has been constructed which is impossible because $T\,\mathrm{d}S \geq 0$ has to be fulfilled. Introducing (2.38) to (2.36) and (2.37), the thermodynamical identities can be expressed by

$$T\,\mathrm{d}S = \partial U_{\mathrm{rev}} + \partial Q_{\mathrm{irr}} + p^{\mathrm{mech}}\,\mathrm{d}V \tag{2.39}$$

$$\mathrm{d}U = T\,\mathrm{d}S - p^{\mathrm{mech}}\,\mathrm{d}V - \partial Q_{\mathrm{irr}}, \tag{2.40}$$

where the newly introduced internal energy change $\mathrm{d}U$ is reduced by the irreversible part which has been dissipated to the surrounding systems. For the reversible case, the latter equations become again (2.36) and (2.37).

Third Law of Thermodynamics

The third law of thermodynamics is often also called NERNST's[11] heat theorem. It is related to the quantum mechanical regime of a given system $\mathcal{S}$. For the special case that the temperature

[9] William Thomson, Lord Kelvin of Largs (June 26, 1824 – December 17, 1907)
[10] Max Karl Ernst Ludwig Planck (April 23, 1858 – October 4, 1947)
[11] Walther Hermann Nernst (June 25, 1864 – November 18, 1941)

reaches very low values in the zero Kelvin regime, the entropy becomes a constant value due to the finite states. The minimum number of possible microstates is one if a system is considered which consists of vacuum only, the corresponding minimum entropy is zero. Hence, the logarithmic value becomes $N_{\sigma 0} \geq 1$ and the minimum of the entropy σ is determined as

$$\lim_{\tau \to 0} \sigma(\tau) = \lim_{\tau \to 0} \ln\left(N_\sigma(\tau)\right) = \ln\left(N_{\sigma 0} \geq 1\right) = \text{const} \geq 0. \tag{2.41}$$

Therefore, the values of the entropy σ at 0 K is expected to be very small, except for some materials where a frozen internal disorder can be observed at temperatures above 0 K, e.g. in glasses and ceramics [69].

Most of the stationary heating phenomena can be described by the fundamental laws of thermodynamic. These equations can be used for the local thermal equilibrium and the local quasi thermal equilibrium if the system is not too far away from the local thermal equilibrium. The transient behavior of a system becomes more important when the investigated systems have raising numbers of uncertainties of their internal state variables. Therefore, the global energy of the system is being minimized according to the second fundamental law of thermodynamics.

Analogously to the observed behavior of fluids and gases, heat transport can be treated as a kind of energy transport, for instance within the electron gas, as it is shown in the following sections.

2.2.1 Heat Flux

It can be observed from various experiments that heat flows from the hotter to the colder side. Since the matter that is involved shows a statistical behavior at micro and nano-scale level in terms of the BROWNian molecular motion, the previous statement can be formulated as: The most probable consequence is that heat flows spontaneously from the hotter to the colder side by diffusion and relaxation mechanisms. This is exactly the definition of the second law of thermodynamics found in [73].

The time derivative of the heat can be expressed by FOURIER's and LAMBERT's[12] law, which is equivalent to the STEFAN[13]-BOLTZMANN law for grey radiators. These laws consider the spatial temperature gradient plus the heat flux density due to the surface radiation to the ambient, respectively:

$$\mathbf{q}_{\text{th}} = - \underbrace{\tilde{\lambda} \cdot \nabla T}_{\text{FOURIER's law}} - \underbrace{\sigma_{\text{SB}} \left(\varepsilon_1 T_1^4 - \varepsilon_2 T_2^4\right).}_{\text{LAMBERT's law}} \tag{2.42}$$

Here, the first term on the right hand side is determined by FOURIER's law, where the thermal conductivity tensor is denoted by $\tilde{\lambda}$ and T is the local temperature in Kelvin. The second term describes LAMBERT's law for grey radiators, where σ_{SB} denotes the STEFAN-BOLTZMANN constant and T_1 and T_2 stand for the ambient and the local surface temperature, respectively. The coefficients ε_1 and ε_2 reflect the efficiency of the absorption and the radiation of the considered surfaces. The "black body" has the absorption and radiation efficiency $\varepsilon_{\text{BlackBody}} = 1$. In most TCAD applications, the radiation can be neglected, except for areas at the surface of semiconductor devices, for instance, in passivation layers and heat sinks.

As the electric and magnetic fields store energy, also the matter stores heat energy. If heated bodies are put into a colder environment, they show a certain thermal relaxation behavior. A

[12] Johann Heinrich Lambert (August 26, 1728 – September 25, 1777)
[13] Jožef Stefan (March 24, 1835 – January 7, 1893)

possible way to describe this relaxation behavior is to assign a quantity to each material, where the value of the quantity determines how much energy can be stored per mass or per mole. This quantity is called specific heat capacitance[14]. Historically, the heat capacitance is distinguished by two types. The first one determines the heat capacitance at constant pressure C_p and the second one describes the heat capacitance at constant volume C_V:

$$C_\mathrm{p} = T\left(\frac{\partial S}{\partial T}\right)_{\mathrm{N,p^{mech}}} = \left(\frac{\partial U}{\partial T}\right)_{\mathrm{N,p^{mech}}} \tag{2.43}$$

$$C_\mathrm{V} = T\left(\frac{\partial S}{\partial T}\right)_{\mathrm{V,N}} = \left(\frac{\partial U}{\partial T}\right)_{\mathrm{V,N}}. \tag{2.44}$$

Here, the heat capacitances C_p and C_V determine the change of the internal energy U with regard to the temperature change where different constraints are applied: constant pressure and constant volume. To obtain the specific heat capacitances c_i, the heat capacitances C_i are normalized to their involved mass m:

$$c_\mathrm{i} = \frac{1}{m}C_\mathrm{i}. \tag{2.45}$$

The unit of the specific heat capacitance is either $\mathrm{J\,kg^{-1}K^{-1}}$ or $\mathrm{J\,mol^{-1}K^{-1}}$ according to the type of the mass used in (2.45) (mass or molar mass). The different values for the specific heat capacitances of a particular material can be easily transformed into each other.

Both heat capacitances of a given material, the one at constant pressure (cf. (2.43)) and the one at constant volume (cf. (2.44)), differ from each other by the identity

$$C_\mathrm{p} - C_\mathrm{V} = T\,V\,\frac{(\alpha^{\mathrm{mech}})^2}{\kappa_\mathrm{T}^{\mathrm{mech}}}, \tag{2.46}$$

where α^{mech} denotes the thermal expansion coefficient at constant pressure, and $\kappa_\mathrm{T}^{\mathrm{mech}}$ represents the isothermal compressibility coefficient of the material [76]. The thermal volume expansion coefficient α^{mech} and the compressibilities $\kappa_\mathrm{T}^{\mathrm{mech}}$ and $\kappa_\mathrm{S}^{\mathrm{mech}}$ are defined as follows:

$$\alpha^{\mathrm{mech}} = \frac{1}{V}\left(\frac{\mathrm{d}V}{\mathrm{d}T}\right)_{N,p^{\mathrm{mech}}}, \tag{2.47}$$

$$\kappa_\mathrm{T}^{\mathrm{mech}} = -\frac{1}{V}\left(\frac{\partial V}{\partial p^{\mathrm{mech}}}\right)_{T,N}, \tag{2.48}$$

$$\kappa_\mathrm{S}^{\mathrm{mech}} = -\frac{1}{V}\left(\frac{\partial V}{\partial p^{\mathrm{mech}}}\right)_{S,N}. \tag{2.49}$$

Here, the α^{mech} determines the relative volume change with regard to temperature changes, $\kappa_\mathrm{T}^{\mathrm{mech}}$ shows the relative isothermal volume change and $\kappa_\mathrm{S}^{\mathrm{mech}}$ the relative isentropic volume change with regard to changes of the local pressure.

Together with the thermodynamic potentials (2.31) and (2.32) and the thermodynamic identities (2.36) and (2.37), another correlation between heat capacitances and compressibilities can

[14] The specific heat capacitance is also called simply "specific heat".

be derived by using the chain rule for differentiation.

$$\frac{\partial(p^{\text{mech}},S,N)}{\partial(V,T,N)} = \frac{\partial(p^{\text{mech}},S,N)}{\partial(V,S,N)}\frac{\partial(V,S,N)}{\partial(V,T,N)} = \frac{\partial(p^{\text{mech}},S,N)}{\partial(p^{\text{mech}},T,N)}\frac{\partial(p^{\text{mech}},T,N)}{\partial(V,T,N)} \tag{2.50}$$

$$= \left(\frac{\partial p^{\text{mech}}}{\partial V}\right)_{S,N}\left(\frac{\partial S}{\partial T}\right)_{V,N} = \left(\frac{\partial S}{\partial T}\right)_{N,p^{\text{mech}}}\left(\frac{\partial p^{\text{mech}}}{\partial V}\right)_{T,N} \tag{2.51}$$

$$\Longrightarrow \qquad \frac{\kappa_{\text{T}}^{\text{mech}}}{\kappa_{\text{S}}^{\text{mech}}} = \frac{C_{\text{p}}}{C_{\text{V}}} = \frac{c_{\text{p}}}{c_{\text{V}}}. \tag{2.52}$$

The equations (2.50) and (2.51) show the equality of the different equivalent methods for differentiation according to the chain rules from Leibniz[15] and with the previous definitions of the compressibilities (2.48) and (2.49). Since the isobar and isochor heat capacitances describe the same region of matter, their ratio in (2.51) is the same as for the corresponding specific heat capacitances.

For isotropic and temperature-independent materials, the left hand side of (2.9) becomes λ times the Laplaceian[16] operator and (2.9) can be written as

$$\lambda \Delta T \;=\; c_{\text{p}}\,\rho_{\text{m}}\,\partial_t T - H_{\text{th}}, \tag{2.53}$$

where the maximum of the thermal conductivity has been published for carbo-nano-tubes (CNTs) and nano wires as $4.0-4.6\times10^4$ W/K in [77,78]. In comparison to that, the thermal conductivity of diamond is typically in the range of $1.0-2.5\times10^3$ W/K [79,80].

To determine the proper heat generation term H_{th} for a particular problem, several proposals have been made for semiconductor and interconnect models. The simplest model is to calculate the power loss with the local electrical field $\mathbf{E}$ and the resulting local current density $\mathbf{J}$ [81,82] by

$$H_{\text{th}} \;=\; \mathbf{E}\cdot\mathbf{J}, \tag{2.54}$$

where $\mathbf{E}$ and $\mathbf{J}$ can be calculated using the appropriate models to describe the observed behavior of the electrical field and the electrical current density. In order to account for the current densities $\mathbf{J}_n$ and $\mathbf{J}_p$ appropriately, the Seebeck[17] effect has to be considered as well, where the phenomenological semiconductor current equations [62,66] can be enhanced by

$$\mathbf{J}_n = -q\,n\,\tilde{\mu}_n^{\text{mob}}\cdot(\,\nabla\Phi_n + P_n\nabla T\,), \tag{2.55}$$

$$\mathbf{J}_p = \;\; q\,p\,\tilde{\mu}_p^{\text{mob}}\cdot(\,\nabla\Phi_p + P_p\nabla T\,), \tag{2.56}$$

where n and p represent the carrier concentrations for electrons and holes, $\tilde{\mu}_n^{\text{mob}}$ and $\tilde{\mu}_p^{\text{mob}}$ denote the mobility tensors for electrons and holes, and P_n and P_n are the Seebeck coefficients for electrons and holes. The quantities Φ_n and Φ_p represent the quasi-Fermi[18] potentials for electrons and holes in semiconductor materials:

$$\Phi_n = \varphi - V_{\text{T}}\,\ln\left(\frac{n}{n_{\text{i}}}\right), \tag{2.57}$$

$$\Phi_p = \varphi + V_{\text{T}}\,\ln\left(\frac{p}{n_{\text{i}}}\right), \tag{2.58}$$

[15] Gottfried Wilhelm Leibniz (July 1, 1646 – November 14, 1716)
[16] Marquis Pierre-Simon de Laplace (March 23, 1749 – March 5, 1827)
[17] Thomas Johann Seebeck (April 9, 1770 – December 10, 1831)
[18] Enrico Fermi (September 29, 1901 – November 28, 1954)

where φ denotes the local potential, V_T is the thermal voltage according to $V_\mathrm{T} = k_\mathrm{B}T/q$, and n_i denotes the intrinsic carrier concentration of the semiconductor material.

For semiconductor devices, the temperature T is often assumed to be the lattice temperature of the semiconductor crystal since the carriers and the lattice can be considered as two systems in thermal quasi equilibrium [66]. For a rigorous treatment of the SEEBECK effect, also the FOURIER law for the heat conduction equation (2.8) has to be adapted to

$$\mathbf{q}_{\mathrm{th},n} = -\tilde{\lambda}_n \cdot \nabla T + P_n T \mathbf{J}_n, \tag{2.59}$$

$$\mathbf{q}_{\mathrm{th},p} = -\tilde{\lambda}_p \cdot \nabla T + P_p T \mathbf{J}_p, \tag{2.60}$$

where the first part is due to FOURIER's law and the second part due to SEEBECK's effect.

2.2.2 ONSAGER's Theorem

A rigorous description of thermal influences on the electrical current and vice versa has been presented by ONSAGER[19] in 1931 [83, 84]. His theory discusses the relations of reciprocity of reversible and irreversible processes, where the coupling of the electrical and the thermal subsystems are investigated. For instance, if the electrical driving force is denoted as $\mathbf{X}_1 = \mathbf{E}$ and the thermodynamic driving force $\mathbf{X}_2$ is expressed as,

$$\mathbf{X}_2 = -\frac{1}{T}\nabla T, \tag{2.61}$$

where T has been identified as the absolute temperature by CARNOT[20] [61,85], the corresponding equation system can be formulated with independent equations as

$$\mathbf{X}_1 = R_1 \mathbf{J}_1 \tag{2.62}$$

$$\mathbf{X}_2 = R_2 \mathbf{J}_2, \tag{2.63}$$

where R_1 and R_2 are the electrical resistivity and the thermal "heat resistance", respectively. The heat resistance is also called thermal resistance R_th in this thesis. The quantities $\mathbf{J}_1$ and $\mathbf{J}_2$ are the electrical and the thermal current, respectively. The thermal current density $\mathbf{J}_2$ is also called heat flow density $\mathbf{q}_\mathrm{th}$. Several thermodynamic experiments over the last 150 years have shown that the electrical current is not independent of the temperature. Therefore, equations (2.62) and (2.63) are coupled. Introducing the standard notation, these equations can be adapted by cross coefficients R_{12} and R_{21} and represent the ONSAGER relations

$$\mathbf{X}_1 = R_{11}\mathbf{J}_1 + R_{12}\mathbf{J}_2, \tag{2.64}$$

$$\mathbf{X}_2 = R_{21}\mathbf{J}_1 + R_{22}\mathbf{J}_2. \tag{2.65}$$

For this equation system, THOMSON proposed the relation

$$R_{12} = R_{21}, \tag{2.66}$$

which is also called "reciprocity theorem" of the ONSAGER relations. However, (2.66) implies that this relation follows from symmetric principles of thermodynamic theory. Hence, the reciprocity theorem neglects the loss during heat conduction and energy conversion and relation (2.66) assumes a balanced energy flow between the two subsystems. Thus, a steady stage is assumed

[19] Lars Onsager (November 27, 1903 – October 5, 1976)
[20] Nicolas Léonard Sadi Carnot (June 1, 1796 – August 24, 1857)

with the request of (2.66) [61], where equilibrium conditions are applicable only within short range.

The principle of microscopic reversibility in (2.66) is less general than the second fundamental law of thermodynamics [83]. For further investigated coupled systems, the currents R_{ij} may have different signs due to the different directions of the energy flows. Therefore, (2.66) is not sufficient enough to fulfill the second law of thermodynamics. Hence, the necessary condition for the equation system consisting of (2.64) and (2.65) to guarantee the second law with $\partial_t S \geq 0$ yields

$$R_{12} + R_{21} \leq 2\sqrt{R_{11}\, R_{22}}. \tag{2.67}$$

This necessary condition has been originally proposed by BOLTZMANN in 1887 [86].

Writing the ONSAGER relations (2.64) and (2.65) as functions of driving forces $\mathbf{X}_i$

$$\mathbf{J}_1 = L_{11}\, \mathbf{X}_1 + L_{12}\, \mathbf{X}_2 \tag{2.68}$$
$$\mathbf{J}_2 = L_{21}\, \mathbf{X}_1 + L_{22}\, \mathbf{X}_2, \tag{2.69}$$

where the necessary condition of type (2.67) remains valid accordingly for L_{ij} as

$$L_{12} + L_{21} \leq 2\sqrt{L_{11}\, L_{22}}. \tag{2.70}$$

To consider the ONSAGER relations in terms of energy, (2.68) and (2.69) can be multiplied by $\mathbf{X}_1$ and $\mathbf{X}_2$, respectively, leading to

$$\mathbf{J}_1 \cdot \mathbf{X}_1 = L_{11}\, \mathbf{X}_1 \cdot \mathbf{X}_1 + L_{12}\, \mathbf{X}_2 \cdot \mathbf{X}_1, \tag{2.71}$$
$$\mathbf{J}_2 \cdot \mathbf{X}_2 = L_{21}\, \mathbf{X}_1 \cdot \mathbf{X}_2 + L_{22}\, \mathbf{X}_2 \cdot \mathbf{X}_2. \tag{2.72}$$

These equations represent the products of the driving forces $\mathbf{X}_i$ and displacements of types of flow $\mathbf{J}_i$. The result of (2.71) and (2.72) can be described as the dissipated energy per volume and per time and reads

$$T\theta = \sum_i \mathbf{J}_i \cdot \mathbf{X}_i, \tag{2.73}$$

where θ is the entropy generation rate per unit volume and follows from the second law of thermodynamics (2.35)

$$\partial_t S \geq 0 \quad \Longrightarrow \quad T\, \partial_t S \geq 0, \tag{2.74}$$
$$T\theta = \frac{\partial (T\, \partial_t S)}{\partial V} = \sum_i \mathbf{J}_i \cdot \mathbf{X}_i \geq 0, \tag{2.75}$$
$$\Longrightarrow \quad \theta = \frac{\partial (\partial_t S)}{\partial V} = \frac{1}{T} \sum_i \mathbf{J}_i \cdot \mathbf{X}_i \geq 0, \tag{2.76}$$

where the entropy generation rate $\partial_t S$ can be determined by the sum of the power densities of all contributing subsystems. Hereby, the parts of the sums can be identified as the power densities of the participating systems which are determined by chemical reactions, the power loss due to heat transfer and JOULE's self-heating, and the power loss due to diffusion processes. The power density of chemical reactions can be expressed by scalar-valued quantities as

$$T\theta^{\mathrm{chem}} = \sum_i J_i^{\mathrm{chem}}\, X_i^{\mathrm{chem}} = \left(\frac{1}{V}\, \partial_t \xi_j\right) \left(-\sum_j \nu_j \mu_j\right), \tag{2.77}$$

where J_i^{chem} is determined by the chemical reaction rate $\partial_t \xi_j$ per unit volume V. The chemical driving force is represented by $X_i^{\text{chem}} = -\sum_j \nu_j \mu_j$, where μ_j denotes the chemical potential and ν_j the stoichiometric coefficient of the participating atom. The electrical power density can be identified by

$$\begin{aligned} T\theta^{\text{el}} &= \sum_i \mathbf{J}_i^{\text{el}} \cdot \mathbf{X}_i^{\text{el}} = \sum_i \mathbf{J}_i \cdot \mathbf{E}_i = \\ &= -\sum_i \mathbf{J}_i \cdot \nabla \varphi_i, \end{aligned} \tag{2.78}$$

where $\mathbf{J}$ and $\mathbf{E}$ are the electrical current density and the electrical field, respectively. The electric field can be expressed by the spatial gradient of an electrical potential φ. For electro-magnetical subsystems, the power density has to be appropriately adapted as discussed in Section 2.2.3. Another important contribution to the global entropy increase is the power loss due to thermal heat flow, which can be expressed in terms of (2.73) by

$$\begin{aligned} T\theta^{\text{heat}} &= \sum_i \mathbf{J}_i^{\text{heat}} \cdot \mathbf{X}_i^{\text{heat}} = \sum_i \mathbf{q}_{\text{th},i} \cdot \left(T_i \nabla \left(\frac{1}{T_i} \right) \right) = \\ &= \sum_i \mathbf{q}_{\text{th},i} \cdot \left(-\frac{1}{T_i} \nabla T_i \right), \end{aligned} \tag{2.79}$$

where $\mathbf{q}_{\text{th}}$ represents the local heat flux density. The second term in (2.79) depicts the thermal driving force according to FOURIER's empirical law. For diffusion processes, the power density can be identified as

$$\begin{aligned} T\theta^{\text{diff}} &= \sum_i \mathbf{J}_i^{\text{diff}} \cdot \mathbf{X}_i^{\text{diff}} = \sum_i \mathbf{J}_i^{\text{mol}} \cdot \left(-T_i \nabla \left(\frac{\mu_i}{T_i} \right) \right) = \\ &= \sum_i \mathbf{J}_i^{\text{mol}} \cdot \left(-\nabla \mu_i + \frac{\mu_i}{T_i} \nabla T_i \right), \end{aligned} \tag{2.80}$$

where $\mathbf{J}_i^{\text{mol}}$ is the mole number per unit area and time of the contributing species i. The driving force of diffusion processes is determined by the gradient of the chemical potential and by the gradient of the temperature.

Since the power density of diffusion processes has been determined for no external forces, an extension for the applied electrical field has to be made by introducing an additional term that depicts the force acting on charged particles inside the simulation domain. Hence, equation (2.80) has to be modified as

$$T\theta = \sum_i \mathbf{J}_i^{\text{mol}} \cdot \left(-\nabla \mu_i + \frac{\mu_i}{T_i} \nabla T_i - Z_i^* \, n_i \, \nabla \varphi_i \right), \tag{2.81}$$

where Z_i^* is the effective valence charge of the species i, n_i is the species concentration per mole, and φ_i is the corresponding electrical potential.

To conclude ONSAGER's thermodynamical treatment, the overall power density is thus given by the sum of the power densities of all contributing subsystems as

$$T\theta = T\sum_i \theta_i \;=\; T\left(\theta^{\text{chem}} + \theta^{\text{el}} + \theta^{\text{heat}} + \theta^{\text{diff}}\right) \tag{2.82}$$

$$\begin{aligned} T\theta = &-\frac{1}{V}\,\partial_t \xi_j \sum_j \nu_j \mu_j \\ &- \sum_i \mathbf{J}_i \cdot \nabla \varphi_i \\ &+ \sum_i \mathbf{q}_i \cdot \left(-\frac{1}{T_i}\,\nabla T_i\right) \\ &+ \sum_i \mathbf{J}_i^{\text{mol}} \cdot \left(-\nabla \mu_i + \frac{\mu_i}{T_i}\nabla T_i - Z_i^*\, n_i\, \nabla \varphi_i\right), \end{aligned} \tag{2.83}$$

where the thermodynamic power density $T\theta$ is determined by the contributing chemical reactions, the electrical burden, heat flows, and molar diffusion processes.

2.2.3 Electro-Magnetic Power Density

The local power density is one of the most important quantities that determine the maximum performance or the maximum number of device operations per time. Moreover, power has been concentrated at certain local regions and has therefore to be controlled appropriately to perform the desired action within the defined requirements. Unfortunately, power has also a dissipative part. Since the electrical and mechanical systems have only finite efficiencies, the usable power is less than the total power consumption. The difference dissipates via thermal conduction, convection, and radiation. In addition to that, the second law of thermodynamics postulates the irreversibility of some of the thermally dissipated heat. Hence, the entropy is steadily increasing on average. The appropriate power density for heat conduction can be derived from the local energy of the electro-magnetic fields, where the energy of these fields is determined by the POYNTING[21] vector. The spatial power source density of the POYNTING vector represents the local energy density [58,59], which can be directly derived from the curl equations of MAXWELL's equations. Multiplying the equations (2.1) and (2.2) from the left with -**H** and **E**, respectively, results in

$$-\mathbf{H} \cdot (\nabla \times \mathbf{E}) = \mathbf{H} \cdot \partial_t \mathbf{B}, \tag{2.84}$$

$$\mathbf{E} \cdot (\nabla \times \mathbf{H}) = \mathbf{E} \cdot \mathbf{J} + \mathbf{E} \cdot \partial_t \mathbf{B}. \tag{2.85}$$

The sum of the left sides of equations (2.84) and (2.85) shows the local source density of POYNTING's vector:

$$\mathbf{E} \cdot (\nabla \times \mathbf{H}) - \mathbf{H} \cdot (\nabla \times \mathbf{E}) \;=\; \nabla \cdot (\mathbf{E} \times \mathbf{H}) \;=\; \nabla \cdot \mathbf{S}, \tag{2.86}$$

with the POYNTING vector **S** as

$$\mathbf{S} \;=\; \mathbf{E} \times \mathbf{H}. \tag{2.87}$$

The sum of the right sides of equations (2.84) and (2.85) represents the equivalents to (2.86):

$$\nabla \cdot \mathbf{S} \;=\; \mathbf{E} \cdot \mathbf{J} + \mathbf{E} \cdot \partial_t \mathbf{D} + \mathbf{H} \cdot \partial_t \mathbf{B}. \tag{2.88}$$

[21] John Henry Poynting (September 9, 1852 – March 30, 1914)

Equation (2.88) represents the local form of the energy conservation equation. The left side of (2.88) shows the local source density of the POYNTING vector which depicts the current change of energy density per time (power density). The right side shows the different components of the contribution. The first term $\mathbf{E}\cdot\mathbf{J}$ is the electric component which represents the JOULE power loss that causes self-heating due to carrier transport[22] mechanisms. The second and the third term depict the change of the energy stored in the electrical and the magnetical field, respectively.

For isotropic and field-independent materials, (2.88) can be formulated as

$$\begin{aligned}\nabla\cdot\mathbf{S} &= \sigma\mathbf{E}\cdot\mathbf{E} + \varepsilon\mathbf{E}\cdot\partial_t\mathbf{E} + \mu\mathbf{H}\cdot\partial_t\mathbf{H}\\ &= \sigma\mathbf{E}^2 + \frac{1}{2}\varepsilon\,\partial_t\left(\mathbf{E}^2\right) + \frac{1}{2}\mu\,\partial_t\left(\mathbf{H}^2\right).\end{aligned} \tag{2.89}$$

With JOULE's power loss equation

$$p = \mathbf{E}\cdot\mathbf{J} = \sigma\mathbf{E}^2, \tag{2.90}$$

(2.91)

and the introduction of w^{el} and w^{mag} as the electrical and magnetical energy densities

$$w^{\mathrm{el}} = \frac{1}{2}\,\varepsilon\,\mathbf{E}^2 \tag{2.92}$$

$$w^{\mathrm{mag}} = \frac{1}{2}\,\mu\,\mathbf{H}^2, \tag{2.93}$$

equation (2.89) can be written as

$$\nabla\cdot\mathbf{S} = p + \partial_t w^{\mathrm{el}} + \partial_t w^{\mathrm{mag}}. \tag{2.94}$$

In comparison with common effects in semiconductor devices, JOULE's power loss has often to be taken into account to describe self-heating. In conjunction with semiconductor devices, there are several other approaches to determine the power loss. Since electrons and holes behave differently in semiconductor materials, the current is conveniently split up in order to account for their different properties. In addition, the potential confinement of the carriers is also different. Therefore, the power densities for electrons and holes can be determined with the appropriate models.

2.2.4 Global versus Local Heating

Beginning with the local point of view, the heat conduction equation (2.9) determines the behavior at the current position and in a certain small region around the current position. In this environment, the heat is generated, for instance, if charge carriers are forced to move and are accelerated according to an external applied electric field. Hence, also the kinetic energy of the carriers increases. According to elastic and inelastic scattering effects at lattice sites, impurities, and surface areas, parts of the carrier's kinetic energy are converted to heat. Therefore, a temperature can be assigned to charge carriers, where these assigned carrier temperatures directly correspond to their energies but does not correlate with the commonly known thermodynamic temperature. Especially for charge transport in semiconductor materials, the carrier temperature assignments are very common [87] to determine carrier energy (cf. hot and cold electrons in [62, 63, 72]).

[22] Transport of electrons, holes, and ions

For small carrier flows in semiconductor devices, the semiconductor substrate temperature and therefore also the channel lattice temperature can be assumed to be equal to the ambient temperature, since self-heating is negligible.

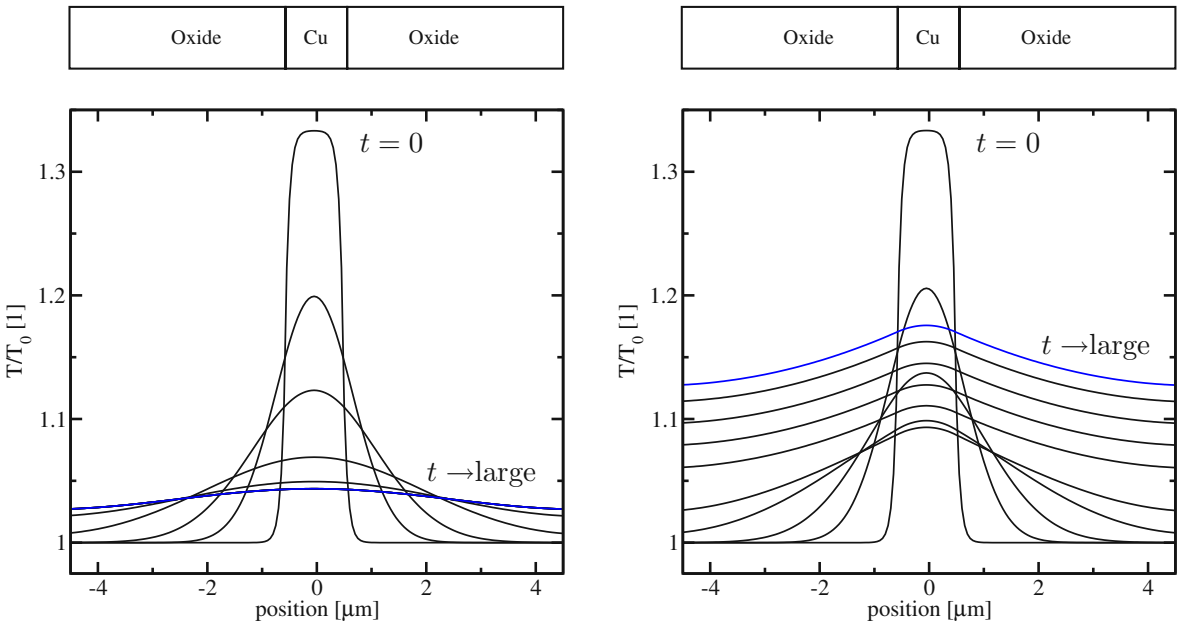
To determine the heating effects, especially for interconnect materials, most involved materials can be assumed to behave like metals. Since highly doped semiconductor materials and also silicides have been introduced to interconnect structures and their resistivity can be adjusted within certain ranges, also their electrical behavior is quite similar to that of metals. Hence, these materials can be treated as metal-like materials, or, if special material-specific effects have to be considered, the model can be extended to account for this particular effect. For instance, highly doped tungsten silicide is often used as a conductor. Hence, it is possible to use freely available regions to integrate fusing elements that exploit electro-migration effects in tungsten silicide [88–90].

As an illustrating example for heating and cooling, a copper plate embedded in SiO_2 dielectrics has been heated up to a certain temperature and the corresponding cooling procedure is shown in Figure 2.2a. This figure depicts an undisturbed thermal relaxation (cooling) where the heat energy and therefore the temperature diffuses to the surrounding oxide layers to become uniformly distributed over the simulation region. For the simulations shown in Figure 2.2, adiabatic boundary conditions are applied where the simulation region does not exchange energy which any other surrounding systems. This can for instance be assumed if this particular simulation domain is embedded in a large periodic structure. In contrast, by considering a continously heated copper plate in Figure 2.2b, the result of the heat conduction equation (2.9) gives0 the evolution of the temperature distribution at different points in time with an additional heat source term H_{th} applied. The heat becomes distributed over the whole simulation domain, where heat conduction is the dominant energy transport mechanisms. Radiation has been neglected in this investigation. As a result, the temperature distribution shows the same behavior at the beginning as without self-heating, but at a certain point in time, the surrounding material is no longer able to absorb the produced heat and the whole structure begins to heat up.

A similar phenomenon can be observed in interconnect structures with included vias. The heat produced by self-heating is preferably transfered through the metal layers and the via structures because the thermal conductivity of the interconnect lines and vias is much higher than those of the barrier layers and the surrounding dielectrics. Since the vias are made of metal and silicided materials, the vias provide a strong secondary heat transfer path in addition to the metal layers, which is often not considered and anticipated appropriately. As a consequence, measurements show that real structures globally heat up during operation and thermal failures may occur which seem to be unmotivated in terms of electrical simulation. With simulators that use suitable thermal models, such possible failures can be identified.

The high supply voltages and increasing clock frequencies are another promoting factor for additional thermal loads because the intervals between the loading and decharging of the intrinsic capacitors are becoming shorter. This requires to operate at higher current densities. In addition, this phenomenon is emphasized when CMOS device structures are considered. During a transition, when the device is turned on or off, both driver transistor are for a short period of time in the on-position. That means, a higher current flows through the devices from the positive supply connector to ground and thus the power loss increases quadratically [91]. This effect can be observed when only frequency scaling is applied for a particular technology.

Rigorous self-consistent electro-thermal investigations of complex interconnect structures have become very time consuming, especially when temperature-dependent material parameters are considered. For these investigations, the three-dimensional interconnect simulator STAP [39, 92,



(a) Thermal relaxation with $H_{\text{th}} = 0$.

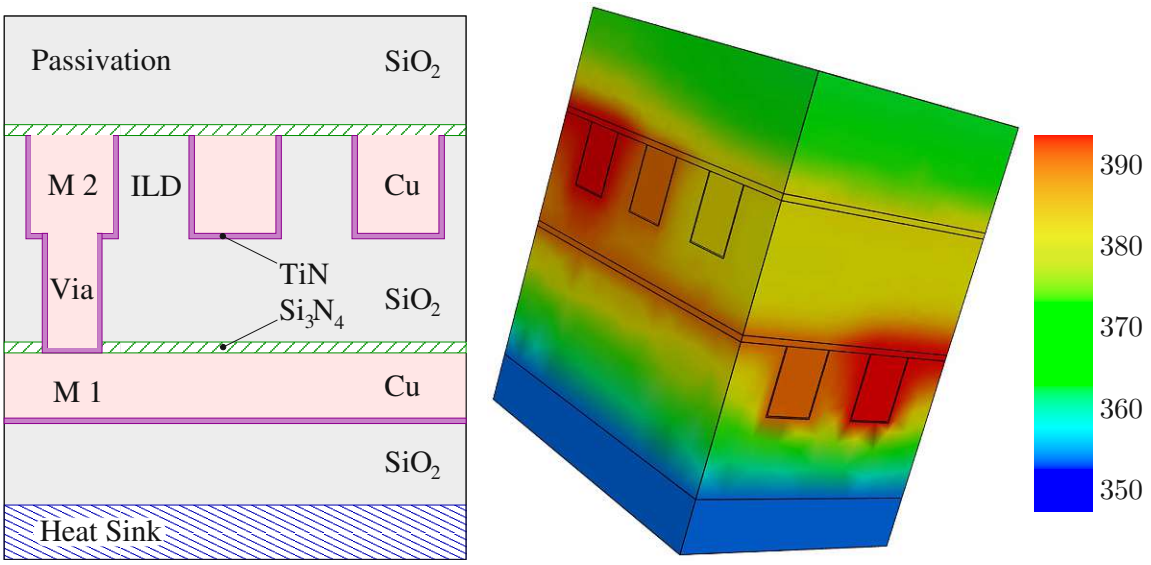
(b) Thermal relaxation with $H_{\text{th}} > 0$.

Figure 2.2: Temperature distribution showing the thermal relaxation of a heated SiO_2-Cu-SiO_2 structure where different heat source terms are applied. The left part (a) shows a simple relaxation without heat source and the right part (b) with an additional power density inside the copper plate as it can be observed in a conducting material due to self-heating.

93] has been developed to extract resistances, capacitances as well as internal quantity distributions such as potential, current density, and temperature, among others.

As a representative example, a typically multi-layered interconnect structure is electro-thermally investigated. Figure 2.3a depicts a structure which consists of several metalization layers which are connected through vias. In between these metal lines, there is SiO_2 as an interlayer dielectric (ILD). The copper lines are coated by TiN barrier layers to avoid copper diffusion into SiO_2. As an etch-stop layer, a Si_3N_4 passivation layer is included due to technology-induced constraints. The heat flows through the paths with the highest thermal conductivities of the different contributing materials. The electrically stressed metal lines produce heat due to JOULE's power loss and the generated heat establishes a certain temperature in the metal lines and the surrounding materials which results in propagation of the heat through the whole interconnect structure. The passivation layer in between the lower metal layer and the upper ILD layer shows a high thermal conductivity path which causes a rapid heating of adjacent regions of that part. As a logical consequence, changes in material compositions could overcome this type of unintentional heat evolution. This change is highly demanded in terms of the electrical and mechanical improvements but underlies many technology-related and economical-related constraints.

As Figure 2.3 shows, the electrical load in semiconductor devices and in interconnect structures produces a considerable amount of heat which has to be included into the design of heat sinks. Due to global heating, many interconnect regions become thermally and mechanically stressed even these which are not directly electrically stressed. Therefore, these regions become extraor-



(a) Material composition. (b) Temperature distribution [K].

Figure 2.3: A typical multi-layered dual damascene interconnect structure made of various materials (a) showing global warming in metalization level M 2 (b).

dinary burdened by further electrical or mechanical stressing effects. Hence, the heat metal lines and metalization layers can be used to shield some regions from external heating by providing an additional heat conduction path with a very low thermal resistivity allowing a fast transport path towards a heat sink.

2.2.5 Heat Sinks and Sources

For the simulation of self-heating, the generated heat is modeled through JOULE's power loss. Hence, each interconnect line with an electrical burden represents a distributed heat source which is modeled via an additional heat source term H_{th} in the local heat conduction equation (2.9).

As initial conditions, the ambient temperature at $T_0 = T(t = t_0)$ is chosen. Typical initial values for T_0 are 300 K for room temperature and 330 K or 350 K for heated device structures which have already reached their stationary operational conditions.

An ideal heat sink provides a constant temperature at a certain part of the device structure. Therefore, the boundary condition for the temperature can be modeled by a DIRICHLET[23] boundary condition

$$\forall \mathbf{x} \in \partial\Omega_1 : T(\mathbf{x}) \; = \; T_0, \tag{2.95}$$

where Ω represents the simulation domain and $\Omega_1 \in \Omega$ a part of the simulation. $\partial\Omega_1$ is the corresponding boundary related to the ideal heat sink. This represents a very good assumption if actively cooled heat sinks are considered.

[23] Peter Gustav Lejeune Dirichlet (February 13, 1805 – May 5, 1859)

Adiabatic boundary conditions can be used, if only a single part of a device is sufficient to describe the device behavior due to symmetry of the device structure [94]. The adiabatic boundary condition can be expressed by homogeneous NEUMANN[24] boundary conditions

$$\forall \mathbf{x} \in \partial\Omega_2 : \tilde{\lambda} \cdot \nabla T(\mathbf{x}) = \mathbf{q}_{\mathrm{th},0}(\mathbf{x}) = \mathbf{0}. \tag{2.96}$$

External heat sources, for instance, at boundaries of the simulation domains $\partial\Omega_3$, require inhomogeneous NEUMANN boundary conditions described by

$$\forall \mathbf{x} \in \partial\Omega_3 : \tilde{\lambda} \cdot \nabla T(\mathbf{x}) = \mathbf{q}_{\mathrm{th},0}(\mathbf{x}), \tag{2.97}$$

where $\mathbf{q}_{\mathrm{th},0}(\mathbf{x})$ is the externally applied heat flux density, this might be for example a heat generating power line or an active cooling element.

Applying additional (fast) heat diffusion paths influences the original heat flux distribution and might result in new and sometimes unwanted heating effects at their surrounding material environments. Therefore, a rigorous investigation of the thermal influence has to be included for modern chip design because even cooling a semiconductor device structure may cause additional mechanical stress as has been outlined in [95].

2.2.6 Electro-Thermal Coupling

As it was shown in the previous sections, the temperature is nearly always involved in parasitic effects and the temperature is mostly a result of electrical load. However, the temperature is a natural result of a system which dissipates power since it represents the change of the internal energy of the system as previously presented at the beginning of this chapter.

A common modeling approach for the heat source term in electro-magnetic systems can be expressed by (2.88) as

$$H_{\mathrm{th}} = \mathbf{E} \cdot \mathbf{J} + \mathbf{E} \cdot \partial_t \mathbf{D} + \mathbf{H} \cdot \partial_t \mathbf{B}, \tag{2.98}$$

where the first term shows the power loss density according to the applied electric field, where the second and the third terms present the internal power density of the electric and magnetic field, respectively.

This rigorous determination of the power density H_{th} couples the electro-magnetic and the thermal subsystems yields (2.9). For special materials which show for instance PELTIER[25] and SEEBECK effects, equation (2.98) has to be appropriately expanded by introducing an additional term accounting for these effects as proposed in [61, 66] and reads

$$-\mu^{\mathrm{th}}\, \mathbf{J} \cdot \nabla T, \tag{2.99}$$

where the THOMSON coefficient μ^{th} is defined by

$$\mu^{\mathrm{th}} = T \frac{\partial P_{\mathrm{th}}}{\partial T}, \tag{2.100}$$

using the thermopower coefficient P_{th}, which is determined by the change of the contact voltage with respect to the temperature change [72]

$$P_{\mathrm{th}} = \frac{\partial V_{\mathrm{c}}}{\partial T}. \tag{2.101}$$

[24] Carl Gottfried Neumann (May 7, 1832 – March 27, 1925)
[25] Jean Charles Athanase Peltier (February 22, 1785 – October 27, 1845)

2.2.7 Volume Expansion

A rather computational expensive but important effect for modern electronic devices is the volume expansion of condensed matter due to a temperature increase. From the microscopic point of view, the BROWNian motion increases with rising temperature. Therefore, the average distance of the particles at higher temperatures is larger than at lower temperatures. Hence, the volume of the unit cell has been expanded. For semiconductor devices that operate at nearly constant and moderate temperatures, a certain thermal volume expansion can be anticipated during the device design.

Because most design rules consider thermal equilibrium conditions between the operating device and its ambient, additional transient heat source terms which result for instance from self-heating can drastically enhance the volume expansions and result in wrong predictions of the device characteristics at certain operation conditions.

Generated heat in modern interconnect structures as well as in semiconductor devices can be taken into account by a certain average heat source distribution but this approach is still limited to experience of the past implementations and experiments. This is a major concern of modern reliability investigations, especially for novel interconnect structures and for high performance devices such as microprocessors and power amplifiers.

To account for the thermal volume expansion behavior, a first order approach

$$\frac{\partial V}{V} = \alpha^{\mathrm{mech}}\, \partial T \tag{2.102}$$

is commonly used [96,97]. Here, the volume of the material is denoted by V and α^{mech} represents the thermal volume expansion coefficient.

Some simulation tools might use higher order *Ansatz* functions, hence (2.47) can be adapted to

$$\frac{V^{\mathrm{expand}}}{V^{\mathrm{normal}}} = \left(1 + \sum_{i}^{\mathrm{N}} \alpha_i^{\mathrm{mech}} (T - T_0)^i\right), \tag{2.103}$$

where N is the order of the *Ansatz* functions for the mechanical expansion model and α_i^{mech} is the appropriate material parameters for the volume expansion model. The reference temperature T_0 is typically either 300 K or 273.15 K. Hence, the values for α_i^{mech} are available for different reference temperatures because this approach refers to a truncated TAYLOR[26] expansion series.

To determine the elongation of a one-dimensional body, e.g. a rod, the one-dimensional expansion coefficient can be derived from the volume expansion coefficient α^{mech} from its definition

$$l = l_0 + \mathrm{d}l \quad \Rightarrow \quad V = V_0 + \mathrm{d}V \tag{2.104}$$

$$l = l_0\,(1 + \alpha_{\mathrm{l}}^{\mathrm{mech}} \mathrm{d}T) \quad \Rightarrow \quad V = V_0\,(1 + \alpha_{\mathrm{l}}^{\mathrm{mech}} \mathrm{d}T)^3. \tag{2.105}$$

Using a TAYLOR expansion for (2.105) yields

$$V \approx V_0\,(1 + 3\alpha_{\mathrm{l}}^{\mathrm{mech}} \mathrm{d}T) + \mathcal{O}\left((\alpha_{\mathrm{l}}^{\mathrm{mech}})^2\right). \tag{2.106}$$

Hence, a coefficient comparison yields for small expansion coefficients the relation

$$\alpha^{\mathrm{mech}} = 3\alpha_{\mathrm{l}}^{\mathrm{mech}}. \tag{2.107}$$

[26] Brook Taylor (August 18, 1685 – December 29, 1731)

The impact of a thermal expansion is manifold. If the volume is able to change its size, and there are no confinements to the volume, the consequence can be either shrinking or expansion. The more general case applies mostly: Confinements like adhesion and constrictions usually determine the volume expansion. Hence, the volume is bared by some other materials, which are not that compressible or elastic as required to allow an unconfined expansion. Therefore, strain and stress occur which are the mechanical driving forces for tension, contraction, torsion, and other effects including a change of the crystal structure, or phases.

2.2.8 Mechanical Subsystem

This thesis deals with mechanical phenomena mainly caused by electro-thermal stress conditions. Since the electrical burden produces heat and the heat non-negligible volume expansion, the mechanical part has to be considered as well. The basic equation used for TCAD purposes is HOOKE's[27] law which has been originally introduced by the words *"Ut tensio sic vis"*[28]. The corresponding formula reads

$$F = c\,u, \tag{2.108}$$

where the absolute value of the applied force F to a body is proportional to its elongation u. Here the constant c determines the stiffness of the body. More generally, HOOKE's law can be formulated for local quantities in a body where the local stress tensor $\tilde{\sigma}^{\text{mech}}$ is associated to the GREEN[29] tensor (local strain tensor) $\tilde{\varepsilon}^{\text{mech}}$ for a given body by

$$\tilde{\sigma}^{\text{mech}} = \tilde{C} \cdot \tilde{\varepsilon}^{\text{mech}}, \tag{2.109}$$

$$\sigma_{ij}^{\text{mech}} = \sum_{k,l} C_{ijkl}\, \varepsilon_{kl}^{\text{mech}}, \tag{2.110}$$

where the proportionality factor is determined by the 4^{th}-rank stiffness tensor $\tilde{C}$ and the strain is defined according to CAUCHY[30] via local displacements

$$\varepsilon_{ij}^{\text{mech}} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right), \tag{2.111}$$

where $\mathbf{u}$ is the displacement or deformation vector and $\mathbf{x}$ the local position.

Using the VOIGT[31] notation [96–98], the ranks of the tensors involved in (2.109) can be reduced due to the symmetry of the material and due to the symmetry according to energy conservation laws [59]. Thus, the number of independent tensor entities reduces from 81 to 36 by material symmetry and further reduces to 21 mutual independent tensor entities due to energy conservation [59, 96]. Therefore, equation (2.109) can be expressed as

$$\hat{\sigma}^{\text{mech}} = \hat{C} \cdot \hat{\varepsilon}^{\text{mech}}, \tag{2.112}$$

where $\hat{\sigma}^{\text{mech}}$ and $\hat{\varepsilon}^{\text{mech}}$ are the vector-valued quantities for the mechanical stress and strain in the VOIGT notation, respectively. Furthermore, $\hat{C}$ represents the stiffness matrix of 2^{nd} rank also in the VOIGT notation.

[27] Robert Hooke (July 18, 1635 – March 3, 1703)
[28] "As the tension, so is the force."
[29] George Green (July 14, 1793 – May 31, 1841)
[30] Augustin Louis Cauchy (August 21, 1789 – May 23, 1857)
[31] Woldemar Voigt (September 2, 1850 – December 13, 1919)

In TCAD applications of modern devices it is often sufficient to deal with static stresses, only. In that cases, the speed of involved particles can be neglected [99]. The mechanical equations have to fulfill general conservations laws [96,97] for energy, momentum, angular momentum, and mass. Thus, the mechanical subsystem can be described by the local conservation laws of energy, momentum, and mass

$$\partial_t w + \nabla \cdot \mathbf{q} = \mathbf{S}^{\text{mech}}, \tag{2.113}$$

$$\partial_t \mathbf{g} + \nabla \cdot p^{\text{mech}} = \mathbf{f}, \tag{2.114}$$

$$\partial_t \varrho^{\text{mass}} + \nabla \cdot (\mathbf{v}\, \varrho^{\text{mass}}) = 0. \tag{2.115}$$

In (2.113) the local energy density is denoted by w, the energy flux density by $\mathbf{q}$, and $\mathbf{S}^{\text{mech}}$ represents the mechanical power density. The latter equation represents the mechanical analogon of the POYNTING vector. Equation (2.114) is the momentum conservation equation, where $\mathbf{g}$ is the momentum density, $\mathbf{f}$ is the local force density, $\mathbf{v}$ the velocity of the moving particles, and $\tilde{p}^{\text{mech}}$ is the momentum flux density which is often called pressure tensor. Equation (2.115) presents the local mass continuity equation, where the specific mass density is denoted as ϱ^{mass}. If mass fluxes have to be considered, for instance in electro-migration analysis, the kinetic pressure tensor $\tilde{p}^{\text{mech}}$ becomes

$$\tilde{p}^{\text{mech}} \;=\; \varrho^{\text{mass}}\, \mathbf{v} \otimes \mathbf{v} - \tilde{\sigma}^{\text{mech}}, \tag{2.116}$$

where the specific mass density is denoted by ϱ^{mass}, $\tilde{\sigma}^{\text{mech}}$ is the stress tensor, and $\mathbf{v}$ is the speed of the moving particles.

Later on, $\tilde{p}^{\text{mech}}$ can be also used as a scalar-valued quantity p^{mech} when the simplified VOIGT notation is used:

$$p^{\text{mech}} \;=\; \text{trace}(\varrho^{\text{mass}}\, \mathbf{v} \otimes \mathbf{v} - \tilde{\sigma}^{\text{mech}}). \tag{2.117}$$

If the flux of mass has not to be considered, the associated velocity of the particles becomes $\mathbf{v} = 0$ and the hydrostatic pressure can be determined by

$$p^{\text{mech}} \;=\; -\,\text{trace}(\tilde{\sigma}^{\text{mech}}). \tag{2.118}$$

The definitions of hydrostatic pressure p^{mech} in (2.117) and (2.118) can be used as a metric which provides a possibility to visualize, to compare within measurements, or to define a figure of merit in an optimization loop [100].

If moving particles are considered, the mechanical analogon to the electrical continuity equation is the mass continuity equation (2.115) and can be treated with the EULERian[32] continuity equation [96]

$$\partial_t \varrho^{\text{mass}} + \varrho^{\text{mass}}\, \nabla \cdot \mathbf{v} + \mathbf{v} \cdot \nabla \varrho^{\text{mass}} \;=\; 0, \tag{2.119}$$

which is the mass conservation equation (2.115). A mathematical coupling between the mass flux and the mechanical stress can be obtain by using the first law for continuity mechanics from CAUCHY

$$\varrho^{\text{mass}}\, (\partial_t \mathbf{v} + \mathbf{v} \cdot \nabla \mathbf{v}) \;=\; \nabla \cdot \tilde{\sigma}^{\text{mech}} + \mathbf{f}, \tag{2.120}$$

where the $\mathbf{f}$ represents the externally applied force density.

[32] Leonhard Euler (April 17, 1707 – September 18, 1783)

2.2.9 Micromechanics in TCAD

The materials involved in typical microelectronic systems are subject to very strong mechanical constrictions due to the rigorous embedding in hard and rigid material compounds. Thus, the microscopic behavior can be assumed to be in steady state in which the mechanical movements can be neglected. Of course, there are still diffusion processes involved, but their time scale, e.g. their diffusion constant, is much smaller than the investigated time regime to determine the mechanical burden. Hence, the thermo-mechanical and electro-thermal investigations can be independently executed very efficiently using tuned simulators for each of the sub-problems. After a certain time, at which the mechanical burden has reached a critical value, the thermo-mechanical sub-system has to be updated to follow a self-consistent iteration scheme for the globally coupled electric, thermal, mechanical problem.

Microelectronical and micromechanical (MEMS) devices often use micro-scaled phenomena of coupled electro-thermal and thermo-mechanical subsystems including also mass flows. Therefore, the assumption of velocity-free material regions remains no longer valid for these systems and a rigorous treatment of the mechanical and electrical phenomena has to be performed.

However, TCAD applications require micromechanical considerations beside rigorous thermo-electrical analysis as recent survey and forecast have shown to overcome scaling-induced problems, e.g. enhanced thermal stress cycles, electro-migration, and current densities [3]. Hence high mechanical demands have appeared not only for devices but also for interconnect structures and chip packages. According to recently announced requirements by the ITRS [3], thermal issues have reached certain critical levels which make an additional micromechanical analysis necessary to fulfill the challenging ITRS goals proposed for the future.

The mechanical strain on various materials shows different effects. Device engineers have successfully developed some applications where the strain can be advantageously used. This strain engineering has been implemented in state-of-the-art technology nodes in semiconductor device fabrication for instance to align the carrier charge mobilities in nMOS and pMOS transistors. In addition, the enhancement factor an be adapted to a certain desired level that the mobility for both, the electrons in the nMOS transistors and the holes in the pMOS transistors have the same values which idealizes CMOS circuits loss during transistors switching [101]. This technique used the advantageous crystal structure of semiconductor materials, for instance Si and SiGe. In this particular example, strained materials are introduced to CMOS structures. Since the charge carriers in these transistors have different mobility tensors, the resistivity as well as the transition times of the transistors can be adjusted to customize the overall performance of the circuit [102,103]. The main application of strain engineering in the semiconductor device regime deals with mobility enhancements and the equalizing of carrier mobilities of nMOS and pMOS transistors in CMOS circuits due to compressive and tensile stress profiles [102, 103].

However, strain engineering is not limited to mobility enhancements in CMOS circuits. Many different materials are currently under development for various applications, e.g. SiGe [103], SiC [104], GaN [105–108]. A promising application for strain engineering is the control of optical and electro-magnetical properties as well as the exploitation of their anisotropy for future device applications [106].

However, modern microelectronic devices are very sensitive to variations of stress levels in certain device layers, e.g. in the channel of a transistor or at edges of thin film dielectrics. The stress changes the lattice configuration slightly and therefore also the bandgap and the mobility inside the channel. Also, the breakdown voltage of the dielectric material can be dramatically influenced [109]. As already mentioned, one consequence of mechanical strain is the lattice de-

formation which can be exploited to enhance the charge carrier mobility or to slightly change the bandgap in semiconductor and dielectric materials [109]. Unfortunately, these two effects appear unexpectedly during fabrication processes or under critical operation conditions where the device characteristics and/or the device performance is permanently changed. For instance, if a rather thick SiGe layer over a Si substrate is deposited, the intrinsic lattice constant of Si is forced to a few tens SiGe atomic layers, where the enforcement to the SiGe lattice distance reduces with increasing distance form the Si-SiGe interface. If in this case additional stress occurs in combination with high temperatures due to high work load. The intrinsic stress due to the lattice mismatch between the SiGe and Si can be loose mechanical contact due to adhesion loss at the interface or due to cracks in the SiGe layer. Another unfavorable effect due to mechanical strain are drifting ions since the deformed crystal lattice provides lower activation energies for ion diffusion and moving ions out of their lattice site. Hence, the temperature as well as the mechanical strain have to be considered carefully to obtain device structures which can act also under high loads and at high temperatures within reliability requirements.

2.2.10 Electro-Mechanical Coupling

Transient calculations of electrical or mechanical problems are challenging on their own. If rigorous coupling of these two systems is considered, the effort increases dramatically, e.g, different requirements of mesh generation for critical regions have to be considered for the electrical simulation as well as for the mechanical simulation. In particular, for regions which do not mainly influence the current distribution or the electric field, a coarse mesh can be applied for the electrical analysis in these regions. However, these regions may be critical ones during the mechanical investigation and might thus be essential for the global device characteristics.

Much effort for TCAD analysis is due to the increasing importance of thermal issues in microelectronic device structures. Therefore, the analysis for the electrical and mechanical problems becomes even more challenging because of the complex simulation tasks and the difficulty to couple the sub-systems self-consistently. Also, mechanical and geometrical constraints as well as boundary conditions are influenced by the temperature and raise the grade of complexity. For instance, if the local temperature increases, it forces the materials to expand and thus to increase the local stress or to perform deformations at weak material interfaces.

Therefore, viscoelasticity can be assumed as long as the material is in the linear regime because after removing the external stress the materials relaxes until a residual stress level. The remaining stress level is relaxes due to diffusion processes of the atoms in the material. Within this small linear regime, the material parameters change only within a small range. But above this certain threshold value, inelastic deformations occur and also the material parameter might change abruptly, as it is shown in Figure 2.4. The materials used in microelectronics show worse characteristics than bulk materials because the thicknesses are in the range of tens of nanometers, where the microscopic grain structures as well as material interfaces drastically influence the material parameters.

In Figure 2.4 the mechanical strain-stress curves of two typical materials are presented, where the normal operation conditions are typically located in the low linear regions. When the temperature increases, the volume expands and additional strain $\varepsilon^{\text{mech}}$ and therefore, stress σ^{mech} occurs and the bias point is shifted from the original linear region towards higher strain values. The maximum stress level for a linear strain-stress relation is marked by point 3 in Figure 2.4 for Fe and Al. For Fe point 2 and 3 coincide.A quite a good linear approximation for the stress characteristics can be applied as long as the stress level is below the “yield strength” (marked

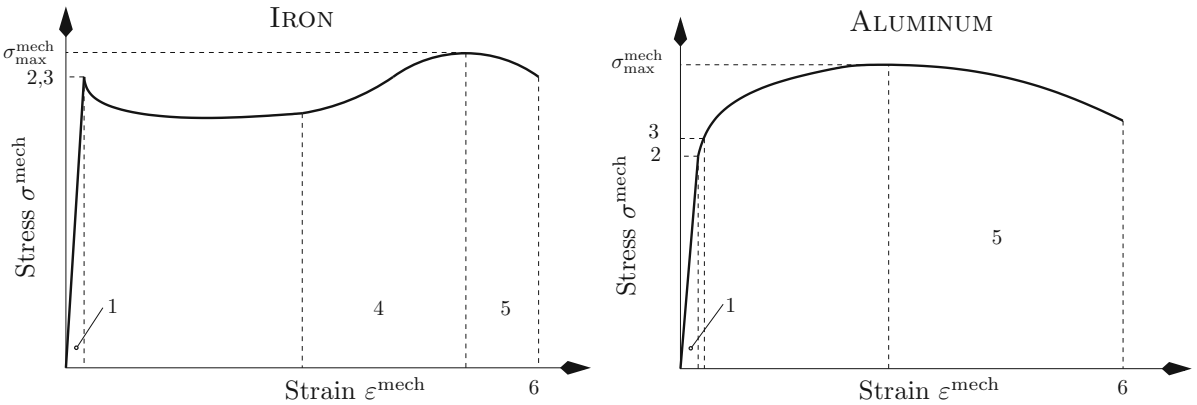


Figure 2.4: Mechanical stress as a function of mechanical strain, showing different regimes of behavior for the ductile materials iron (Fe) and aluminum (Al). The Following distinctive points can be determined for different materials: σ_{max} ultimate strength, 1. linear regime, 2. yield strength 3. proportional limit for stress 4. strain hardening region 5. necking region 6. rupture point

as point 3). Above this yield strength point, the stress characteristics show a highly non-linear behavior. For most metals and some other materials the elastic limit and the yield strength are essentially the same [110, 111]. Hence, beyond this point plastic deformations occur due to reordering of the grains and the crystal structure can be observed. Further increase of strain pushes the stress towards its maximum at σ_{max}^{mech}. After the maximum stress level has been reached, further strain enhancement causes material necking until the material ruptures.

To see the impact of increasing temperature within mechanical simulations, both the electro-thermal and the thermo-mechanical equation system have to be solved for each time step. Because of different requirements for the simulation of electrical and mechanical problems, a combined simulation within a single simulator requires a lot of memory which often exceeds the physically available amount in high-end simulation nodes. Therefore, the simulations have been separated to a self-consistent simulation flow, as shown in Figure 2.5. This self-consistent approach for solving electro-thermal and thermo-mechanical problems provides as benefits that already existing simulation tools with well-established and calibrated simulation models can be used, and in addition to that these simulators are already optimized and well-tuned for their specific problems. As a drawback, since this approach is an iterative one, cutting down to sub-

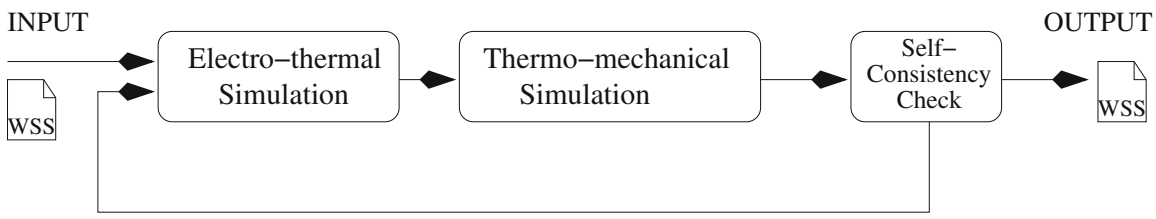


Figure 2.5: Schematic overview of the self-consistent electro-thermo-mechanical simulation tool flow.

problems results in longer execution time because for each iteration the complete loop has to be carried out until a stationary state has been reached.

To shorten the simulation time to the absolute necessary, an additional "self-consistency check" has been introduced which checks the update of the simulation result (cf. Figure 2.5). For very small updates of the temperature or the constraints due to mechanical changes, the simulations results will not be affected too much. Hence, the error if the consistency loop is not closed is negligible. However, as an additional safety measure, the self-consistency loop can be closed at certain user-defined check points.

2.3 Material Properties

The bulk properties of materials used in today's Si-based semiconductor devices like conductivities and permittivities are well known. Due to the steady enhancements over the last decades and because the material behavior has changed in terms of purity, doping, and microstructure, the previously well known values of the material properties have to be reconsidered to appropriately account the recently observed discrepancies. However, the values of the bulk properties are still a good approximation of the real values and are thus often used in material models. Once the materials become thinner than a certain threshold size the behavior changes and the material can no longer be considered a macroscopic object. In this case new models have to be applied, which differ from the well known standard bulk models. When the material dimensions are above the threshold value, the new models are required to converge to the well known bulk models.

For anisotropic materials the current density $\mathbf{J}$ is proportional to the electric field $\mathbf{E}$ by a tensor-valued quantity as

$$\mathbf{J} = \tilde{\sigma} \cdot \mathbf{E}, \tag{2.121}$$

where $\tilde{\sigma}$ is the electrical conductivity tensor. In most cases the material can be assumed to be isotropic and so the conductivity can be assumed scalar-valued. Hence, (2.121) reads

$$\mathbf{J} = \sigma \mathbf{E}. \tag{2.122}$$

This assumption is valid as long as the material has no priority directions as can be observed for instance in cubic-crystallized materials. The historical definition of σ (cf. [58, 112]) includes averaged values as

$$\sigma = \frac{n_\mathrm{e}\, q^2\, \lambda_\mathrm{MFP}}{2\, m_\mathrm{e}\, v_\mathrm{m}} \tag{2.123}$$

where the density of free electrons n_e, the length of the mean free paths of electrons λ_MFP, and the average mass of the electrons m_e are material-dependent. Furthermore, q denotes the elementary charge and v_m the thermal velocity of the electrons. Moreover, it is obvious that the average values of the length of the mean free paths, the thermal velocity, and the density of free electrons are temperature-dependent due to the thermal expansions of the material and the resulting mechanical stresses as well as thermally induced diffusion. Therefore, the commonly given values for the electrical conductivity are only valid within a certain temperature range.

When the layer thickness of the material is reduced to a thin film of only several atomic layers, the electrical conductivity becomes anisotropic. This also happens when grains determine the current flow inside polycrystalline materials. Due to considerable research efforts in material science and

in semiconductor process technology [31, 113–115] most of the materials can be appropriately deposited in such a way that the anisotropic part of the tensors becomes negligible for many applications [116].

To deepen the understanding of the material behavior and to improve reliability, accelerating tests [33,117–119] have been introduced, which stress the materials at high temperatures, causing aging effects within a very short time period. This procedure enables life-time tests at the very beginning of the life cycle and at very low costs costs compared to field experiments.

2.3.1 Interconnect Materials

The performance of digital semiconductor devices is determined by two issues: the switching behavior of the semiconductor itself and the signal delays due to the interconnects, which include metal lines, vias, the necessary protective coatings, barrier layers, and parasitic capacitances. Both parts are of the same order of magnitude whereby the contribution of the interconnect lines has gained more significance at every new technology node [3, 10].

Shrinking the technology node affects at first the transistor gate lengths and the area of the transistors, but also its surrounding elements such as contacts, vias, and interconnect structures. Since the dimensions of the interconnects have to be scaled down accordingly, the small additional protective coatings, barrier layers, and vias become more and more important [5, 120]. For instance, if a metal line with a rectangular-shaped cross-section of one square micrometer is considered, a 25 nm thick coating reduces the active metal cross-section by 5 %. If the interconnect line is scaled down to 90 nm, the coating thickness cannot be reduced that much. Hence a minimum thickness of approximately 5 to 10 nm still remains. This geometry would result in a reduction of the active cross-section by 11 to 21 %, respectively.

However, the cross-sections of the conductive materials are only one of several factors for the increasing resistivity. As Figure 2.6 shows, scaling down the dimensions for interconnects according to the ITRS [3,121] results in a considerable increase of the interconnect resistivity. The different contributions for lowering the conductivity as a function of the interconnect line dimensions are also shown. The figure shows the two main phenomena for the increase of the line resistance: The dominant factor is the reduction of the mean free path of the charge carriers, which can be observed by increased surface scattering due to line narrowing. A second important factor is the internal microstructure of the material, which can be described with a grain boundary model (cf. Section 2.3.2).

The following part of this section deals with the resistance increase due to the confinement of the trajectories of the electrons. The reason for the confinement of the trajectories of the electrons in metal lines is the internal microstructure (grains) and the fact that the macroscopic geometry of the metal lines comes into the range of the dimensions of the microstructure of the material. Both the grain boundaries and the surfaces of the solids are barriers for the charge carriers. Moreover, at barriers and surfaces the crystal structure is distorted and the impurity concentration is rather high compared to the bulk material. Therefore, the charge carriers scatter also at these impurity sites.

Since the deposition process cannot be guaranteed to result in a single crystalline material, it has to be considered polycrystalline. Due to the advances in process technology, especially for metal lines, the type and the shape of the distribution of grains can be controlled in wide ranges. However, the presence of grains cannot be avoided, but their shape can be adjusted appropriately, at least for certain metals. With this technology enhancement, many former serious problems

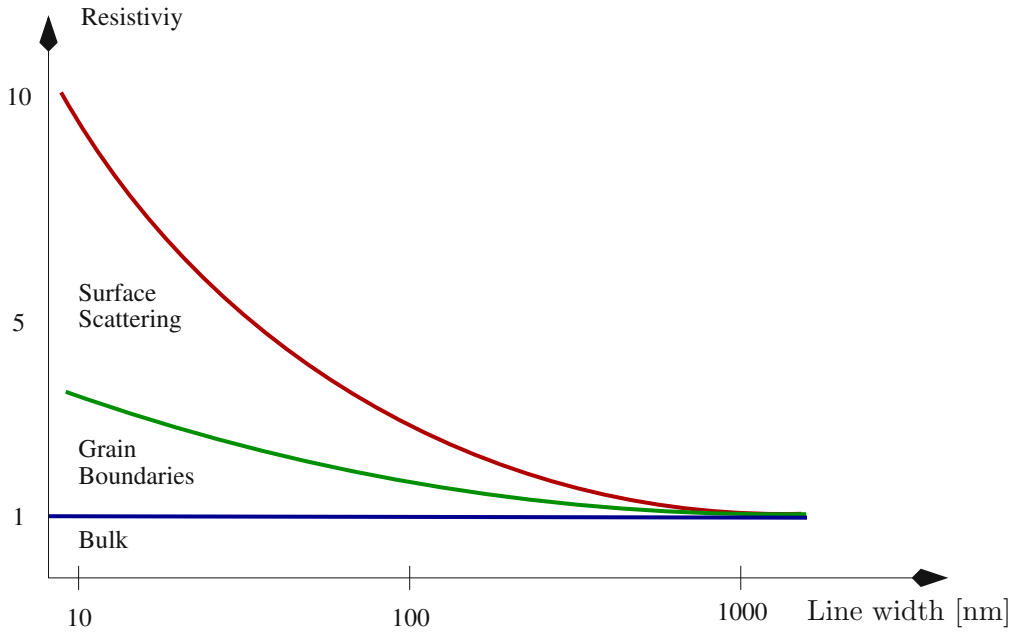


Figure 2.6: Relative resistivity increase for copper as a function of line width [3].

can be simply avoided or minimized. Hence, the impact of the polycrystalline behavior can often be neglected, for instance if the number of grain boundaries is small or when current flow is perpendicular to the average grain boundary surface. Therefore, metals can be mostly considered as single-crystalline solids which simplifies the numerical investigation of transient electro-thermal problems enormously.

The part of the resistivity that results from the grain structure can be reduced to a certain minimum. However, the geometry introduces absolute physical constraints according to the maximum resistance allowed and the expected current load for the interconnects. These requirements are given by the technology node and the design rules and cannot be changed. Thus, a resistance increase due to increased surface scattering events has to be taken into account from the beginning of the design stage. As Figure 2.6 shows, the bulk resistivity of some metals (e.g. copper) is only valid for thick metal lines. For thin wires, a significant resistance increase is observed, which has to be considered during the design phase of the integrated circuit.

To determine the difference between the resistance of a bulk material and the resistance of the same material as a thin film, the scattering effects have to be considered according to the narrowing effect.

For a one-dimensional resistivity increase due to confinement of charge carriers, for instance in a flat parallel-sided slab, Fuchs[33] [122] derived the expression

$$\frac{\rho_0}{\rho} = 1 - \frac{3}{2}\frac{1-p_{\mathrm{sc}}}{\kappa}\int_1^\infty \left(\frac{1}{x^3} - \frac{1}{x^5}\right)\frac{1-\exp\left(-\kappa x\right)}{1-p_{\mathrm{sc}}\exp\left(-\kappa x\right)}\,\mathrm{d}x, \tag{2.124}$$

where

$$\kappa = \frac{d_{\min}}{\lambda_{\mathrm{MFP}}} \tag{2.125}$$

[33] Klaus Fuchs (December 29, 1911 – January 28, 1988)

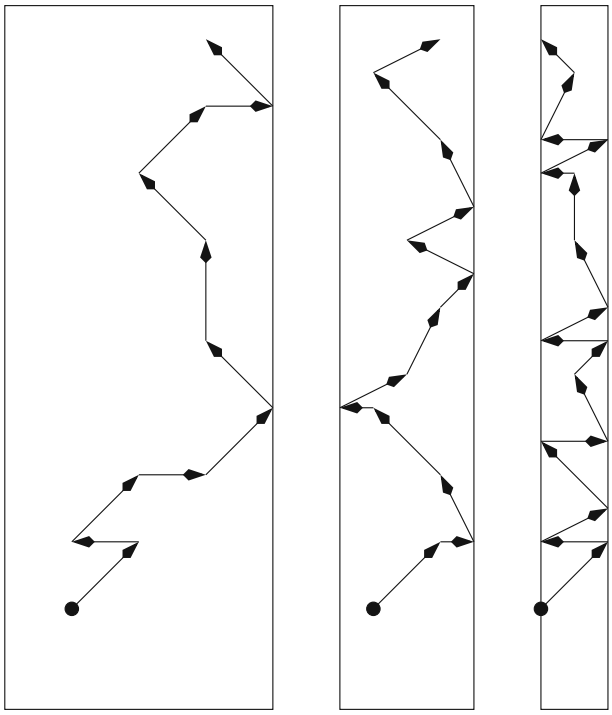


Figure 2.7: Resistance raise due to increased scattering events resulting from line narrowing of interconnect path dimensions.

represents the ratio between the smallest distance in the thin film d_{min} and the mean free path of the charge carriers λ_{MFP} and the fraction of charge carriers that elastically scatter at the material surface is denoted by $p_{\text{sc}} \in [0, 1]$. The mean free path of electrons in bulk Cu has been reported to be in the range of 39 nm [121,123,124] to 41.2 nm [5] in a temperature range of 300 K and 273.15 K. The mean free path for electrons in bulk Al is much shorter, for instance 14.8 nm at room temperature [124].

For thick material layers as well as for thin film layers (compared to the length of the mean free path), (2.124) can be simplified according to [125,126] as

$$\frac{\rho_0}{\rho} = 1 + \frac{3}{8}\,\frac{1-p_{\text{sc}}}{\kappa}\,, \qquad \kappa \gg 1, \tag{2.126}$$

$$\frac{\rho_0}{\rho} = \frac{4}{3}\,\frac{1-p_{\text{sc}}}{1+p_{\text{sc}}}\,\frac{1}{\kappa \ln\left(\frac{1}{\kappa}\right)}\,, \qquad \kappa \ll 1. \tag{2.127}$$

The approximation for $\kappa \gg 1$ is still used for interconnect materials because the technology nodes still operate at feature sizes with $\kappa \geq 1$.

However, to facilitate further down-scaling parasitic effects have to be reduced by optimizing the materials in terms of geometry and purification. For instance impurities aggregate at grain boundaries and material interfaces. Hence, device fabrication becomes more and more expensive due to the requirements of highly pure materials used in the device structures. Further enhancements of process technology nodes are very costly and require the consideration of new materials to improve the device characteristics as well as to reduce the fabrication costs. For instance, sili-

cided metals are a good alternative for gate contacts and for barrier materials to protect silicon. This type of material compound is very common in today's semiconductor devices.

Metals and Silicides

To improve the transistor switching characteristics high current densities have to be provided by interconnect lines. Because the dimensions of the interconnect lines cannot be increased as required, the performance becomes limited by signal delays due to the transistor transfer characteristics as well as by signal delays due to the parasitic resistances, capacitances, and inductances.

One of the requirements for shrinking a technology is that the voltage and the current remain at the same order of magnitude in order to provide backward compatibility to former circuit designs. But when the current remains constant at shrinking device feature sizes the current density increases quadratically with feature size reduction.

In order to provide good conduction properties for good device characteristics, a low sheet resistivity of the interconnect lines is not sufficient. Moreover, a low contact resistivity to other parts of the device, e.g. to bonding wires, vias, and semiconductor regions, is required. The phase state of conducting materials has to be mechanically stable over a wide temperature range, which also implies that the diffusion of ions into adjacent material regions has to be negligible. This requirement is very important for narrow interconnect lines, where the diffusion of some metals provide additional trap sites in the dielectrics and additional dopants in semiconductor materials, which would result in a long-term change in the device characteristics. One of these effects is called "contamination". It reduces reliability and lowers the quality of the device characteristics. The second effect is called "poisoning", which not only reduces but can destroy the device characteristics immediately due to changes in the doping profile. However, both effects destroy the desired device characteristics if a longer period of time is considered.

To provide appropriate barrier and protective layers the compatibility of the materials has to be clarified according to the requirements of the device structures to avoid performance reductions due to mechanical and electrical material constraints. This can be done for instance by matching volume expansion coefficients. Furthermore, another serious concern related to the introduction of new materials into an existing process is the cost of ownership (CoO). Since certain materials require special treatments in terms of safety for the device structures, the fabrication equipments and the environmental laws, additional costs have to be considered as well in advance. However, if there are more benefits than costs, the materials are introduced to the semiconductor device structure if the following requirements can be fulfilled [127]:

- The new materials have to avoid unintentional reactions with other materials, for instance with dielectrics or semiconductors in order to provide a stable material stack during device operation.
- High process yield is required to minimize the material-induced and process-related fabrication failures.
- Grain engineering allows the designers to modify and adjust the statistical distribution of the grain shapes and structures.
- It is important that the material can be easily patterned, which offers cheap and fast technology processes to deposit and etch the material.

- The new materials have to be stable in oxidizing chemical environments and oxidizable at certain defined ambient conditions to provide well-controlled material patterning processes.
- A good adherence is mandatory to avoid unintentional material dissolutions.
- Low residual stress levels from fabrication processes are desirable to allow mechanical relaxation of surrounding materials for a mechanically stable material stack.
- Smooth surfaces and interfaces to other materials are required to obtain well-defined interface conditions and low impurity concentrations at material interfaces for increased reliability.
- A rather long life-time of the material stack is required to meet the reliability requirements given by the device specifications.

For silicides (salicides) additional requirements apply [127]:

- The material has to remain stable throughout processing to avoid diffusion into other materials as well as to preserve the stoichiometric material compositions.
- Minimal reactions with metalization layers, e.g. Al, W, Cu, Ta, and Ti to provide stable material stacks and to offer stable diffusion barrier layers for certain metals is necessary.
- Minimal junction penetration to offer good contacts due to the Si consumption. A low Si consumption is required to preserve a proper operation in Si region and in the semiconductor junctions.

Al has served very well for several decades as an interconnecting material and for bonding pads. Al shows a high conductivity but has the disadvantage that it generates a native oxide (Al_2O_3), which is very stable, similar to SiO_2. However, Al_2O_3 is thermally and chemically much more resistant than SiO_2. Therefore, preventing oxidation of Al is very important during the fabrication process. Yet, even though the process is very well controlled, the higher demands related to higher current densities, brought up the problem of electro-migration, which has forced several companies to changes the interconnect base material to Cu. Bulk Cu has a much lower tendency of electro-migration than Al. On the other hand, the Cu integration into the technology nodes requires more effort due to the higher diffusivities and solubility of Cu into the standard materials used in the interconnect stack. To prevent the diffusion of Cu into Si, SiO_2, and other materials, additional coating and barrier layers have to be introduced. However, the introduction of such barrier layers posed new challenging effects that had to be overcome. The adhesion of copper on typical barrier layers is very weak, which results in high-diffusive paths at material interfaces with Cu. Hence, these weak interfaces reduce the activation energy for ion diffusion significantly, increasing the electro-migration effects and thereby decreasing the reliability of Cu-based technology. Due to new materials for barrier layers at material interfaces to Cu, the reliability of Cu interconnects can be better controlled within a certain range to meet the circuit design requirements.

Therefore, the interconnects have protective layers around the Cu. In addition to this measure, the lowest layer contacting the Si surface is still made of W, as known from the Al technology. The reason for that - once again - is reliability. If Cu diffused into the dielectric in a higher interconnect stack level, the Cu could contaminate and thereby reduce the dielectric reliability in terms of resistivity, break through voltage, and other parasitics: Nevertheless, the device structure would still be functional. If, e.g., Cu atoms diffused into the Si regions, a significant

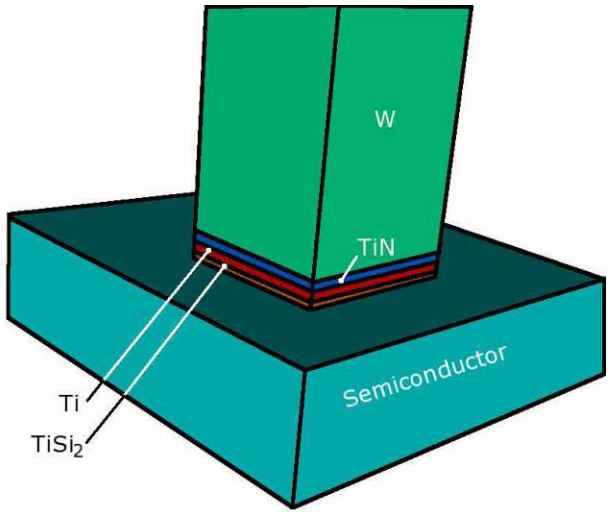


Figure 2.8: Tungsten via containing several materials as diffusion barriers.

shift of the threshold voltage is observed. As a result, the transistor would no longer be working properly resulting in complete device failure. Hence, to ensure that Cu atoms cannot contaminate the Si regions, the lowest via level is made of a less contaminating material than Cu. Figure 2.8 shows a typical via made of W and Ti compounds as used in standard Al technology nodes, and as first-level vias in Cu technology nodes. For this structure it is important to note that the contact to the semiconductor consists of $TiSi_2$. On top of the $TiSi_2$ contact layer is a Ti layer, which comes from the fabrication process in which Ti is deposited on top of the Si substrate. The $TiSi_2$ layer is formed either by diffusion,

$$2\,Si + Ti \longrightarrow TiSi_2, \tag{2.128}$$

or by explicit $TiSi_2$ deposition, following [127]

$$TiCl_4 + 2\,SiH_4 + Si \longrightarrow TiSi_2 + SiClH_3 + 3\,HCl + H_2, \tag{2.129}$$

$$TiCl_4 + H_2 \longrightarrow TiCl_2 + 2\,HCl, \tag{2.130}$$

where (2.129) shows a reaction in which also a consumption of solid Si occurs. By applying the appropriate temperature and optimal HCl concentration, the Si consumption can be avoided according to [127] by

$$TiCl_4 + 3\,SiH_4 \longrightarrow TiSi_2 + SiClH_3 + 3\,HCl + 3\,H_2. \tag{2.131}$$

However, this reaction is only stable for a certain chemical environment and certain temperatures. The chemical reaction can be further controlled by adjustments of the concentrations of silane (SiH_4) and hydrogen:

$$TiCl_4 + 2\,SiH_4 \longrightarrow TiSi_2 + 4\,HCl + 2\,H_2 \tag{2.132}$$

$$Si + 2\,HCl \longrightarrow SiCl_2 + H_2. \tag{2.133}$$

In this reaction, the Si consumption underneath the $TiSi_2$ layer can be controlled by the concentration of silane, which is quite stable at a typical processing temperature of 800 °C and a pressure of approximately 250 Pa (2 Torr).

On top of the $TiSi_2$ layer, a thin Ti layer is deposited that is covered by a TiN film. The TiN serves as a diffusion barrier for the W via. In optical stacks, TiW can be used instead of TiN to reduce reflections of light at interfaces.

For the Cu technology nodes, requires additional protection is required, since the size of the Cu atoms is smaller than that of W. Furthermore, Cu shows a higher tendency to diffuse into the dielectrics. The first attempt was the application of thicker layers of TiN, but at the early stages of Cu technologies, the adhesion of Cu metal to TiN was much weaker than expected [128]. With down-scaling and increasing current densities, the weak material interface caused too many failures, and so alternative material compounds were investigated. As a logical consequence, TiN was replaced by other nitrates. The most suitable of them (TaN) improved interface adhesion, but was not quite satisfactory. Hence, an additional Ta layer has been introduced to the Cu technology node, providing a sealing film over the Cu interconnect structures. Between Cu and TaN, an alloy of (TaCu) [33] — due to the high-temperature phase during the fabrication, in which both materials diffuse into each other — builds up. Recent arrangements have been proposed where the sealing TaN layer is replaced by a Ta layer only. Compared to the high resistivity of TiN, the Ta layer provides the advantage of much higher conductivity. Hence, the Ta layer offers an additional conductive path, which becomes importance for further down-scaling in interconnect structures. The lower limit of the thickness of the Ta layer, however, is approached. Hence, the resistance increase in smaller interconnects is still considerable, calling for research into alternative materials suitable for the future technology nodes.

A new approach has been reported which uses Mn as barrier layer for Cu interconnects [129,130]. In this case, as interconnect material CuMn alloy is deposited. With increasing temperatures, the Cu grains reach their final size and shape, and the Mn impurities accrete at material interfaces, but mostly at the interconnect surface. Due to diffusion and accretion processes of Mn impurities, the resulting thickness of the Mn barrier layer is in the range of a few atomic layers (12 to 40 Å). Hence, this procedure offers a good alternative to the previously used CuTa solutions, where the thickness of the thin Ta layer is limited to 25 to 50 Å with relatively a high standard variations, compared to the Ta layer thickness.

The Al and the Cu technology use Si_3N_4 as etch-stop layer, where the mechanical properties, especially the hardness of Si_3N_4, forced the technology developers to find better materials. In particular, a material was sought more suitable for chemical-mechanical polishing (CMP) than SiO_2. Because of the extreme hardness of SiC, this material is well fitting to these needs but the manufacturing process is very difficult. On the other hand, as experiments have shown, SiO_2 is still a good alternative, and now used as top layer in the interlayer dielectrics (ILD) stack even though it originally had been substituted by other alternative materials. The use of these new materials introduced too many problems and challenges during the CMP process so that a return to the original material was indicated. With SiO_2 as top ILD layer, the original technique for CMP can be applied and the well known effects such as dishing can be considered by the same design rules as before.

The requirements for the use of metals in interconnect structures can be summarized as follows:

- A high electrical conductivity is mandatory to minimize the power loss due to high electrical loads carried by the interconnect lines, especially for thin shallow structures such as vias in the lower metalization levels.
- A high thermal conductivity is desired to provide heat conduction paths with low heating due to the heat transport and to control the heat flow through the whole semiconductor device from the heat sources to the heat sinks.

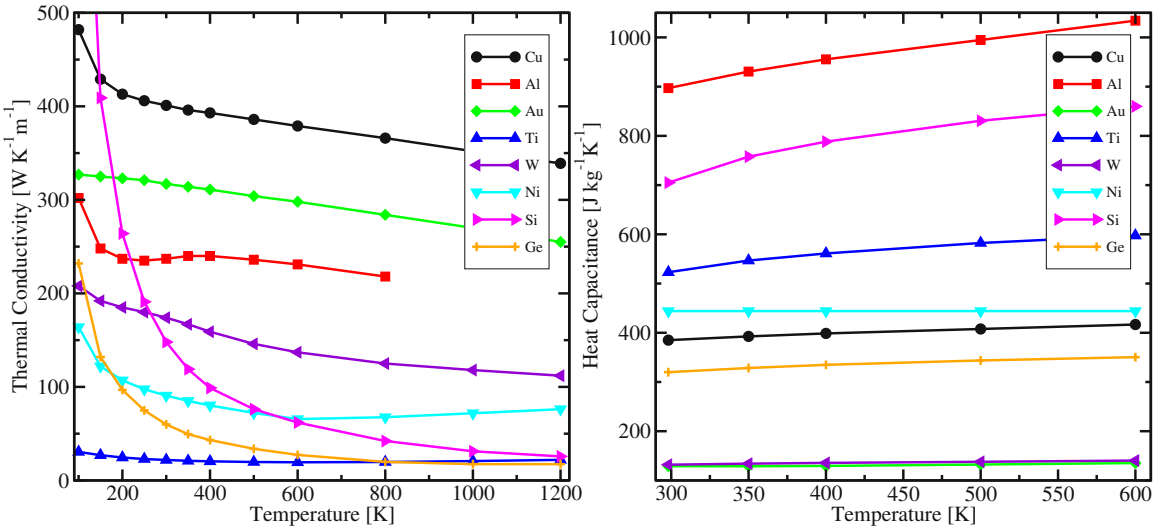


Figure 2.9: Thermal conductivity and specific heat capacitance of various common interconnect materials compared to Si and Ge. [131–138]

- The microstructure of the metal should be controllable within a certain range to reduce the scattering effects in order to facilitate a more precise estimate of the thermal impact of this effect, which then can be considered in the design process.
- Thermal stability over a wide range is also a very important feature that offers on the one hand a stable material during fabrication, and a reduced diffusivity into other materials, and on the other hand a mechanical rigid solid. However, this property is not available from just any material. For instance, B and Cu atoms show high diffusivities in many materials, especially into dielectrics and semiconductors.

Table 2.1: Characteristic electrical parameters of typical materials used interconnect structures and contacts.

Material $T = 300\,K$	σ_0 $[1/\mu\Omega cm]$	α_σ $[10^{-3}\,K^{-1}]$	References
Ag	0.613-0.629	4.1	[139, 140]
Cu	0.588-0.645	3.65-6.8	[4, 33, 139–142]
Au	0.425-0.49	4.0	[139, 140, 142]
Al	0.33-0.4	4.3-5.0	[39, 140, 142]
W	0.2	-	[142]
Ta	0.057-0.072	-	[33, 142]
Ti	0.012	-	[142]
$TiSi_2$	0.0625-0.0769	-	[127, 143, 144]
WSi_x	0.0125-0.0333	8.9	[47, 127, 143]
$TaSi_2$	0.0182-0.0285	-	[127, 143]
TiN	0.0167-0.025	4.1	[39, 127]

Tab. 2.1 presents parameters for the electrical conductivity for materials which are commonly used in interconnect structures. Compared to Tab. 2.1, Figure 2.9 shows a comparison of thermal conductivities of common interconnect materials, set into relation with the semiconductor materials Si and Ge. As can be directly seen from Figure 2.9, the thermal conductivity of metals depend only slightly on temperature. With rising temperature the semiconductors Si and Ge show a marked decrease of thermal conductivity. This effect can be explained by the main heat transport mechanism in metals and metal-like materials such as semiconductors. At low temperatures, the electron gas in metals has a certain average energy and — with a certain bias applied — the electron gas transports the information such as electrical current or heat with very little loss to the opposite side. Hence, the electrical and the thermal conductivity can be modeled by the WIEDEMANN[34]-FRANZ[35]-LORENZ[36] law, where the ratio between the electrical and the thermal conductivity is proportional to the absolute temperature. Conducting materials with a perfect lattice have no resistance, hence infinite electrical conductivity [142]. According to the WIEDEMANN-FRANZ-LORENZ law, also the thermal conductivity would be infinite. However, the crystals are not perfectly periodic and the crystal planes are also not perfectly aligned. Therefore, the electrons are scattered and, as temperature increases, also scattering increases due to several additional effects.

The according specific heat capacitances are presented in Figure 2.9, which shows that the tendency of the semiconductors and the different metals from Figure 2.9 is quite the similar.

However, a rigorous electro-thermal analysis, especially in the high frequency domain, has to include electro-magnetic effects like the skin effect, which reduces the conductivity due to a limited electro-magnetic field penetration into the metal [145–147]. The penetration depth d is expressed as

$$d = \sqrt{\frac{2}{\omega\sigma\mu}}, \tag{2.134}$$

where ω is the angular frequency, σ the electrical conductivity, and μ the magnetical permeability. As a consequence the resulting local current density $J(\mathbf{x})$ can be approximated with the analytical solution for a cylindrical solid [59]

$$J(\mathbf{x}) = J_0 \exp\left(-\frac{|\mathbf{x} - \mathbf{x}_{\text{Surface}}|}{d}\right), \tag{2.135}$$

where J_0 is the DC current density and $\mathbf{x}_{\text{Surface}}$ is the closest point of the surface of the conductor. The temperature dependence in (2.134) and (2.135) is implicitly present and determined by the well known temperature dependencies of the materials parameters used.

Semiconductors as Conducting Materials

Determining the optimum layout of interconnect lines is a quite complicated task since the thermal properties of the underlying materials and devices have to be considered. Therefore, and because the deposition and pattering processes for metals operate at elevated temperatures, the thermal budget is the fundamental constraint. If, for instance, the thermal budget is exceeded, the temperature profile causes thermally induced diffusion processes which alter the underlying device structures. For instance the doping profile can change its shape or certain materials

[34] Gustav Heinrich Wiedemann (October 2, 1826 – March 23, 1899)
[35] Rudolph Franz (December 16, 1826 – December 31, 1902)
[36] Ludwig Valentine Lorenz (1829 – 1891)

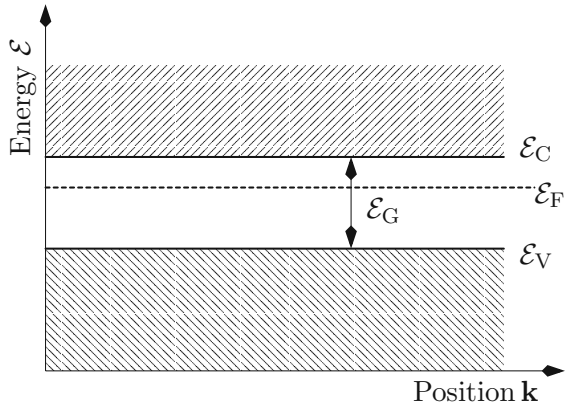


Figure 2.10: Idealized band diagram which shows a semiconducting material with the different energy levels resulting from the band structure. The FERMI level $\mathcal{E}_F$ which determines the energy level where occupation probability is one half. If $\mathcal{E}_F$ is closer to the conduction band than to the valence band, the material behaves more like a semiconducting material than an insulator.

can diffuse into other materials at contacts or at protective interface layers. To overcome these types of problems associated with metals, semiconductors can be used instead as contacts and interfacing materials between the lowest level where the semiconductors device structures are located and the first metalization level. The use of semiconducting materials at the interfacial layer includes for instance contacts to semiconductors regions via polySi or silicides (Co, Ni, and Ti silicides) for gate contacts or interconnect lines made of polySi. This is possible since the properties of semiconducting materials can be adjusted within a wide range.

Semiconducting materials are nearly ideal insulators at very low absolute temperatures and show a drastically decreased resistivity due to thermal activation of electrons of the semiconductor material itself and of impurities [65,66,136–138] at moderate temperatures (room temperature 300 K. Hence, these materials offer also an interesting alternative for conductors for certain applications where the current load is not too high. To further increase the conductivity, impurities can be implanted into the crystal lattice to shift the FERMI level according to the demands. Since the conductivity can be adjusted over a wide range, conductors and resistance elements can be realized.

Dielectrics

Non-conducting, insulating materials are called dielectrics and provide a band gap $\mathcal{E}_G$ which is typically larger than 4 eV. Figure 2.10 shows a typical band edge diagram for a semiconducting material. Here, the energy levels $\mathcal{E}_C$, $\mathcal{E}_V$, and $\mathcal{E}_F$ are the conduction band energy, the valance band energy, and the FERMI level, respectively. The FERMI level is defined as the energy for which the occupation probability is exactly 1/2. The activation of an electron from the valence band to the conduction band requires the energy of the band gap. This type of energy can be either of potential, kinetic, or thermodynamic nature.

Natural quantities to describe dielectrics are the conductivity σ, the band gab $\mathcal{E}_G$, the relative dielectric constant ε_r, the break-down voltage, and the melting, respectively the boiling point for fluid dielectrics. For modern semiconductor devices, the conductivity and the break-down

voltage have lost importance because the design can control the internal voltage distribution very well and the conductivity of the insulating material are considered together with their leakage behavior including tunneling effects in gate dielectrics.

For thicker dielectric layers, the capacitances are more important than the conductivities because the impact of cross talk between different interconnect lines is mainly determined by the capacitive coupling. This phenomenon can be investigated only if the final chip layout has been designed.

The relative dielectric constant ε_r is the most important quantity in microelectronics to characterize the insulation material since the capacitive coupling of two structures should be either ideally if the current flow has to be controlled, e.g. in a transistor or has to be avoided to reduced cross talk between interconnect structures. However, the constant has certain limits in both directions: If no matter is present the relative dielectric constant has its lower bound with $\varepsilon_r = 1$. The upper bound is given by the crystal structure of the material. Some materials, such as Perovskites and PZTs [140], provide extremely high relative dielectric constants. As a drawback, these materials often show a quite low stability in terms of the repolarisation and temperature. However, in certain applications these materials can be applied as well.

SiO_2 has been used in a wide range of applications because it is easy and cheap to produce and is rather stable in electrical and thermal terms, and very chemically resistant. Therefore, SiO_2 is very often used for instance as insulation material in control gates in transistors where a rather high dielectric constants would be required. But due to the low costs and simpleness of the fabrication of SiO_2, this material is still used as gate dielectrics and as passivation and insulation layers in interconnect structures to encapsulate the interconnecting lines from each other. In the latter example, a very low dielectric constant is the optimum for the overall device performance.

There are many materials which provide better electrical behavior than SiO_2, but none of them can be as reliably produced within existing economical limits as SiO_2. Therefore, the Si technology is very commonly used and has generated a considerably big market for such fabrication machinery, which even further reduces its CoO. If new material compounds have to be considered in terms of fabrication, additional materials have to be acquired, which are mostly very rare in high purity. Furthermore, the deposition and etching of such materials often requires new chemical environments and new machinery to handle these chemical reactions.

Despite of the huge costs, the enhanced electrical requirements given by the semiconductor road map demand the introduction of new materials which have either lower or higher relative dielectric constants ε_r to fulfill the industry's needs for future down scaling.

Sofar, enormous efforts have been made to supply the newly developed technology nodes with novel materials that require only minor changes to the standard Si process flow. Material types, which have succeeded in reducing CoO are the low-κ and high-κ materials. The offer different ε_r values compared to SiO_2 and are used to adjust the capacitive coupling through material selection.

The dielectrics can be grouped according to their chemical structure in oxides, nitrides, carbides, halogenides, polymers, and organic materials. In addition, there are plenty of mixtures and doped material which provide advantageous material properties for certain purposes.

The group of oxides include the well known compounds SiO_2, Al_2O_3, and germanium oxide which can be either GeO or GeO_2 where germanium dioxide is thermally more stable. Other commonly used oxides for new semiconductor structures are BeO, ZrO_2, HfO_2, and Ta_2O_5 which are mainly used as high-κ materials within FEOL structures like gate stacks for transistors or capacitors for memory cells. Special types of oxides are the high-κ compounds Perovskites and lead zirconium

titanites (PZT). They provide high values of ε_r but have a very limited thermal budget because above the CURIE[37] temperature the spontaneous polarization vanishes according to a mechanical relaxation of the crystals.

The nitride group includes TiN, TaN, and Si_3N_4 which excels with their hardness. Unfortunately, the member materials are quite brittle compared to most of the oxides. Important advantages of nitrides are that nitrides can be built on top of a metal layer and that according to the stability of the nitride compound, the nitride layer can be used to seal certain regions for instance to avoid the diffusion of a particular metal to its surrounding semiconducting or insulating materials.

Carbides are another group of dielectrics where SiC is the most important representative. Because its advantageous crystal structure, this material can be used as a substrate material like Si, Ge, or Al_2O_3. However, SiC is very brittle, extremely hard, and chemically very robust. Hence, it is also used for BEOL structures for instance as etch stop layers in interconnect stacks.

Beside the already mentioned materials types, there are plenty of polymers and organic compounds which include polyimide, poly-tetra-fluorine-ethylene (PTFE), organosicate glasses, and other polymers. Those materials are often used as low-κ materials in BEOL structures as interlayer dielectrics (ILD) and some even as substitute for semiconducting materials.

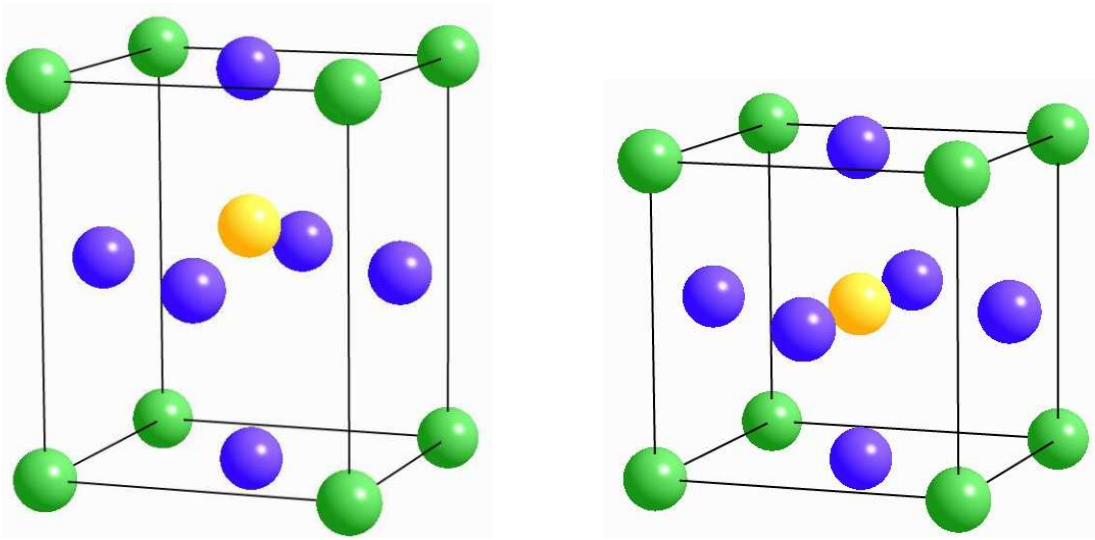
For BEOL structures dielectric layers are often doped to improve particular properties such as to harden the material compound, decrease the relative permittivity, or to reduce the diffusion constant for a certain atom species [148]. Typical representatives for this type are SiON, SiOC, SiOF, SiCN. They appear in the interconnect structures of leading edge high performance devices.

A critical issue in alternative materials is their temperature stability both during fabrication and during operation. For instance the phase stage of the Perovskites and PZT crystal structures that provides the high ε_r value is only thermally stable below the CURIE temperature T_c. For these materials the CURIE temperature determines the temperature limit for operation and the thermal budget during device fabrication. Figure 2.11 shows the principal assembly of a unit cell of a certain PZT material. In the mechanically relaxed stage, the crystal shows a face centered cubic structure where the Ti atom is exactly located in the center of the cubic unit cell. However, under certain conditions of pressure and temperature, the unit cell deforms in such a way that the Ti atom has too little space in the center of the unit cell and flips therefore either slightly to the upper or to the lower side. Hence, the space charge does no longer vanish but shows a spontaneous polarization. This stage provides a meta-stable energetic minimum of the crystal structure.

This is demonstrated in Figure 2.11 where an applied electric field in the vertical direction enables the Ti atom to slip from the upper side to the lower side of the center of the unit cell. This effect of flipping the Ti atom provides the high ε_r value. However, with every flip of the Ti atom, energy is absorbed by the crystal and causes hysteresis loss and with increasing number of flips the ε_r number will be slightly reduced due to mechanical relaxations. Nevertheless, the number of possible flips is enormous according to the current reliability concerns according to the ITRS. But if the temperature is increased above a certain threshold value (CURIE temperature), the thermal energy is sufficient for the advantageous crystal structure to mechanically relax. As a consequence, the high ε_r vanishes and drops back to approximately 1 in the global energetic minimum of this crystal structure.

Low-κ materials can be used to reduce capacitive coupling like cross talk or influence charges in adjacent interconnect lines. Materials with ε_r values lower than 2.5 are called extreme low-κ

[37] Pierre Curie (May 15, 1859 – April 19, 1906)



(a) $T < T_c$ (b) $T > T_c$

Figure 2.11: Temperature impact on the crystal structure of the high-κ material PZT. Below the CURIE temperature T_c, the crystal structure of PZT offers not enough space for a cubic centered Ti or Zr atom (a). Hence, spontaneous polarization is observed. However, above T_c the thermal energy is sufficient to enable a stable energetic minimum in which the cubic centered atom is at its forseen position in the center of the atomic unit cell.

(ELK) materials. Typical ELK materials reach values around 2.4 by using doped SiO_2. Examples are SiOC [148–151], SiON, or SiOF [152, 153]. Alternatively, values in the regime between 1.6 and 1.9 have been reported using air gaps [154–156]. Polymers like aromatic polymers [21] reach values of $\varepsilon_r = 2.7$. A typical range for organic silicate glasses OSG is $\varepsilon_r = 2.3 - 3.1$ [21], whereby the low values for the latter materials can be obtained if a porous low methyl variant of OSG is used [21].

As additional layers for etch stop and passivation purposes, the materials Si_3N_4 and SiCN can be used where carbon doped nitride offers a lower ε_r than the commonly used Si_3N_4 [151]. To account for the high mechanical stress in these material stacks used in BEOL structures, additional layers of Al_2O_3 can significantly reduce the mechanical stress but have higher ε_r values than SiO_2 [157, 158].

Barrier Layers

In addition to the already described materials which are used as dielectrics, semiconductors, and interconnects, there has to be another material type introduced which protects the other materials from metallurgical reactions and diffusion. This is required to prevent for instance poisoning of certain regions in semiconductor layers and consequently destruction of the required material properties [31]. The diffusion into dielectrics would cause additional traps in the surface region of the dielectrics and yield high coupling capacitances, higher leakage, and reduced break down voltages [148]. Materials which offer these required capabilities are called barrier materials. They can be either metals, dielectrics, or compound materials but they have to be chemical robust

and stable in terms of chemical and thermal burdens. Common materials are silicides for their compatibility to the Si regions in the lowest level to protect the Si from the materials which would change the doping profiles. Common materials to protect the dielectrics are nitrides, phosphides, and borides.

Barrier layers are also often used for contacts to provide a good conducting interface from the semiconductors to the via metal. As Figure 2.8 depicts, there are a couple of layers necessary to provide an appropriate transition from a semiconductor layer, which is mainly Si, to the via of the first metalization layer, which is mainly W. In between, there are a couple of barrier layers such as a silicide layer to contact the Si region and a Ti layer which comes from the deposition of $TiSi_2$. On top of those there is a TiN layer which actually protects the W layer from the via to diffuse towards the Si regions.

Similar protective coatings are necessary for the metal Cu. This is because the Cu atoms show a high diffusibility and a rather high solubility especially in Si and SiO_2. Industrial experience has shown that good protective properties can be observed with special alloys, nitrides, phosphides, and borides like TiN, $TiSi_2$, Si_3N_4, TaN, Mn-Si-O compounds, as well as CoWP [32, 158, 159], NiMoP [160], and NiMoB [159].

2.3.2 Polycrystalline Materials

Material properties of polycrystalline materials are often modeled in a similar fashion as single crystals thereby neglecting the internal microstructure. Unfortunately, most of the available deposition technologies only provide polycrystalline or amorphous rather than single crystal deposition. The only deposition process which provides single crystal growth in the processing reactor is the atomic layer deposition (ALD) which provides a very low growth rate and is therefore rather expensive with respect to time. Nevertheless, if crystalline structures have to be used, for instance in semiconducting materials, expensive single crystal deposition or growth methods have to be applied anyways. For most other applications, however, amorphous and polycrystalline regions are sufficient for device operation. For instance a crystalline structure of interconnect lines, contacts, and dielectrics is generally not required. However, the down-scaling of the semiconductor devices has also resulted in a certain downscaling of the interconnect structures. With lower line dimensions, however, additional effects which need to be negligible have now to be considered as well. For instance the grain boundaries of all polycrystalline materials contain unsaturated bounds, which can bind impurities. As a consequence energetic barriers form at these sites. Another crucial effect is the enhanced diffusion due to a lowering of the activation energy at these grain boundaries and material interfaces.

Typical properties of grain boundaries are that they have higher thermal conductivity than the surrounding material [161]. In addition, they can also have a higher melting point than the bulk material because there may exist some constellations (phase stages) where the grain boundary regions have higher thermal stability than the surrounding bulk-like material and can therefore act as a stabilizing element in the material structure [161]. If these stabilizing grain boundary elements are located in a material with a certain regular density, they mainly determine the behavior of high temperature regions. A typical energy diagram of such grain boundaries is depicted in Figure 2.12 where the common energy level has additional energetic barriers at the grain sites, which increases the specific number of scattering effects per time and therefore the resistivity of the material. At large dimensions of the interconnects lines with respect to the grain sizes, the shape and the distribution of the grains is not critical because the average over the hole structure is quite homogeneous. If very small interconnect dimensions have to be considered,

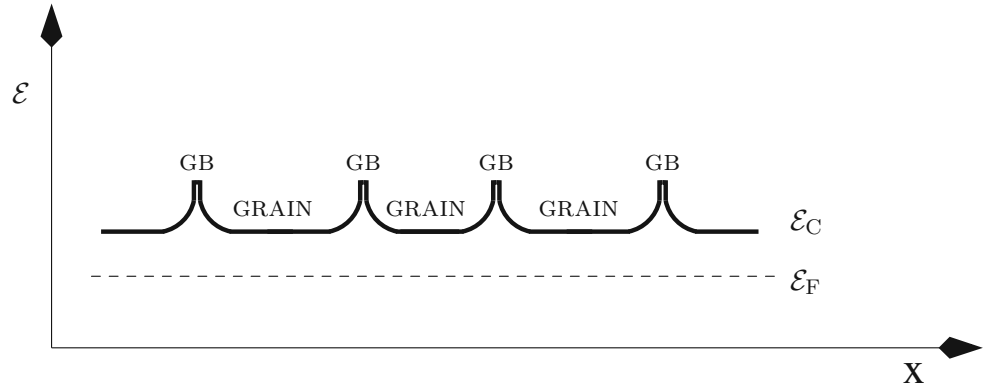


Figure 2.12: Energy band diagram of some grain boundaries showing energy barriers for charge carriers.

the single structure would consist of a few grains with different size and shape. Under these assumptions, the resistivity of this particular structure can be very different from the others. Hence, variable line resistances can result in different electrical burden and a different behavior of the whole circuit.

Figure 2.13 shows two pieces of polycrystalline Si with two different doping concentrations. However, the grain size shows in both subfigures nearly the same size distribution from a macroscopic point of view. From a microscopic point of view, for instance in a window 0.1×0.1 microns, the area would include either 1 or a few grains. If the resistivity would be considered of this fictive area, the difference would be a factor 2 or even more.

To describe materials from a macroscopic point of view, there are two main approaches. The first approach describes the conductivity or the resistivity of the materials, which requires the knowledge of the microstructure. If the material cannot be assumed to be an isotropic bulk material, the observed behavior has to be described with anisotropic material models. However, the microstructure is often not well known. Therefore, characteristic material layers were measured to provide characteristic material parameters such as the temperature coefficients for the

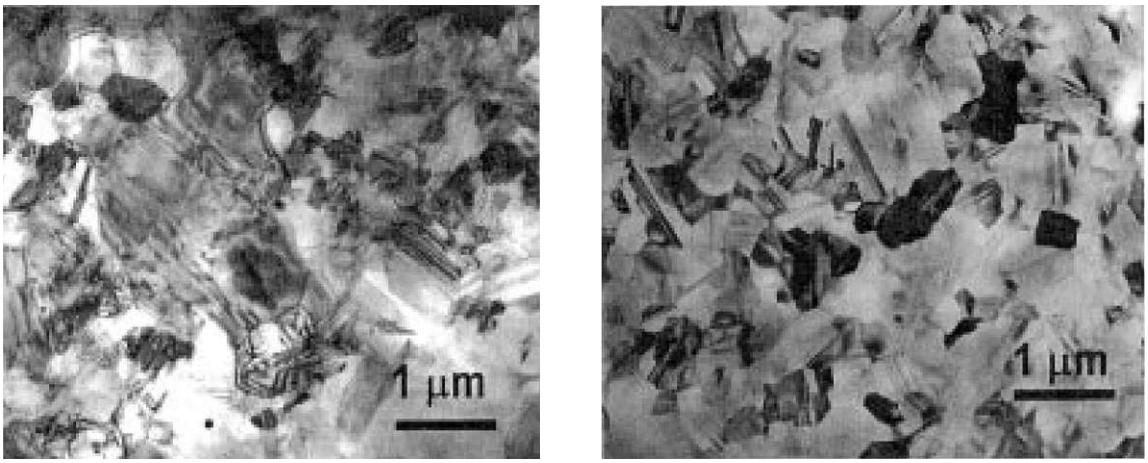


Figure 2.13: Grain structure of polycrystalline Si [162]

conductivities or the sheet resistance, which is mainly determined by the fabrication process. This provides an important advantage for the design in terms of computational effort for the evaluations of the models.

The sheet resistance $R_\square$ is defined as resistance for a specific layer thickness t_R as

$$R_\square = \frac{\varrho}{t_\mathrm{R}}, \tag{2.136}$$

where the specific resistivity is denoted by ϱ. Hence the overall resistance R can be calculated by

$$R = \varrho \frac{l}{A} = R_\square \frac{l}{w}, \tag{2.137}$$

where l is the length, w the width, and $A = t_\mathrm{R}\, w$ is the cross section of the layer. By using the sheet resistance $R_\square$ and the layer thickness for the appropriate technology, the model has only to account for the lateral dimensions such as length and width of the material chunk. With this information, the model uses information about the layout only, because the effects from the technology-related issues are determined by the sheet resistance.

A common method to account for difficult thermal effects for instance during self-heating is to apply a correction term to the well known parameters, which change with temperature, power density, or electric field. If for instance the temperature dependence is modeled, a polynomial approach is often used

$$\varrho = \frac{1}{\sigma} = \frac{1}{\sigma_0} \left(1 + \alpha_\sigma (T - T_0)\right), \tag{2.138}$$

where the material parameters are appropriately adjusted within a certain temperature range. Here, the electrical resistivity is modeled as a first order polynomial in temperature. The temperature coefficient α_σ is usually obtained by fitting (2.138) to measurement data. It has to be noted that the coefficients for these models are valid in a certain bias point only. If the bias point changes, for instance the temperature increases to a higher level, the first order approximation of the previous temperature would yield to a negative resistivity which is physically impossible. Another important side effect when using this type of model is that with increasing temperature the phase stage of the material might change and for instance the microstructure of the material changes. Effects such as movements of grain boundaries [116, 163, 164] or recrystallization [127] may occur. Hence, the values for the conductivities and their thermal coefficients changes as well.

Such polynomial models are found in literature for nearly every material which shows rather smooth behavior without any abrupt property transitions. However, if the impact of a particular observable is too drastic, a different model has to be developed and to be applied to describe the observed behavior more accurately.

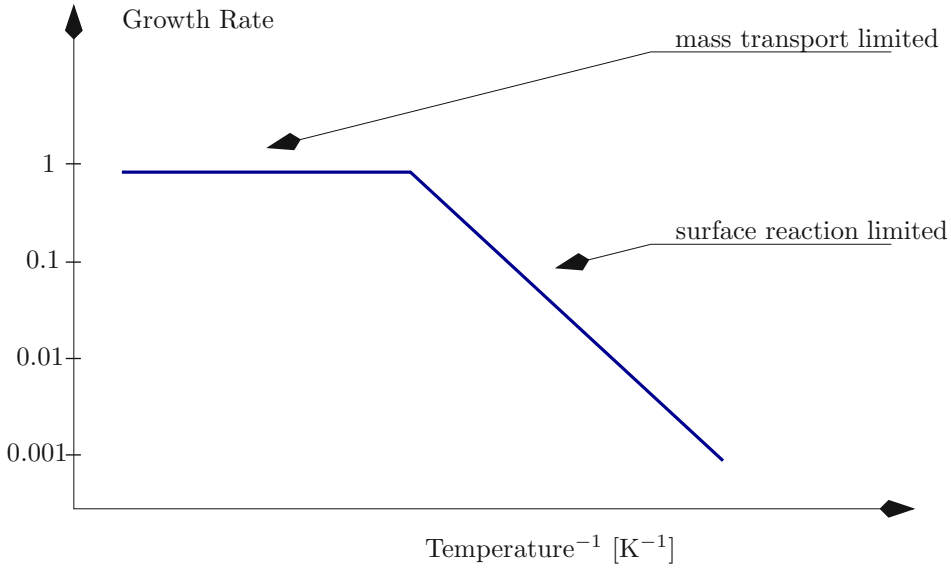


Figure 2.14: Deposition rate as a function of temperature.

2.4 Material Deposition

This section presents the most important processes that are necessary to deposit materials mainly used for interconnects. Starting with the principles of chemical vapor deposition CVD, this section gives an overview over the different possibilities for deposition of Si and Cu.

2.4.1 Chemical Vapor Deposition

One of the most effective methods to deposit material in microelectronics is to use a chemical reaction at the surface and of the underlying materials by reactant diffusion into the target material. This method is called chemical vapor deposition (CVD) and requires the underlying materials to act as catalytic materials. If a material does not, that material should at least do not react with the deposited material. However, some materials do not provide one of these mandatory requirements and therefore demand an additional barrier or seed layer, which provides the needed properties to both material layers.

The type of the deposition can be classified into three major categories: a mass transport limited regime, a surface reaction limited regime, and a reactant concentration limited regime. As Figure 2.14 and Figure 2.15 depict, the mass transport and the surface reaction are influenced by the temperature, where all three limitations depend on the geometry. These regimes have different constraining factors which reduce the maximum deposition rate and thus the growth rate of the material in the reactor [25, 127, 165].

- In the mass transport limited regime, the maximal support of masses is given by the chemical reactor and constrained by the external settings of the reactor; for instance the gas flux density is set by the engineers to control the growth rate in the reactor and the constellation between the reactor size, the set pressure, and the flux through the reactor determine the mass supported by the particular reactor system.

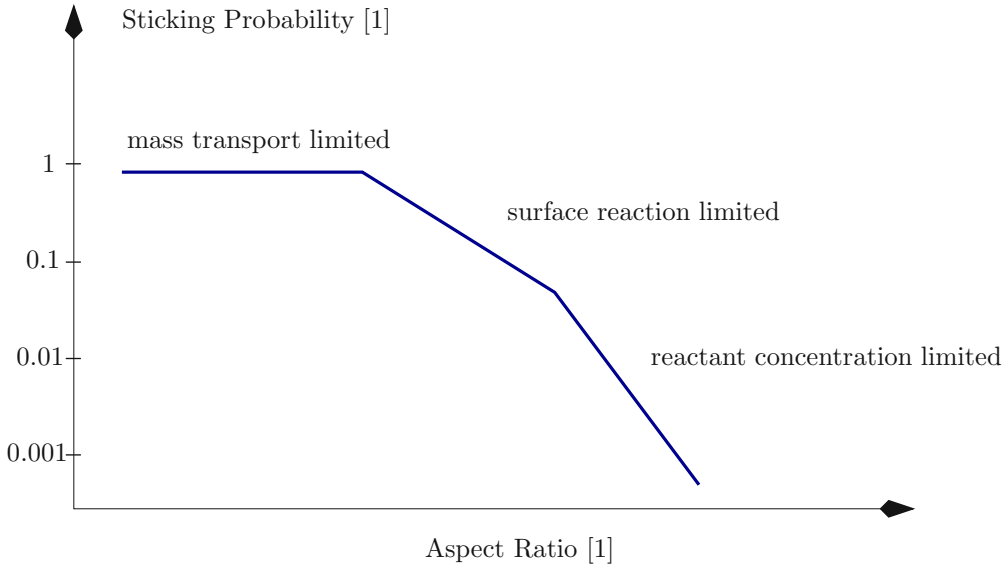


Figure 2.15: Sticking probability as a function of the aspect ratio.

- In the surface reaction limited regime enough reactants are available but the reaction at the surface (and sometimes in the bulk) is limited mostly by the temperature or the available catalytic particles, for instance elements at the surface area. Hence, the amount of reaction which can be performed is decreased either due to decreasing reaction coefficients [24] or by the surface area on which the reaction can take place [127].

- The reactant concentration limited regime has to be considered if the shape of the surface changes the gas fluxes, for instance inside deep trenches. In that case, the gas exchange can only be performed at regions close to the surface. Hence, in deeper regions of the trench a depletion of active reactants can be observed which yields exponentially decreasing reaction rates. However, the impact of this effect can be reduced if the deposition is performed at lower pressures (LPCVD). In that case, the by-products of the reaction in the trench have to be considered because they cannot be easily transported from the surface to the exit of the reactor. Therefore, they cause a considerable decrease of the growth rate due to collisions of the by-products with active reactants coming form the material source.

Once the materials are deposited, the possibilities to change the microstructure of the material are very limited because the internal material structure has mainly been defined by the deposition process. However, there are different methods, e.g. rapid thermal annealing (RTA) [127], dopant implantations [166,167], and mechanical and chemical methods like CMP (chemo-mechanical polishing) [168–170], which can slightly modify the microstructure. However, all these *a-posteriori* methods affect the regions at the surface or a limited region underneath the surface of the material only. For instance, if the Si surface is oxidized by O_2 or H_2O to obtain SiO_2, the corresponding chemical reactions use Si from the surface. Therefore, the thickness of the Si layer is reduced, which is often not desired for certain applications [25]. Hence, for this case, the material has to come from an external gas source to preserve the previous deposited layers. However, these chemical reactions follow mostly complex pyrolytically reactions and produce a lot of highly reactive byproducts [171].

Huge efforts have been made to describe the result of the material deposition in advance. However, due to the different and highly complex chemical reactions inside a reactor, the predictability is still limited. Several approaches have been proposed to deal with these problems. There are two main approaches for the simulation of material deposition. One approach is cell-based [167, 172] and has been introduced to describe etching of Si and the deposition of Tungsten and Silicon [172]. For the Si deposition, a CVD process of Silan has been considered in [172, 173]. Since the description for two and three-dimensional structures increases in complexity and memory consumption, a level-set approach has been proposed as a second approach [174, 175], which is presented in Chapter 5.

2.4.2 Deposition of SiO_2

The use of Si in microelectronic devices is very convenient because Si builds a native oxide on top of the Si surface following the oxidation reactions

$$\mathrm{Si} + \mathrm{O_2} \longrightarrow \mathrm{SiO_2}, \tag{2.139}$$

$$\mathrm{Si} + 2\,\mathrm{H_2O} \longrightarrow \mathrm{SiO_2} + 2\,\mathrm{H_2}, \tag{2.140}$$

where the SiO_2 is built by Si consumption from the surface. After the first oxide layer is built, oxygen and water have to diffuse to the Si-SiO_2 interface to grow new oxide. With this type of oxidation, very thin layers of SiO_2 can be deposited. However, thick oxide layers take a long time to build with this process technique and use a certain amount of Si as a source of the oxide layer.

Therefore, another technique has been introduced which transports all its sources via a carrier gas to the reactor where they react at the wafer surface, using the wafer surface as a catalytic material. A deposition process using silane SiH_4 has been commonly established, which follows the reaction equation

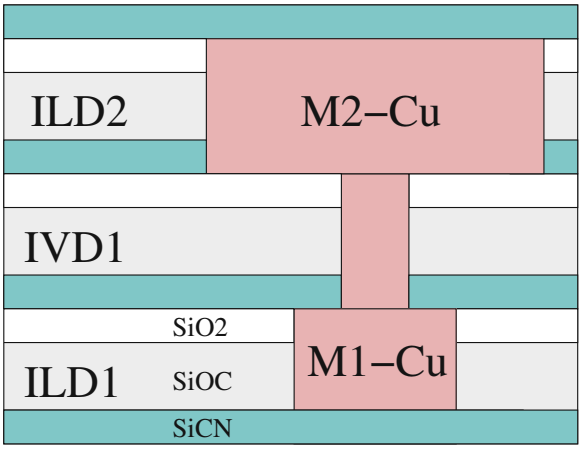
$$\mathrm{SiH_4} + \mathrm{O_2} \longrightarrow \mathrm{SiO_2} + 2\,\mathrm{H_2}, \tag{2.141}$$

where the SiH_4 reacts with oxygen at the hot wafer surface and SiO_2 is built together with H_2. A optimal temperature for this reaction process is in the region of 1300 K [25]. The by-products from (2.139)-(2.141) are able to diffuse through the oxide to the plain Si and react with Si as oxygen or as water according to (2.139) and (2.140). This diffusion of oxygen and water can be controlled by regulating the temperature of the reaction process (2.141). However, despite of the regulative measures, some of the Si is alway consumed. In addition, a considerable concentration of H_2 is built during the deposition of SiO_2 and has to be taken into account for reliability issues during the further processing and the device operations [25, 127].

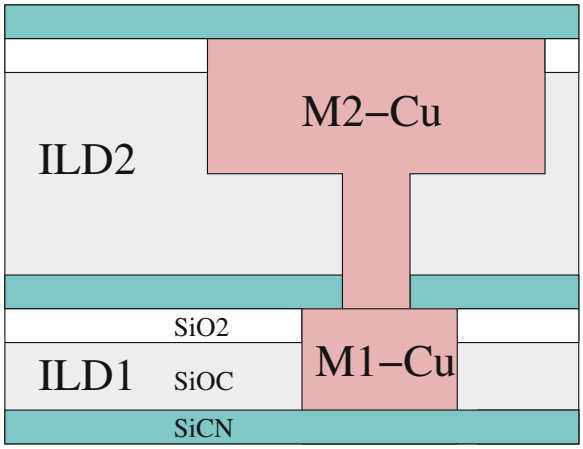
The previously presented methods have shown how a SiO_2 layer can be deposited by using Si from the target material (wafer). However, if Si consumption is not allowed at the surface, a more complex deposition method is required. A possible alternative, which provides that requirement is TEOS (Tetra-ethoxy-silane, $Si(C_2H_5O)_4$). The deposition of SiO_2 with TEOS uses a pyrolytic chemical reaction at a hot wafer surface in a LPCVD process (low pressure chemical vapor deposition) and follows the chemical reaction [25, 127]

$$\mathrm{Si(C_2H_5O)_4} \xrightarrow{1000\,\mathrm{K}} \mathrm{SiO_2} + 4\,\mathrm{C_2H_4} + 2\,\mathrm{H_2O}, \tag{2.142}$$

where the semiconductor device structures on the wafer are heated at a temperature of approximately 1000 K. The reactant TEOS is transported from a material reservoir to the reactor via a



(a) Single damascene structure.



(b) Dual damascene structure.

Figure 2.16: A cross-section through a modern multi-layered interconnect structure showing a single damascene structure (a) and a dual damascene structure (b).

carrier gas typically consisting of 92% N_2 and 8% H_2. At the hot surface of the wafer, the pyrolytic dissociation reaction (2.142) takes place. Typical SiO_2 growth rates at these environment conditions are 100 Å/min [25] up to 1000 Å/min [127]. The growth rate can be controlled within a certain range by varying the temperature, pressure, and the TEOS concentration in the carrier gas. Typically, the deposition reactions follow an ARRHENIUS[38] law [176]

$$R = A \exp\left(-\frac{\mathcal{E}_\mathrm{A}}{k_\mathrm{B}T}\right), \tag{2.143}$$

where the reaction rate R depends exponentially on the activation energy $\mathcal{E}_\mathrm{A}$. The proportionality constant A depends on the surface shape and the chemical reaction which takes place and must therefore be determined separately for each different deposition type. The proportionality constant A often depends on the temperature through a square-root law [127]

$$A = A_0 \sqrt{T}. \tag{2.144}$$

2.4.3 Deposition of Cu – Damascene Processes

Another important process technology is the electro plating for the Cu interconnect metal structures. In this technology, the structure is patterned by etching the shape of the structure in the underlying inter-layer dielectric (ILD) materials. After the patterning, a very thin barrier layer is deposited on top of the etched structure [32, 158–160] (e.g. Ta/TaN, TiN, CoWP, NiMoP, NiMoB). On top of that layer a seed layer is deposited which supports better adhesion of the Cu on the underlying material and acts as catalytic material during the plating process as well. Typical materials for these seed layers are compounds which include the material Pd or other compounds of polymers and organic materials.

The original deposition process (damascene process) was designed to process each layer on its own. Hence, the vias and the metalization levels had different process steps and demand a

[38] Svante August Arrhenius (February 19, 1859 – October 2, 1927)

sequence of cleaning, material deposition, CMP, and another cleaning step for each layer. A Cu technology using this sequence for its metalization levels as well as for its inter layer dielectrics (ILDs) and inter via dielectrics (IVDs) is called single damascene process. A simplified cut through a single damascene processed wafer is depicted in Figure 2.16a, where a typical structure, showing a level-oriented material compositions, is presented. Here, each level requires its own cap layer or etch stop layer, a separate ILD layer, and at the top there is a need for a material — for instance SiO_2 — that can be polished together with the interconnect metal copper.

Alternatively, a dual damascene technology combines certain similar process steps to one reducing the number of process steps and thus the time and costs required to build the back end of line (BEOL) stack. Hence the dual damascene process fabricates the IVD and the metalization layer at once. Obviously, as can be seen in Figure 2.16b, the metalization level has an increased thickness because the interconnect lines as well as the vias are included in a single metalization level.

Chapter 3

Thermal Models

This chapter discusses the significance of the temperature dependence in model for material parameters. However, it should be noted that the computational effort might increase significantly if the particular models discussed in this chapter are included in simulation tools. Therefore, the design engineer should be able to decide whether to give priority to the run-time performance or to the accuracy of the simulation.

3.1 Electrical Conductivity

The standard models for the electrical conductivity have been presented in Section 2.4. These models give very good results for bulk materials, where a single-crystal like behavior or an averaged material behavior can be assumed over the simulation domain. If thin film materials are considered, however, the previously presented description of bulk properties often lacks accuracy especially if the models should cover high-temperature effects. Therefore, additional material models have to be developed, which are able to describe the physics more fundamentally. Such models are intended to be implemented in simulation tools to improve the predictability of material models.

Applying single-crystal material models for bulk-sized materials chunks might can also suit well for semiconductor materials, if the internal microstructure can be neglected. However, if an additional impurity concentration can observed, a significant amount of extra energy barriers are built, as shown in Figure 3.1. Looking at the polycrystalline structure with a broader view the shape of the energy band looks more like Figure 2.12, where the typical ratio of the grain region and the grain boundaries is depicted. Here, the grains are separated by an energetic barrier consisting mostly of mismatched atoms that are located with a certain offset to the optimal lattice sites of an ideal crystal. Due to the diffusion and the segregation tendency of the impurities in the semiconductor materials, these impurity atoms move towards the interfaces of the grains and build structures as shown in Figure 3.1 at grain boundaries and at the surfaces.

To account for this type of behavior, the polycrystalline structure has been considered for contacts in transistors where additional barriers occur besides the work-function difference [177,178]. This model can be applied to other material parts where polycrystalline Si appears, for instance in contacts [177]. Consequently, a model has been developed by Mandurah and others, which accounts for these energetic barriers so as to derive a macroscopic model that can be calibrated if the necessary information is provided by measurements (cf. (3.3) and [177–181]). Based on this model Nathan and Baltes propose in [79] to use the assumptions of Mandurah from [179,180]

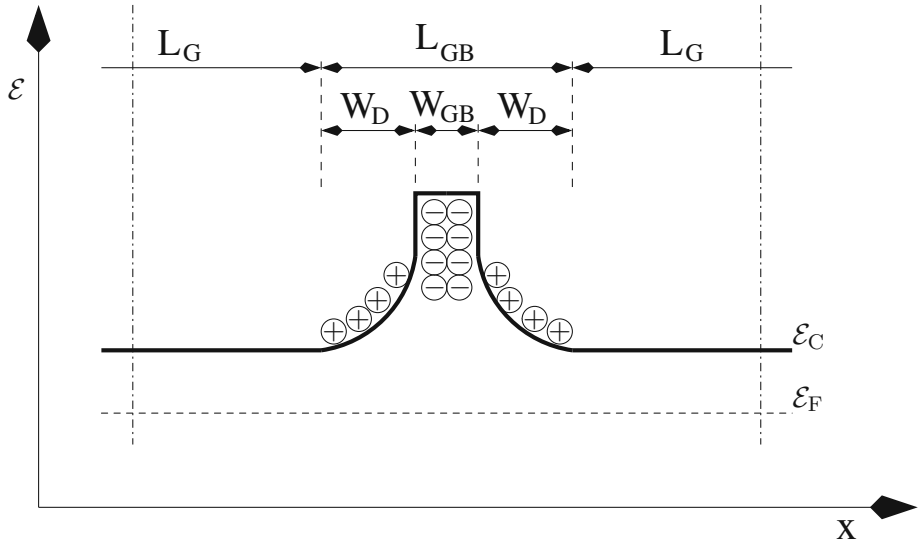


Figure 3.1: Energy band diagram of a single grain boundary showing trapped negative carriers at the boundary and the accumulated positive carriers towards the grain boundary zone.

and extend the grain boundary model by including an average-sized grain to obtain a more reasonable conductivity model.

Applications for such models include solar cells, polycrystalline diodes for temperature measurements, fuses, and sensors in general. The advantage of this model is to predict the behavior of the polycrystalline semiconductor materials in regions where single crystal growth cannot be provided. In this case, the electrical behavior is different from devices built from single-crystal materials. For instance, the electrical resistivity highly depends on the temperature and thus also on the reverse current of a diode made of polycrystalline Si. When comparing the standard model with measurements of polycrystalline materials, the original polygonal model with a reference value σ_0 and thermal correction term

$$\sigma = \frac{\sigma_0}{1 + \alpha_\sigma (T - T_0)}, \tag{3.1}$$

yields acceptable results for a very limited temperature range. Here, σ_0 denotes the electrical conductivity at the reference temperature T_0, α_σ is the material-dependent temperature coefficient of first order for the electrical conductivity, T is the absolute temperature. Depending on the shape of the real conductivity, the temperature coefficient α_σ has a certain range of validity. If wider ranges are considered, a second order temperature coefficient β_σ can be introduced by

$$\sigma = \frac{\sigma_0}{1 + \alpha_\sigma (T - T_0) + \beta_\sigma (T - T_0)^2}. \tag{3.2}$$

Regardless of the temperature bias point (reference temperature T_0) assumed, specific different values for σ_0, α_σ, and β_σ are required. They can be obtained by model calibrationa with measuremnts or data listed in literature.

Yet this expansion has its limits. For applications where this second-order temperature coefficient model is not sufficiently accurate, the microscopic structure of the material has to be included into the conductivity model. Especially if material compounds are considered, e.g. polycrystalline silicon with high doping for a better conduction, or silicided metals ($TiSi_2$, WSi_x), the thermal

impact on the conductivity can no longer be described by these polynomial functions. With inclusion of the microstructure into the electrical models, the accuracy of the models can be increased. As proposed in [79], a conductivity model which accounts for grain boundaries [177, 179, 180] is combined with the drift diffusion model for the plain grains.

The grain boundary model for the electrical conductivity σ_{GB} considers a ballistic transport over the barriers which can be described by the doping concentration N, the interface trap density N_{t}, and by the temperature T. The other parameters are more or less constants for a certain process technology. Following [179, 180], the model for the grain boundary conductivity reads

$$\sigma_{\mathrm{GB}} = \frac{q^2\, L_{\mathrm{GB}}\, N \exp\left(-\dfrac{\mathcal{E}_{\mathrm{GB}}}{k_{\mathrm{B}}T}\right)}{\sqrt{2\pi\, k_{\mathrm{B}}T\, m^*_{n,p}}} \left[\frac{\exp(-b_1)}{1 - c_1\, k_{\mathrm{B}}T}\, \frac{N_{\mathrm{t}} + c_1\, k_{\mathrm{B}}T\, N\, W_{\mathrm{GB}}}{N_{\mathrm{t}} + N\, W_{\mathrm{GB}}}\right], \tag{3.3}$$

where the coefficients b_1 and c_1 are defined as

$$b_1 = \frac{4\pi\, W_{\mathrm{GB}}}{h} \sqrt{2\, m^*_{n,p}(q\xi - \mathcal{E}_{\mathrm{GB}})}, \tag{3.4}$$

$$c_1 = \frac{4\pi\, W_{\mathrm{GB}}}{h} \sqrt{2\, m^*_{n,p} \frac{1}{(q\xi - \mathcal{E}_{\mathrm{GB}})}}. \tag{3.5}$$

The first term in (3.3) determines the base value of the conductivity due to the ballistic transport over the energy barrier $\mathcal{E}_{\mathrm{GB}}$. However, this base value is corrected by two additional factors, where the first correction factor is to due the temperature dependence of the width of the depletion zone close to the grain boundary and the second correction is due to the temperature dependence of the trap density in the interface region between the grain boundary and the grain. In these equations, the energy $q\xi$ is the grain boundary energy level defined relative to the conduction band [79]. The effective mass for electrons and holes are denoted as $m^*_{n,p}$, W_{GB} is the grain barrier width, L_{GB} the overall length of the grain boundary including the depletion zone, and the width of the depletion zone is defined as $W_{\mathrm{D}} = N_{\mathrm{t}}/2N$ according to [179].

To determine the global conductivity σ of a polycrystalline material, Nathan and Baltes [79] used the Matthiessen rule to combine the electrical conductivities of the grain region σ_{G} obtained from (2.14) and (2.15) and the grain boundary region σ_{GB} as

$$\frac{L}{\sigma} = \frac{L_{\mathrm{G}}}{\sigma_G} + \frac{L_{\mathrm{GB}}}{\sigma_{\mathrm{GB}}} \tag{3.6}$$

where L represents the overall length of a unit cell,

$$L = L_{\mathrm{G}} + L_{\mathrm{GB}}, \tag{3.7}$$

L_{G} denotes the average length of a grain and L_{GB} the average width of a grain boundary. However, to consider also the third dimension, (3.6) can be extended by using the same assumptions in three dimensions. Hence, a cube with an edge length of $L_{\mathrm{G}} + L_{\mathrm{GB}}$ is considered, which consists of a cubic-shaped grain with the edge length L_{G} and the remaining part represents the contribution of the barrier. Considering a cross sectional cut perpendicular to the current density, the sheet conductivity for electrons and holes $\sigma_{\square}$ can be expressed by

$$\sigma_{\square}\,(L_{\mathrm{G}} + L_{\mathrm{GB}})^2 = \sigma_{\mathrm{G}}\, L_{\mathrm{G}}^2 + \sigma_{\mathrm{GB}}\left[(L_{\mathrm{G}} + L_{\mathrm{GB}})^2 - L_{\mathrm{G}}^2\right]. \tag{3.8}$$

Thus, the overall conductivity σ with the corrected terms is

$$\frac{L}{\sigma} = \frac{L_{\mathrm{G}}}{\sigma_{\square}} + \frac{L_{\mathrm{GB}}}{\sigma_{\mathrm{GB}}}. \tag{3.9}$$

However, the use of this model has also its limitations. The model has shown to be rather sensitive to changes of the parameters. For instance the average grain size has a significant impact on the value of the conductivity of this model. Due to different process steps and fabrication formulas, the parameters for this model can change significantly, which requires a separate parameter extraction for each particular setting of the machines used in the fabrication of the microelectronic device. Nevertheless, the model provides a better insight in to the microstructure of polycrystalline materials using a meso scale model.

3.2 Electrical Permittivity

The electrical permittivity tensor $\tilde{\varepsilon}$ describes how the electric field $\mathbf{E}$ is related to the electric flux density $\mathbf{D}$,

$$\mathbf{D} = \tilde{\varepsilon} \cdot \mathbf{E}. \tag{3.10}$$

For isotropic materials, the permittivity tensor reduces to a scalar-valued quantity and the direction of the electric flux density and the electric field is the same. Hence, for isotropic materials the electrical permittivity can be approximated

$$\varepsilon \;=\; \varepsilon_0\, \varepsilon_\mathrm{r} \left(1 + \alpha_\varepsilon (T - T_0)\right), \tag{3.11}$$

where the temperature coefficient α_ε is determined by

$$\alpha_\varepsilon = \frac{1}{\varepsilon_\mathrm{r}} \frac{\mathrm{d}\varepsilon_\mathrm{r}}{\mathrm{d}T} \tag{3.12}$$

and can be derived from the Clausius[1]-Mosotti[2] equation [140]

$$\frac{N\,\alpha^\mathrm{pol}}{3\,\varepsilon_0} = \frac{\varepsilon_\mathrm{r} - 1}{\varepsilon_\mathrm{r} + 2}, \tag{3.13}$$

where N is the dipole density, α^pol the polarizability, and ε_r the relative permittivity. This equation describes the relative permittivity as an implicit function of the polarizability and the dipole density. The latter is an inherent property of the material and does normally not change as long as the phase stage is not altered. Hence, the derivative with respect to the temperature T is

$$\frac{\mathrm{d}\varepsilon_\mathrm{r}}{\mathrm{d}T} = \frac{(\varepsilon_\mathrm{r} - 1)(\varepsilon_\mathrm{r} + 2)}{3} \left(\frac{1}{N} \frac{\mathrm{d}N}{\mathrm{d}T} + \frac{1}{\alpha^\mathrm{pol}} \frac{\mathrm{d}\alpha^\mathrm{pol}}{\mathrm{d}T} \right). \tag{3.14}$$

Due to mass conservation, the thermal volume expansion can be assumed to be equal to the negative temperature coefficient of the dipole density because the number of atom inside a atomic unit cell remains constant. The only assumption here is that the material persists in the phase stage. Hence, α_ε is

$$\alpha_\varepsilon = \frac{(\varepsilon_\mathrm{r} - 1)(\varepsilon_\mathrm{r} + 2)}{3} \left(\frac{1}{\alpha^\mathrm{pol}} \frac{\mathrm{d}\alpha^\mathrm{pol}}{\mathrm{d}T} - \alpha^\mathrm{mech} \right). \tag{3.15}$$

According to the magnitude of the relative dielectric constant, materials can be divided into two groups: high-κ materials, which have a larger ε_r than SiO_2, and low-κ materials, which have a

[1] Rudolf Julius Emanuel Clausius (January 2, 1822 – August 24, 1888)
[2] Ottaviano Fabricio Mosotti (April 18, 1791 – March 20, 1863)

Table 3.1: Typical relative dielectric constants for various high-κ materials.

Material	ε_r [1]	$\mathcal{E}_G$ eV	References
SiO_2	3.9	9.0	[182–184]
	3.9	8.0 – 9.0	[25, 185, 186]
	3.9 – 4.6 (nitridation)	–	[127, 187, 188]
SiO_2(TEOS)	4.1	–	[21]
Si_3N_4	7.5	5.0	[184, 189]
	7.0 – 7.9	5.0 – 5.3	[79, 182, 183, 185]
	4.0 (Si-rich) – 8.0 (N-rich)	–	[25, 79]
SiCN	5.0	–	[20]
TiO_2	40.0	3.5	[184, 189]
	39.0 – 170.0	3.0 – 3.5	[182, 186, 190]
ZrO_2	23.0	5.8	[183]
	12.0 – 25.0	5.0 – 7.8	[182, 184, 185, 190]
HfO_2	25.0	5.7	[184, 185]
	16.0 – 40.0	4.5 – 6.0	[182, 183, 190]
Al_2O_3	9.0	8.7	[185]
	5.0 – 12.0	8.7 – 9.0	[25, 182, 190]
Y_2O_3	15.0	5.6	[183, 185]
	4.4 – 18.0	5.5 – 6.0	[186, 190]
Ta_2O_5	25.0	4.4	[184, 189]
	23.0 – 26.0	4.4 – 4.5	[182, 183, 185, 190]
$ZrSiO_4$	12.6	6.0	[190]
	3.8 – 12.6	4.5 – 6.0	[182, 183, 186]

lower ε_r number. Materials with higher relative permittivities than SiO_2 have a higher dielectric displacement field (electric flux density) as SiO_2 if the same electric field is applied. Hence, the capacitive coupling between the two opposite sides of the materials is much tighter as with SiO_2. Therefore, such materials are used to increase the capacitance where the thickness of the materials can be reduced to obtain the same capacitance as with SiO_2.

However, capacitive coupling is often considered as a parasitic effect, for instance if an array of interconnect lines is considered, where each of them is used to transmit a different signal. Capacitive coupling means in this context that an electric signal in one of these lines influences an electric signal in the other lines. This is refered to as cross talking. If bus structures for data or addresses are considered too much cross talk might result in logic device failures. Hence, for these devices a reduction of the coupling is required. Materials with lower ε_r values (low-κ materials) offer an alternative to reduce the capacitive coupling. A major drawback of all high-κ and low-κ materials is their mechanical weakness and the complicated and expensive fabrication. The high ε_r value of high-κ materials vanishes at a certain temperature, where a phase change takes place, reducing ε_r to 1. This temperature is called CURIE temperature in analogy to thermo-magnetic effects. Hence, the thermal budget for these materials is very limited. The low-κ materials are either porous and very hard or soft, where both properties are extrema for the fabrication. For instance, if a soft interlayer dielectric material is processed by a chemical mechanical polishing (CMP) process step in Cu technology, the abrasion of the new materials is considerable higher than

Table 3.2: Typical relative dielectric constants for various low-κ materials.

Material	ε_r [1]	References
SiO_2	3.9	[25, 182–186]
SiOC	2.4 – 3.4	[20, 148–150]
porous SiOC	2.5	[150]
SiOCH	2.7	[152]
organo-silcate glasses (OSGs)	2.3 – 3.1	[21]
aromatic polymers	2.7	[21]
with air gaps	1.6 – 2.5	[154–156]

that of Cu and SiO_2. The mechanical weakness of this soft material requires additional protective layers for a normalized CMP procedure. Because many of the low-κ materials are compounds of severals material, they also limit the thermal budget for the fabrication processes. Otherwise the amorphous and porous materials re-anneal and the advantageous properties vanish.

3.3 Thermal Conductivity

According to FOURIER's law (e.g. to Chapter 2) the heat flow between two regions having different temperatures is determined by the temperature gradient and a proportional factor λ. The value of the thermal conductivity for metals is related to the electrical conductivity because the electron gas in the metal transports the heat as well as the electrical current. WIEDEMANN and FRANZ found that the ratio between the electrical and thermal conductivity for a metal is proportional to the absolute temperature

$$\lambda = L\,\sigma\,T, \tag{3.16}$$

Later on, this proportionality factor L was identified by LORENZ as

$$L = \frac{\lambda}{\sigma\,T} = \frac{\pi^2}{3}\frac{k_B^2}{q^2} = 2.44\times10^{-8}\,\frac{V^2}{K^2}. \tag{3.17}$$

Equation (3.16) is referred to as the WIEDEMANN-FRANZ-LORENZ law and considers only the contribution of the heat transport capabilities of the electrons. Due to the different material properties, each material has its own LORENZ number, which differs from the theoretical value and generally depends on the temperature. Nevertheless, (3.17) is a very good approximation for metals at temperatures above the DEBYE[3] temperature Θ_D [142]. Better agreement with measurements can be obtained using an adjusted model

$$\lambda(T) = \lambda_0 + L\,\sigma\,T, \tag{3.18}$$

where the term λ_0 can even be of the same order of magnitude as the WIEDEMANN-FRANZ-LORENZ term. For temperatures below Θ_D, a second-order term might have to be included, according to the model

$$\lambda(T) = \lambda_0 + L\,\sigma\,T + \beta_\lambda\,T^2, \tag{3.19}$$

[3]Petrus Josephus Wilhelmus Debye (March 24, 1884 – November 2, 1966)

Table 3.3: Typical values for the thermal conductivities of various materials.

Material	λ [W/K]	References
Si Nanowires	$4.0 - 4.6 \times 10^4$	[77, 78]
Diamond	$1.0 - 2.5 \times 10^3$	[79]
Cu	$3.80 - 4.13 \times 10^2$ (300 K)	[39, 191]
	$3.63 - 3.92 \times 10^2$ (400 K)	[39, 191]
	$3.55 - 3.86 \times 10^2$ (500 K)	[39, 191]
	$3.49 - 3.79 \times 10^2$ (600 K)	[39, 191]
Al	2.37×10^2 (300 K)	[33]
	2.40×10^2 (400 K)	[33]
	2.37×10^2 (500 K)	[33]
	2.32×10^2 (600 K)	[33]
n-polySi	$0.16 - 0.41 \times 10^2$	[80, 192–195]
p-polySi	$0.17 - 0.20 \times 10^2$	[194]
Si	2.66×10^2 (200 K)	[79, 136]
	1.56×10^2 (300 K)	[79, 136]
	1.05×10^2 (400 K)	[79, 136]
	0.80×10^2 (500 K)	[79, 136]
Ge	0.95×10^2 (200 K)	[79, 136]
	0.60×10^2 (300 K)	[79, 136]
	0.44×10^2 (400 K)	[79, 136]
	0.33×10^2 (500 K)	[79, 136]
GaAs	0.44×10^2	[79]
Al_2O_3	0.21	[25]
SiO_2	0.014	[25]

where for temperatures below the DEBYE temperature Θ_D the parameter λ_0 is often negligible. Another approach to describe both temperature regimes is to use an empirical polynomial model of second order

$$\lambda(T) = \frac{\lambda_0}{1 + \alpha_\lambda (T - T_0) + \beta_\lambda (T - T_0)^2} \tag{3.20}$$

which is a TAYLOR series for the thermal resisitivity $1/\lambda_0$ at the reference temperature T_0. Here, λ_0 is the thermal conductivity at the reference temperature T_0 and the coefficients α_λ and β_λ are the corresponding first- and second-order temperature coefficients.

As a first approach the thermal conductivity can be assumed to follow the WIEDEMANN-FRANZ-LORENZ law, also for nonmetalic materials. However, to improve the model accuracy for semiconducting and insulating materials, a polynomial model may be used. A comparison of typical values of the thermal conductivities of common materials is given in Tab. 3.3 and Figure 2.9, where Tab. 3.3 shows a list of common materials ordered by descending thermal conductivities and Figure 2.9 gives an overview of the temperature dependence of various materials compared to Si and Ge.

3.4 Heat Capacitance

The heat capacitance is modeled according to a commonly used empirical formula introduced in [196, 197]

$$c_{\mathrm{p}} = A + BT + CT^2 + DT^3 + \frac{E}{T^2}, \tag{3.21}$$

where appropriate coefficients for this equation exist in tabular form and in diagrams, see for instance [131–138, 197–199] and Figure 2.9. Equation (3.21) yields very good results in simulations when compared to measurements. However, these coefficients are valid for the specified temperature range only.

3.5 Volume Expansion

The models which consider the thermal volume expansion have similar problems with the material structure as the models for the electrical and electrical conductivities. However, the thermal volume expansion coefficient is estimated with an equation given for bulk material as

$$\frac{\partial V}{V} = \alpha^{\mathrm{mech}}\, \partial T. \tag{3.22}$$

A rigorous mechanical investigation would require the appropriate volume expansion coefficients for the contributing materials. Those parameters can be either obtained from rigorous Monte Carlo calculations or from measurements. The coefficients given in Tab. 3.4 were obtained from measurements of bulk materials and give sufficient approximations. The microelectronic devices considered in this thesis show a slightly different behavior because the volume expansion described in (3.22) is mechanically constricted by additional material layers. The mechanical equation system is applied to materials which are embedded in rather stiff bulk material which normally do not allow expansion. This results in mechanically highly stressed material regions, especially if high temperature gradients are present. The thinner the material layers are the more sensitive they are and tend to relax by cracking. A crack in a layer often results in a fatal failure of the complete system, which is assumed to be the worst case. To avoid this type of failure, an appropriate investigation is required which also includes the mechanical subsystem.

If a block consisting of different materials tends to expand it is limited in its movement by the surrounding materials in which it is embedded. Hence, mechanical stress develops according to the values given in Tab. 3.4 for the thermal volume expansion coefficients.

The relation (2.111) considers expansion as well as contraction and torsion, thus, all phenomena where atoms are moved. However, if movements of atoms or atom clusters have to be considered, the mesh for those regions is very critical and can cause tremendous problems. For instance, in a region where a layer is contracted due to surrounded materials with a higher stiffness, the mesh is contracted as well. Due to limited numerical precision or approximations made in the material models, a single mesh point might move through a mesh edge between two other mesh points. This results in negative volumes and negative coupling coefficients in the finite element method and causes numerical problems, because the resulting system matrices are no longer positive definite and the solver have commonly severe convergence problems or even fails.

Table 3.4: Typical values for thermal expansion coefficients of various materials.

Material	α^{mech} [1/K]	References
Al	2.5×10^{-5}	[200]
Cu	$1.4-4.92\times10^{-5}$	[79, 142]
Mo	1.5×10^{-5}	[142]
W	$0.45-1.3\times10^{-5}$	[79, 142]
Ta	$0.65-1.92\times10^{-5}$	[142, 200]
Ti	$7.6-9.8\times10^{-6}$	[79]
Ag	5.7×10^{-5}	[142]
Au	4.32×10^{-5}	[142]
Al	6.78×10^{-5}	[142]
Si	2.33×10^{-6}	[79]
polySi	2.6×10^{-6}	[201]
SiO_2	$0.25-1.4\times10^{-6}$	[25, 79, 201]
Si_3N_4	$1.1-4.2\times10^{-6}$	[25, 79]
Al_2O_3	$5.6-7.1\times10^{-6}$	[25, 79]
TiN	6.5×10^{-6}	[200]

3.6 Mechanical Stress

If several materials are combined within a microstructure, mechanical forces occur if the volumes of the different materials change. Several effects may enforce volume expansion, for instance chemical reactions, phase interchanges and recrystallization, mass migration, and thermal expansion and contraction. In this thesis, the mechanical stress due to thermal expansion is considered. The mechanical stress σ^{mech} can be separated into a static and a thermal stress component [33] as

$$\tilde{\sigma}^{\text{mech}} = \tilde{\sigma}^{\text{stat}} + \tilde{\sigma}^{\text{therm}}, \tag{3.23}$$

where σ^{stat} represents the static stress component, which is impressed and mostly fixed by the fabrication processes, for instance material deposition, annealing, and packaging. The thermal component of the stress σ^{therm} is a transient quantity influenced by the ambient temperature and other state variables of the system. The corresponding hydrostatic pressure p^{mech} is then defined as the trace of the mechanical stress tensor σ^{mech},

$$p^{\text{mech}} = -\frac{1}{n}\,\text{trace}(\tilde{\sigma}^{\text{mech}}) = -\frac{1}{n}\sum_{i}^{n}\sigma_{ii}^{\text{mech}}, \qquad n = 3, \text{ for } \mathbb{R}^3, \tag{3.24}$$

which represents an averaged value and can be used for significant comparisons, as a figure of merit for optimization purposes, and to visualize the mechanical stress in a microelectronic device structure.

The local force density which acts on the material can be described by the mechanical stress tensor σ^{mech}, which can be derived from the mechanical strain tensor $\varepsilon^{\text{mech}}$ using the LAMÉ[4] formalism to model (2.110) as

$$\sigma_{ij}^{\text{mech}}(T) = B_{ij}^{\text{mech}}\,\alpha^{\text{mech}}(T-T_0)\,\delta_{ij} + \lambda^{\text{Lame}}\,\varepsilon_{ij}^{\text{mech}}\delta_{ij} + 2\,\mu^{\text{Lame}}\,\varepsilon_{ij}^{\text{mech}}. \tag{3.25}$$

[4]Gabriel Lamé (July 22, 1795 – May 1, 1870)

The quantities λ^{Lame} and μ^{Lame} are the LAMÉ constants, which can be expressed by YOUNG's modulus E^{mech} and POISSON's ratio ν^{mech} [96, 97] as

$$\lambda^{\text{Lame}} = \frac{\nu^{\text{mech}}\, E^{\text{mech}}}{(1+\nu^{\text{mech}})(1-2\,\nu^{\text{mech}})}, \tag{3.26}$$

$$\mu^{\text{Lame}} = \frac{E^{\text{mech}}}{2\,(1+\nu^{\text{mech}})}. \tag{3.27}$$

For orthotropic materials the number of independent components for the stiffness and stress tensor is reduced due to energetic considerations and symmetry in the crystals [96, 97, 202]. Hence, the number of independent components of the stiffness tensor (forth rank) is reduced to nine and for the stress tensor (second rank) to six. Hence, the mechanical problem can be expressed in terms of vectors and matrices [202] according to the VOIGT notation.

3.7 Interconnect Reliability

To estimate the life time of electronic products, several different characteristic quantities have to be extracted. However, the estimation has still a considerable variance because this measure evaluates the worst case only. The goal is to obtain the mean time to failure (MTTF) for standard operation conditions by extrapolations based on data of MTTF for elevated temperatures. To enforce a change of material properties, it requires a certain amount of external energy to cause harm to the device structure, if the energy exceeds a critical level.

The worst case for interconnect lines is a direct current, where the maximum of the electrical current density J_{PEAK} of a certain interconnect element is given by the approximation

$$\sup\left(J(\partial\mathcal{V})\right) \;=\; J_{\text{PEAK}} \approx \frac{I_{\text{PEAK}}}{A}. \tag{3.28}$$

Here, the maximum current density is estimated as the maximum of the expected current I_{PEAK} through the interconnect divided by the local area A of the interconnect. The advantage of this approximation is that the peak value of the electrical current can be measured and verified for worst-case considerations. Other characteristic quantities are the mean value and the root mean square value of the current density J

$$\mathbf{J}_{\text{AVG}} \;=\; \langle \mathbf{J} \rangle = \frac{1}{T}\int_0^T \mathbf{J}(t)\,dt, \tag{3.29}$$

$$J_{\text{RMS}} \;=\; \sqrt{\langle |\mathbf{J}|^2 \rangle} = \sqrt{\frac{1}{T}\int_0^T |\mathbf{J}(t)|^2\,dt}, \tag{3.30}$$

respectively. Here, the values are averaged over a characteristic time span, for instance a full period for oscillating signals.

There exist several additional constraints which affect the reliability of interconnect structures, such as the maximum allowed electric field for dielectric break down or the minium resistivity of the dielectric to limit leakage currents between different interconnect structures. However, these constraints are currently not of high interest because these requirements are just certain constraints of process technology nodes and can be taken into account in advance during the design phase.

3.7.1 Electro-Migration

Current transport involves two types of charge carriers: electrons and holes (defect electrons), and ions and vacancies (defect ions) as depicted in Figure 3.2. While the first charge carrier type is normally used for the device operation, the second produces disadvantageous effects and causes tremendous failure. For instance, if ions are transported through the interconnect lines the resulting ion current has also to be considered as mass flux.

Atoms on ideal crystal lattices need a high activation energy to change the position on the lattice (cf. Figure 3.2). Unregularities in the crystal significantly descrease the activation energy. Such promoting factors are interstitials and the grain structure due to irregular crystal growth, and impurities in the grain as well as at grain boundaries at the material interfaces (cf. Figure 3.3). These factors are advanced by additional energy provided by elevated temperatures in the structures due to self-heating and external heat sources, where the increasing current densities reach critical values at which self-heating effects dominate the reliability issues.

At elevated temperatures, the ions and defects have an increased diffusivity, which is further enhanced by external applied current densities. The atoms move in the direction given by the electrons. At a certain point in time, a microscopic hole (void) is seeded on the one side and a pocket (hillock) is formed on another side. The void can cause an opening of an interconnect line while the hillock might shorten two adjacent interconnect structures. Both scenaria yield fatal errors.

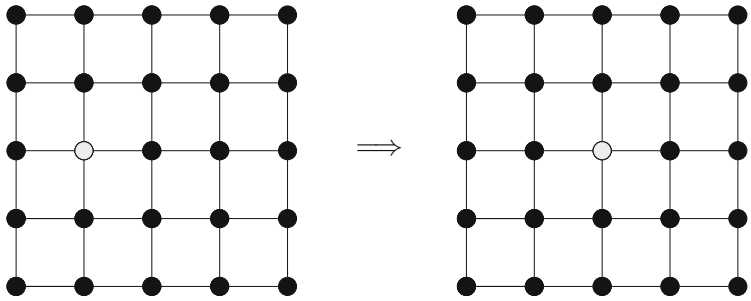


Figure 3.2: Movement of vacancies by ion diffusion into lattice vacancies.

The dynamics of moving atoms and their vacancies can be described by a drift-diffusion model similar to that for electrons in semiconductor materials in (2.14) and (2.15), but with different mobilities and diffusion coefficients. A model for the mobility of vacancies $\mu_{\text{vac}}^{\text{mob}}$ has been proposed in [203] as

$$\mu_{\text{vac}}^{\text{mob}} = \frac{a^2 f}{k_{\text{B}} T} \exp\left(-\frac{\mathcal{E}_{\text{A}}}{k_{\text{B}} T}\right), \tag{3.31}$$

where a is the lattice constant, f is the attempt frequency for electro-migration, and $\mathcal{E}_{\text{A}}$ the activation energy. The diffusion coefficient is either a constant or an empirical formula to account for the dependence on the local stress [36, 204, 205]. The driving force $\mathbf{F}_{\text{e}}$ can be determined by

$$\mathbf{F}_{\text{e}} = \text{Z}^* q \mathbf{E}, \tag{3.32}$$

where Z^* is the effective valence number for a given certain material [203, 206]. Measured values for Z^* and $\mathcal{E}_{\text{A}}$ can vary by 300% and 250%, respectively [203, 207]. Therefore, each material for a certain technology node needs a separate calibration to account for all effects which can occur during the fabrication process and during operation.

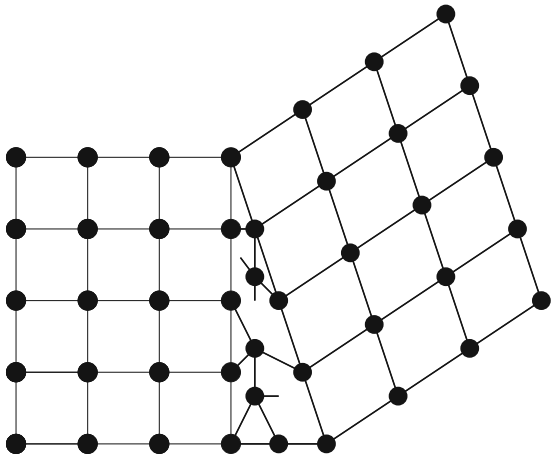


Figure 3.3: Vacancies at grain boundaries.

Regions in which the mechanical stress is notably important are areas where high current densities occur simultaneously with high temperature gradients [100]. In these regions with very high current densities and elevated temperatures, electro-migration can be observed [208]. The occurrence of this failure type is accelerated with increasing temperature. This effect can be used to estimate reliability and life time. However, this accelerated test method is only an estimation for the worst case.

The previously presented models have assumed that the shape, the orientation, and the distribution of the size of the grains does not affect the stress distribution. Unfortunately, the grain boundaries are layers of a finite thicknesses and the mechanical constants vary along these structures. Due to the different crystal orientations in adjacent grains, the same material in different grains behaves slightly different even if the same mechanical load is applied [209]. For instance, the diffusion along paths through the interconnect results in different current densities according

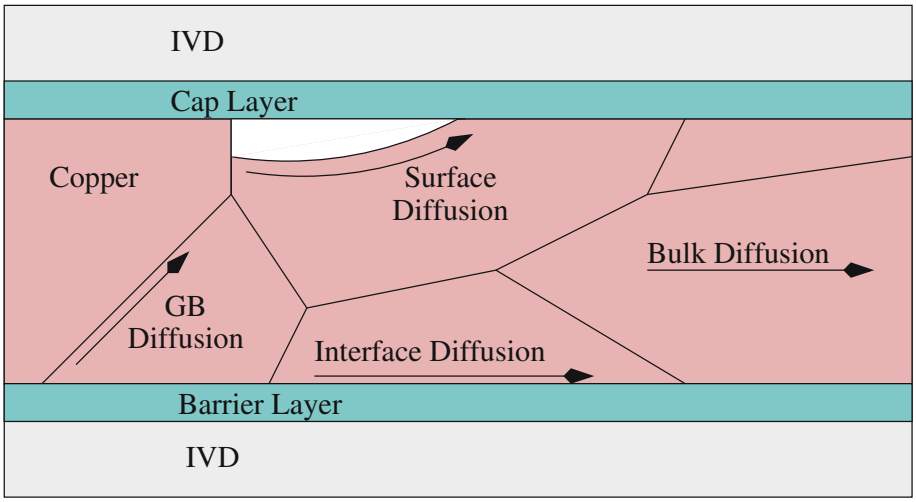


Figure 3.4: Diffusion paths in a typical copper interconnect line. The diffusion paths determine the mass flux due to ion diffusion along distinguished paths.

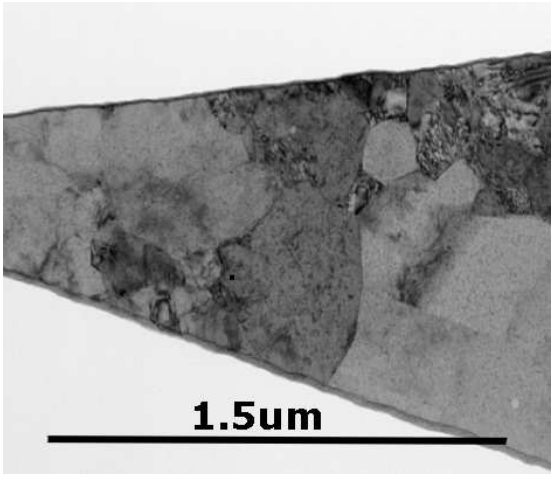


Figure 3.5: TEM picture of a Cu interconnect test structure showing a wide range of different grain sizes due to confinements of the geometry [32].

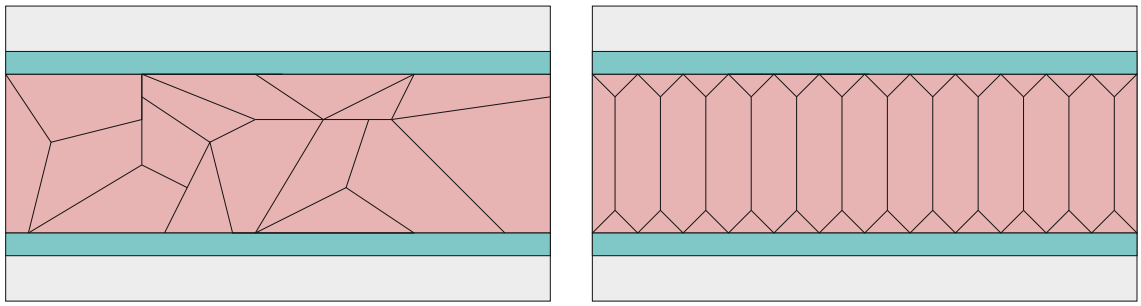
to the corresponding local crystal structure. Apparently, the diffusion coefficients depend also on an activation energy [33,163], which itself is a function of the material adhesion and the regularity of the lattice, as shown for Cu in Figure 3.4. A cross section of a Cu interconnect structure is shown in Figure 3.5, where no information about the distribution of the grain sizes is given. Nevertheless, process technologies have been developed to control the grain sizes and distribution of the shape of grains in a certain regime to provide more uniformity in the interconnect lines for highly sophisticated microelectronic devices.

The uncertainty of the position and the size of grain boundaries are new critical parameters for reliability investigations. Unfortunately, these parameters cannot be determined excactly in advance. However, a statistical description of these parameters can be applied and provides rather good agreement with measurements but does not account for the degradation effects in the metal, the moving grain boundaries, the movements of the defect location, or recrystallization [164].

As Tab. 3.5 shows, copper ions at lattice sites in bulk-sized grains have the tendency to diffuse not as much as near the surface. For ions located near the interfaces (grain boundaries or material interfaces) the diffusion constant nearly doubles. For copper atoms at surfaces the activation energy is only one third of that in the bulk material. Hence, material regions near surfaces and material interfaces have to include the ion diffusion of copper rigorously to predict their reliability properly.

Table 3.5: Typical activation energies $\mathcal{E}_A$ along different paths for diffusion in Cu [33,158,207, 210–212].

Diffusion Path	$\mathcal{E}_A$ [eV]
surface	0.5 – 0.7
material interface	0.8 – 1.25
grain boundary	1.2 – 1.25
bulk (within the grains)	2.1



(a) Disordered grains. (b) Bamboo grain structure.

Figure 3.6: Different grain structures for Cu interconnect lines.

The material Al has been substituted by Cu to increase the intrinsic activation energy for electro-migrationfrom 0.6 eV for bulk Al to 1.2 eV for bulk Cu [207]. Cu has a higher melting temperature. Incidentally, this substitution should have caused less problems with electro-migration. However, the material interface conditions were not considered in the very beginning of the transition. In the common Cu technology, the barrier and cap layers for interconnect lines increase the mechanical weakness of the interfaces to Cu, which cause an additional reduction of the activation energy.

Another contribution to reliability reduction can be found in the unfavorable microstructure of the materials used in interconnects. The microstructure of Cu is formed during electro-plating, where the final shape of the grains is determined by exogenous parameters. The microstructures of two different electro-plating fabrication process are depicted in Figure 3.6a and Figure 3.6b, where in Figure 3.6a a technology is presented where the microstructure of Cu cannot be controlled. A better result can be achieved if the process temperature, and the seed and barrier layers can be appropriately controlled. As a result, the grains evolve homogeneously with similar shape and with nearly the same size distribution, as shown in Figure 3.6, where Figure 3.6b shows a 'bamboo' structure, which has the benefit that the current through such structured interconnects flows perpendicular to most parts of the grain boundaries. This provides less energy to the atoms to move along the grain boundaries inhibiting an additional ion current.

The most significant contribution to electro-migration is due to surface diffusion, which is also a main factor for reliability issues in Cu interconnect lines. The bamboo structure in Figure 3.6b could easily be enhanced further by chemical-mechanical polishing (CMP), where the upper zones of the bamboo structures can be removed. Hence, the removed upper side is a step towards an ideal microstructure by minimizing the number of grain boundaries not perpendicular to the average current density vector. Removing the most upper region of Cu, the original Cu structure has to be deposited with a higher thickness as required. The additional Cu is removed by a CMP process. With that procedure, the upper critical zone is removed from the copper surface. The bottom of the Cu surface as well as the side walls are treated by special seed and barrier layers, which prevent at one hand the diffusion of Cu into the dielectrics. On the other hand these layers act as catalysts during the initial Cu deposition to form an "ideal" initial microstructure.

During operation, elevated temperature and high current densities, it can happen that the vacancy concentration due to electro-migration inside a critical material region exceeds a certain limit (vacancy concentration $\approx$ number of atoms per unit cell of the basis interconnect material). In that case, a void has been seeded in that region [213]. This behavior is called inverse clustering

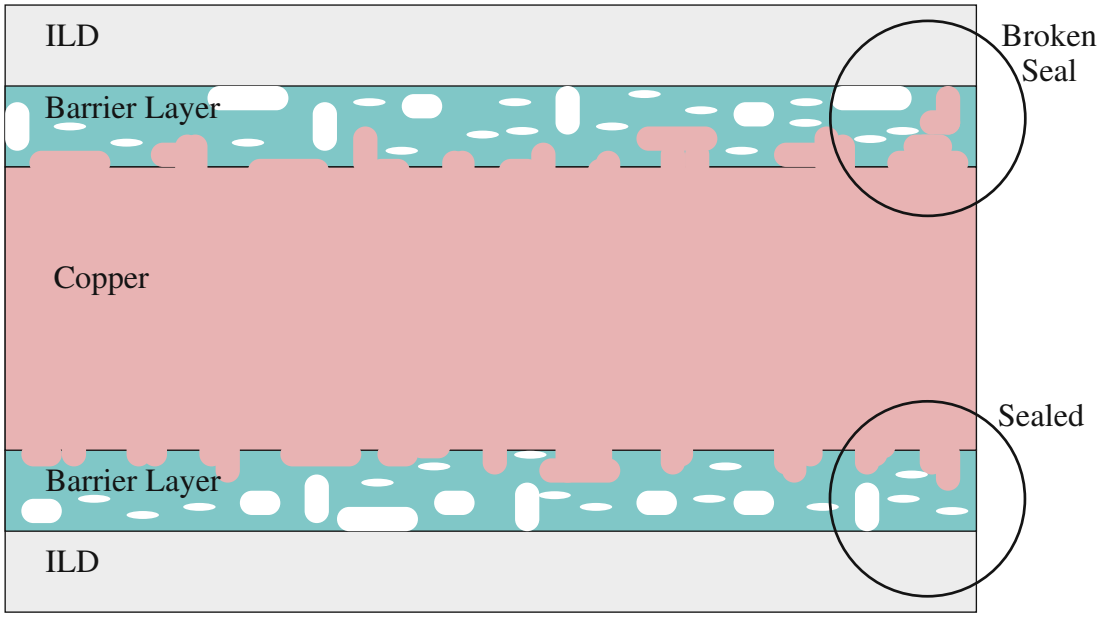


Figure 3.7: Novel barrier material type, showing the benefits and drawbacks of material porosity, e.g. of CoWP. To prevent the diffusion of Cu atoms into the ILD and dielectric layers a guaranteed sealing of the (porous) barrier materials have to be enforced.

because the modeling approach considers the accumulation of precipitates at the beginning in terms of a vacancy concentration in the microscopic level and at the intermediate state via voids as a macroscopic description. With a seeded void, an additional surface has been introduced into the material region which has been previously assumed to be bulk-like. This surface results in a region with reduced activation energy and therefore promotes the diffusion of Cu atoms, thus accelerating the growth of the void. A growing void yields a material transport from the void site to a physical barrier material where a pocket (*hillock*) is built. The voids itself may cause an interruption of the interconnect line, while the materials pocket may cause a shortening between two adjacent interconnect structures. Providing a reservoir of the migrating materials is an improvement [214], but only of limited help for the interconnect reliability. While the time that is needed to seed and form a void is considerably high, the time needed to move the voids and form hillocks and big voided regions is orders of magnitude less than the time for the seeding and forming process.

Better results for the improvement of reliability can be obtained if the roots of the problem are rigorously tackled. Those are the high current densities, the reduced line dimensions, the resulting elevated temperatures, and the weak material interfaces with their poor adhesion on certain materials causing a considerable problem in addition to the natural material diffusion, the current density, and the high temperatures. The main reason for enhanced surface diffusion and electro-migration has been identified as the unfavorable crystal structure close to the material interfaces. If one can provide an (ideal) crystallized Cu layer up to the material interface, the Cu atoms would be located at more or less ideal lattice sites and thus would require a higher activation energy to move.

The introduction of new interface materials such as Ta and TaN have provided promising results by slightly rising the activation energy and lowering the migration tendency. However, the fundamental problem of not ideal crystal structures with high diffusion constants has not been

tackled. Only with the introduction of new porous barrier materials, the electro-migration effects has been significantly reduced. If the pores of these materials exceed a certain size, Cu atoms have the possibility to move into those pores. This yields a starting crystallization very close to the surface, actually in the barrier layer (cf. Figure 3.7). That provides a good alternative to Ta and TaN coatings. The deposition of such porous materials requires a guaranteed seal barrier coating. With completely sealed coatings, the Cu atoms diffuse into the ILDs, causing serious problems as already discussed.

Unfortunately, the CoO of these porous materials is considerably large. Thus, the implementation of such materials might take some time in the common Cu technology nodes. With the increasing number of high-κ and low-κ materials, where most of them are porous materials, also the porous barrier layers will be available in the near future.

3.7.2 Mean Time To Failure

The failure distribution of microelectronic devices follows a WEIBULL[5] distribution. At the beginning, early failure may occur due to fabrication faults. At the end of the device's life time, failure is mostly due to material fatigue of one or more parts of the device. To determine the reliability especially for metals, BLACK [215,216] has proposed a modified ARRHENIUS law, which can be arbitrarily adapted and calibrated to special failure mechanisms. BLACK's equation reads

$$\text{MTTF} = A\,|\mathbf{J}|^{-n} \exp\left(-\frac{\mathcal{E}_\text{A}}{k_\text{B}T}\right), \tag{3.33}$$

where the local mean time to failure (MTTF) is inversely proportional to a certain power n of the local current density and directly proportional to $\exp(-\mathcal{E}_\text{A}/k_\text{B}T)$, with $\mathcal{E}_\text{A}$ as the activation energy. The proportionallity constant A is geometry-dependent, and must be determined by measurements for different interconnect geometries.

However, BLACK's equation describes an empirical observed material behavior and is thus not valid for arbitrary use. It requires a separate calibration for each different failure mechanism. Nevertheless, BLACK's equation is still commonly used to estimate the reliability with respect to the mean time to failure [5,33,217].

[5] Ernst Hjalmar Waloddi Weibull (18 June 1887 – October 12, 1979)

Chapter 4

Optimization for Technology CAD

"Man darf nicht das, was uns unwahrscheinlich und unnatürlich erscheint, mit dem verwechseln, was absolut unmöglich ist."

Carl Friedrich Gauß[1]

This chapter first discusses the different optimization techniques and strategies that are commonly used in modern optimization applications. The second part of this chapter deals with the industrial requirements for optimization as well as its challenges for TCAD applications. The third part shows the need of an optimization framework and the resulting concepts which fulfills the presented requirements.

In this work the term optimization is used as a search for a minimal or maximal value for an objective function (also called score function) within certain defined constraints. It is a widely used practice that optimization problems are formulated as a minimization task of an objective score function

$$f_{\text{Score}}(\mathbf{x}) \rightarrow \min, \quad x \in \Omega_{\text{G}}. \tag{4.1}$$

To perform a maximum search, the formalism can be transformed to a minimum search for negative values of an objective score function [219]

$$f_{\text{Score}}(\mathbf{x}) \rightarrow \max \quad \Longleftrightarrow \quad -f_{\text{Score}}(\mathbf{x}) \rightarrow \min. \tag{4.2}$$

Despite of the different score functions for the optimization, the mathematical convergence criteria for both optimization algorithms remain valid [219][2]. The optimization problems discussed in this thesis are finite-dimensional optimizations of the following type: A given n-dimensional variable vector $x \in \Omega_{\text{G}} \subseteq \mathbb{R}^n$ of an p-dimensional objective score function $f_{\text{Score}} : \mathbb{R}^n \rightarrow \mathbb{R}^p$ has to be optimized globally in order to obtain a resulting vector $\mathbf{x}_{\text{opt}}$ which minimizes the value of the score function in a certain domain $\Omega_{\text{G}} \in \mathbb{R}^n$. Equation (4.1) can also be expressed by using the following equivalent notation:

$$\mathbf{x}_{\text{opt}} \in \left\{ \mathbf{x} \,|\, \forall \mathbf{y} \in \Omega_{\text{G}} : f_{\text{Score}}(\mathbf{x}) \leq f_{\text{Score}}(\mathbf{y}) \right\}. \tag{4.3}$$

[1] Carl Friedrich Gauß (April 30, 1777 – February 23, 1855),
"Do not confound what appears unnatural with the absolutely impossible." [218]

[2] However, an objective score function with negative values often yields different results due to different implicitly taken assumptions in certain implementations of optimizers.

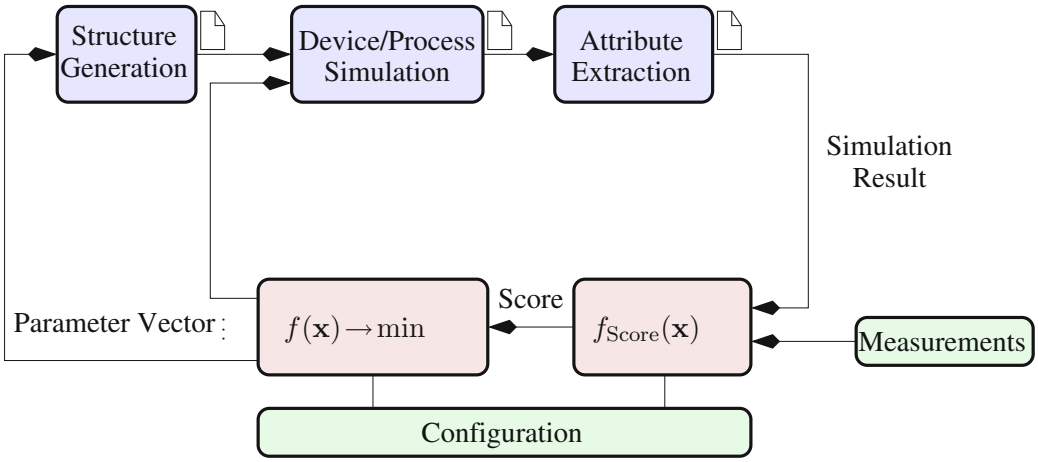


Figure 4.1: Generic optimization loop for multiple purposes.

In this definition, the function $f_{\text{Score}}(\mathbf{x})$ denotes a continuous objective score function which has to be minimized. To determine a mathematically suitable criterion for minimization, the objective function $f_{\text{Score}}(\mathbf{x})$ has to apply a metric which maps the simulation result in a scalar-valued quantity. The industrial requirements requires an objective score function of the form

$$\mathbf{y} = f_{\text{Score}}(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}^p. \tag{4.4}$$

However, this type of optimization would require optimization tools that are capable to operate with PARETO[3] sets which is not commonly available in optimization tools. To overcome this kind of problem, a weighted norm can be applied to the score function and reads

$$f_{\text{Score}}(\mathbf{x}) = \sqrt[p]{\sum_{i=0}^{p} \alpha_i |f_{\text{Sim},i}(\mathbf{x})|^p} : \mathbb{R}^n \rightarrow \mathbb{R}, \tag{4.5}$$

where the result of this score functions is a real number, which can be optimized using the property that $\mathbb{R}$ is an ordered field. If the weight parameters α_i are set to 1, (4.5) can be written as

$$f_{\text{Score}}(\mathbf{x}) = \|f_{\text{Sim}}(\mathbf{x})\|_p. \tag{4.6}$$

In typical TCAD applications such objective score functions represent sequences of simulation software tools and therefore $f_{\text{Score}}(\mathbf{x})$ can also include some necessary post-processing steps. A typical data flow of an optimization run is shown in Figure 4.1 where certain different input parameters can be applied to each simulation tool separately. In this depicted example, the simulation flow consists of a tool for the generation of the device geometry, the device simulator, and at the end a tool for the extraction of objective parameters from a finite-dimensional simulation result in order to compare it with reference data, which can be either constant values, analytical functions, or quantities in tables obtained from measurements.

In order to limit the optimum search to a certain domain and to weight or exclude certain parameter constellations, the input parameter space can be constrained by lower and upper bounds as well as by a finite number of constraint functions. In the following chapter Ω_{G} denotes

[3] Vilfredo Federico Damaso Pareto (July 15, 1848 – August 19, 1923)

a convex[4] and closed finite-dimensional domain $\Omega_{\mathrm{G}} \subset \mathbb{R}^n$ which can be further constrained by functions $g_i(x)$ which yields in the most general case a non-convex shape and can be expressed as

$$\Omega_{\mathrm{G}} = \{x \,\epsilon\, \mathbb{R}^n \mid g_i(x) \leq 0, \;\; i = 1, ..., m\}. \tag{4.7}$$

The constraint functions $g_i(x)$ can map for instance some physical constraints to the input parameter domain, represent some technological or economical constraint from the fabrication processes, or these functions can be used to avoid parameter constellations which are not allowed, either by specifications or due to patent laws. However, these functions have to be individually chosen for a particular optimization problem. Constraints for Ω_{G} can be applied *a-priori* in contrast to constraints for output parameters. Therefore, the resulting domain for valid output parameters $\Omega_{\mathrm{F}} \subseteq \mathbb{R}^p$ is defined as the physical feasible values and the corresponding constraint function reads

$$f_{\mathrm{Sim}} \in \Omega_{\mathrm{F}}, \tag{4.8}$$

where the valid values of the output domain Ω_{F} are given by the nested function of the score function f_{Score} applied to the final results of the function f_{Sim} which describes the sequence of the different simulation tools used.

While *a-priori* constraints limit the search domain only, *a-posteriori* constraints restrict the simulation results, which requires to calculate a complete simulation sequence to obtain a single result. This is thus very costly in time. Therefore, one tries to transform the constraints for the output in constraints for input parameters.

For some cases, where the constraints of simulation results have to be included into the constraint functions, an estimation can be performed to approximate the simulation results in advance. If the calculations of the original function is very time-consuming compared to the time for evaluating the approximation, this method provides the benefit of saving time by excluding certain *a-priori* known not valid simulation results.

With the domains (4.7) and (4.8) the initially constrained minimization problem (4.1) can be reformulated by using barrier or penalty functions $P_k(\mathbf{x}, \mathbf{y})$ in order to obtain an unconstrained surrogate optimization problem [219, 220].

$$P_k(\mathbf{x}, \mathbf{y}) \to \infty, \qquad k = 1, 2, \ldots \qquad \begin{cases} \forall \mathbf{x} \notin \Omega_{\mathrm{G}} \\ \forall \mathbf{y} \notin \Omega_{\mathrm{F}} \end{cases}, \tag{4.9}$$

which provide the possibility to use the original optimization framework with minor changes which can be specified by the user. In (4.9) $\mathbf{x}$ symbolizes the input parameters and $\mathbf{y}$ the output parameters or the results of the simulation and the score function. The barrier and penalty functions try to account for the behavior of the output $\mathbf{y}$ as accurately as possible in order to save computation time. Then the penalty problem reads

$$f_{\mathrm{Pen}}(\mathbf{x}) = \|f_{\mathrm{Score}}(\mathbf{x})\|_p + P_k(\mathbf{x}, \mathbf{y}) \to \min. \tag{4.10}$$

To conveniently apply such functions to a particular problem, the penalty function can be adapted to user defined constraints. For instance, there exist several different approaches for barrier and penalty terms. In the following, the barrier and penalty functions P_k are defined using a sequence of penalty parameters r_k, where

$$r_k > 0, \qquad k = 1, 2, \ldots \qquad \text{and} \qquad \lim_{k \to \infty} r_k = +\infty. \tag{4.11}$$

[4] Confer Section A.1 for the definition and some remarks on convexity.

Since the transition of $r_k \to \infty$ in (4.11) is numerically impossible, various finite approximations have been proposed in literature. However, the use of large numbers can result in serious convergence problems, because of the numerical calculations of the gradients when regions are considered which are located very close to the domain boundaries.

The formulation for barrier and penalty functions often considers a certain margin of the valid parameter domain. To prevent the search algorithm from moving too close to the domain boundary, a barrier function is applied that reaches the value infinity at the boundary. The penalty function charges a certain fine for the function if the search algorithm is outside of the specified domain. The inexact penalty functions have are vanishing inside the allowed domain. There are several methods to implement such barrier and penalty functions:

- The exact penalty function [221–223] vanishes inside the specified parameter domain and reaches a certain value greater zero outside the domain:

$$P_k(\mathbf{x}) = r_k \sum_{i=1}^{m} \max\{0, g_i(\mathbf{x})\}, \tag{4.12}$$

  where the corresponding constraint functions g_i apply with $g_i(\mathbf{x}) \leq 0$ inside the valid parameter domain and > 0 outside.

- The exact quadratic penalty function [224, 225] is quite similar to the previous one, but shows generally a quadratic increase with the distance from the domain boundary:

$$P_k(\mathbf{x}) = r_k \sum_{i=1}^{m} \max^2\{0, g_i(\mathbf{x})\}, \tag{4.13}$$

  where the corresponding constraint functions g_i apply. Again, $g_i(\mathbf{x}) \leq 0$ inside the valid parameter domain and > 0 outside.

- A logarithmic barrier function [226–228] offers the possibility of directing the parameter search to particular subdomains, in which the score function is superposed by a logarithmic function in the whole domain. Hence, the minimum of the sum of the score function and the logarithmic barrier function is located in the subdomain in which the optimum is. At the boundaries and outside the domain, the barrier function reaches infinity according to the definition of $P_k(\mathbf{x})$:

$$P_k(\mathbf{x}) = \begin{cases} -\dfrac{1}{r_k} \displaystyle\sum_{i=1}^{m} \ln[-g_i(\mathbf{x})], & \mathbf{x} \in \Omega_{\mathrm{G}} \\ +\infty, & \text{otherwise} \end{cases}, \tag{4.14}$$

  where the corresponding constraint functions g_i apply for $g_i(\mathbf{x}) \leq 0$ inside the valid parameter domain and > 0 outside.

- With an inverse barrier function [224], the search region inside the parameter domain can be predefined similarly to the logarithmic barrier function, but with a differently shaped approximation:

$$P_k(\mathbf{x}) = \begin{cases} -\dfrac{1}{r_k} \displaystyle\sum_{i=1}^{m} \frac{1}{g_i(\mathbf{x})}, & \mathbf{x} \in \Omega_{\mathrm{G}} \\ +\infty, & \text{otherwise} \end{cases}. \tag{4.15}$$

  Here, the constraint functions g_i are applied, where $g_i(\mathbf{x}) \leq 0$ inside the valid parameter domain and > 0 outside.

- An inexact exponential penalty function [229] offers an efficient method to priorize a particular subdomain of the parameter domain without dealing with infinity. However, the value of the barrier function increases rapidly when the search algorithm leaves the valid parameter domain.

$$P_k(\mathbf{x}) \;=\; r_k \sum_{i=1}^{m} \exp[r_k\, g_i(\mathbf{x})], \tag{4.16}$$

  where the constraint functions satisfy $g_i(\mathbf{x}) \leq 0$ inside the valid parameter domain and > 0 outside.

However, the user has always to choose the appropriate barrier or penalty functions in order to account for his particular needs and to check the convergence behavior of the whole optimization algorithm in advance. For instance if the score function and the contributions from the barrier and penalty function differ by many orders of magnitude, the discretization and gradient calculations algorithm of the optimizer might run into numerical problems in terms of precision and accuracy.

According to the principal behavior of the score function within the optimization problem, an appropriate barrier or penalty function modifies the original optimization problem in the same way as the score function would, but provides an additional term to the score function which allows to exclude certain domains from the original parameter space or to priorize certain subdomains for example if several optimal values are expected. Since many of the available optimization strategies do not inherently support constraint functions, additional barrier and penalty function are often used in non-linear optimization problems where only a certain set of optimization strategies are available for utilization with in a framework.

4.1 Basic Issues on Optimization

From a mathematical point of view, optimization is a rather straight forward task for standard problems, for instance if a set of linear or non-linear equations is considered. However, the optimization of TCAD problems has to deal with highly complex score functions (objective functions). These score functions include pre-processing steps like geometry construction, meshing, and process simulation, the actual simulation with its manifolds of different models, and at the end, there is a set of post-processing steps to obtain the appropriate quantities from the simulation results. From these data sets, the quality has to be determined with respect to a user-defined reference. These sets of equations and models cannot be sufficiently determined to obtain a best strategy for an optimization because the different input parameters can cause to switch the behavior of the models due to different operation conditions. Therefore, the global behavior of the simulation results changes. Hence, the convergence of the simulation cannot be guaranteed for all parameter sets of the domain of the input parameters. For instance, changing the geometry might require a different meshing algorithm to obtain a sufficiently good resolution for reasonable simulation results. On top of that, numerical difficulties may arise within every part of the optimization loop.

From a mathematical point of view, the exchange of the score function f_{Score} with its negative values $-f_{\text{Score}}$ transforms a minimization to a maximization of f_{Score}, showing the same convergence criteria [219]. However, some implicit assumption of certain properties of the score function yields a limited range of applications. For instance, a very common procedure is to use the absolute value of the score function $|f_{\text{Score}}|$. As a consequence, the information about

the sign of the score function vanishes. Whether this is the case or not should be denoted in the particular optimizer's documentation. However, the user has to be aware of these different procedures within the optimizer and to take it into account by using appropriate user-defined score functions.

Improvements of the optimization algorithm in terms of convergence can be achieved by iterative methods [230] (cf. NEWTON's[5] iteration scheme as a propaedeutic example). However, these methods are limited by a certain domain in which the optimization algorithm shows the property of high convergence. With a starting point outside of this defined domain, the convergence speed can often not be determined, and sometimes the convergence of the optimization algorithm is not even guaranteed. An additional challenge in optimization is that there is no knowledge whether the objective function is twice continuously differentiable because many optimization algorithm assume this property.

There are several different optimization methods, where each having its benefits and drawbacks. The optimization methods can be dived into several groups: local and global An additional classification of optimization strategies separates the techniques in incomplete, asymptotically complete, complete, and rigorously complete optimization algorithms. An optimal optimizer would combine the benefits of the iterative optimization algorithms and the genetic and randomized search, certain techniques for stochastic, heuristic, and genetic approaches [231] can be combined to build new optimization strategies in order to increase the probability of finding a good optimum within a short period of time. Other improvements to speed up the optimization is to use parallel optimization [232] and to select the appropriate optimization algorithm [233].

Typical optimizations take a lot of time. Therefore, reducing execution time for optimization is very important to provide results within reasonable time. There are several principal methods to speed up a optimization run:

- Parallelization of the optimization algorithm:
  This methods is advantageous when a huge amount of data has to be processed in each optimization iteration step. For instance a complex response surface method (RSM) optimization requires a huge amount of computational power and memory to calculate the parameters and coefficients for the next iteration. In this case, a parallelization makes sense because the computational effort and execution time is of the same order of magnitude as the evaluation of a particular parameter. However, the parallelization of the optimization algorithm often does not decrease the overall optimization time all that much, because the evaluation of particular parameter sets usually takes much longer (minutes to weeks) than the calculation of the next parameter sets (seconds to minutes). Hence, the execution time of the optimizer program can often be neglected compared to the time which is necessary to evaluate the parameters by its score function, which includes the entire simulation sequence.

- Parallelization of the evaluations of parameter sets:
  The evaluations of parameter sets can be generally computed concurrently because the proposed parameter sets from the optimizer do not depend on each other. For instance, if a optimizer has to calculate a gradient in a certain point of the parameter domain, a certain amount of necessary parameter sets for evaluation are submitted concurrently. Hence, these parameter sets can also be evaluated in parallel and, if possible, on different computational nodes. However, if the calculation of gradients are computational expensive, gradient-free optimization algorithm are advantageous.

[5] Sir Isaac Newton (January 4, 1643 – March 27, 1727)

- Parallelization of the solvers in the simulation tools:
  To parallelize the solver in the simulation tools increases the performance of each parameter evaluation during the optimization. However, the influence of whether the simulation tools use parallel solvers is very limited because many vendors distribute their software with fixed capabilities and features.

- Selection of a good optimization strategy and an appropriate score function:
  To choose the appropriate optimization strategy is one of the most critical tasks during the optimization setup because the strategy defines several additional parameters and at the end also the convergence speed. However, the score function has to be chosen according to the type of the problem and according to the selected optimization strategy to provide fast and accurate optimization results. If additional constraint functions are necessary, they have to be aligned to the score function to avoid numerical problems within the optimizer. If, for instance, the numerical values are too different which can cause problems with the precision which yields numerical noise and therefore wrong optimization results.

The main goal of an optimization is to find the parameter set that yields the global optimum in a minimum number of iterations. However, this search is often very complex and would require to check the complete parameter space in all dimensions which often leads to serious problems in terms of finite resources of computational power and time. To overcome this type of problem, the optimization can be split into three major parts where the first part deals with the separation of the most significant and less significant parameters. This part requires a good knowledge of the problem class which has to be optimized. Once the parameters are separated with respect to their importance and significance, the optimization of the significant parameters can be started. This leads to certain good parameter sets which can be fine tuned in the third part where all parameters are included in the optimization. Appropriate constraint functions and close intervals for the parameters can drastically reduce the computational effort for the parameter search.

4.1.1 Optimization Loop

An optimization loop for industrial applications includes several aspects (e.g. of robustness, speed, etc.) in addition to the numerical optimization. However, the basic structure of the optimization is the same for all applications. A typical optimization flow is depicted in Figure 4.1, where the different parts of the optimization are presented. According to the different levels where the optimization is applied, the interactions with the optimization loop and the results of the different optimizations are different. An optimization in a typical TCAD software environment includes several applications as listed below:

- Device characteristics to improve performance and to reduce parasitic effects as well as self-heating effects

- Single process steps to obtain more realistic device structures for device simulations

- Sequences of single steps, for instance several process steps (material deposition, mask deposition with optical proximity correction, material etching, and mask stripping)

Typical optimization tasks in a semiconductor device fabrication environment requires:

- Optimization of a single process step to increase yield for this particular step or to identify critical parameters to apply them for reliability calculations.

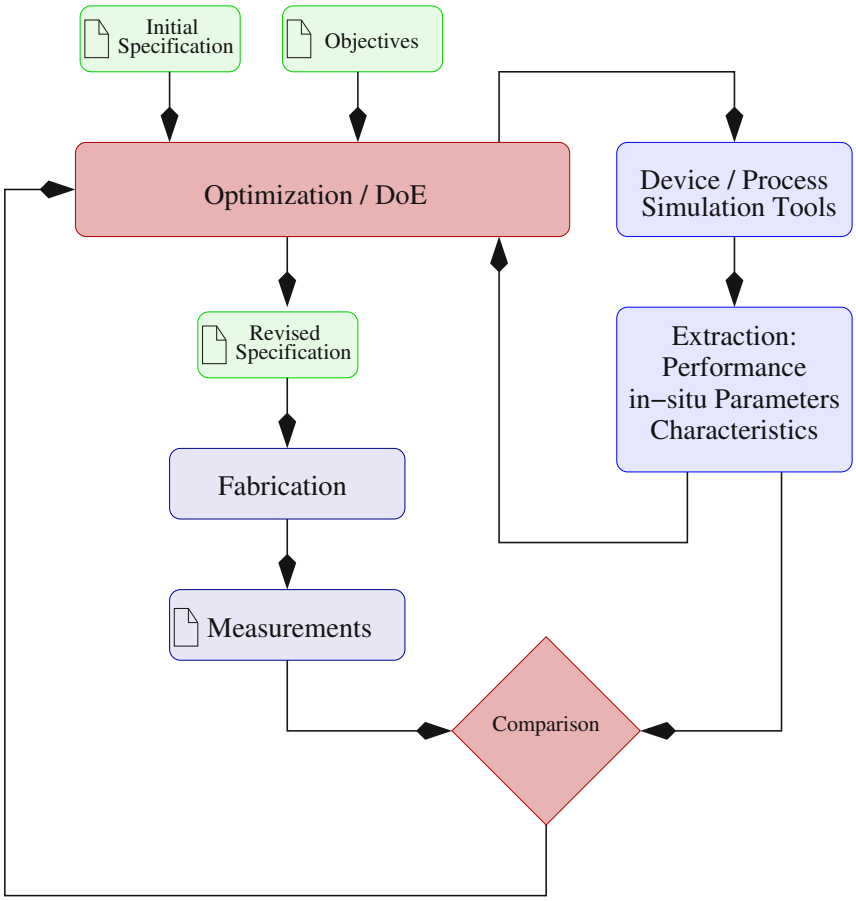


Figure 4.2: Integration of optimization loops in device design and device fabrication process

- Optimization of a technology is a steady process during the whole life cycle of the particular semiconductor device. This includes improving yield, and reliability of the device, and to port the process to a smaller-scaled technology node to increase the number of devices per lot.
- The most complex task is to optimize the complete formula of a technology, including the resulting device characteristics which often require several weeks to months from fabrication to the final results of the measurements of characteristics and reliability.

To decrease the time from the start of optimization until the final results are obtained, certain aspects can be optimized in advance — not on hardware but – with appropriate simulations tools — by using software. This offers the developer several weeks' gain in time for each optimization run they can perform with software tools before doing the experiments. Although, the final measurements for the verify the device characteristics and reliability cannot be avoided, the monetary benefits due to time savings can be considerable.

4.1.2 Industrial Requirements

The requirements for optimization, especially for industrial applications, are fast and easy setup procedures, speedy and accurate simulation results, and, in addition, one of the most challenging tasks: cross-checking mechanisms providing the user with information on whether there are any mistakes within his applied constraints, and within the chosen command sequences, or within the optimization setup.

Up to a certain level, a graphical user interface (GUI) can provide such features. Another fact that has to be considered is that some of the assumed (commonly used) settings made by the GUI may not be always wanted by experienced users. These well experienced users often prefer a text-based input deck to specify their real needs. For that particular use case, the GUI does not help the user very much. Therefore, in the concept phase of the GUI, the decision has been made in favor for well trained technicians and experts and for people which are not so experienced. Hence, the concept covers both categories of users.

However, an important feature requested from many design engineers is to have a pool of ready-to-start templates which offer a very expedient introduction into the optimization methods as well as a rapid initial setup of their optimization tasks, even for very complicated configurations. Gradually, the original templates are often adapted in order to meet more specific and more complex demands. The big benefit in that use case is to have a very fast set-up of the optimization run and a very steep learning characteristics of the design engineers with only a few minutes for the introductory and the set-up task.

Thus, the design decision at the Institute for Microlelectronics has been made in favor for experts with a large pool of such templates and example files. Since most of the simulators have to be adjusted by experts for the appropriate technology and IT equipment, additional installation of templates and examples would not be too time consuming and is therefore not an issue. With these features and a plenty of templates and examples files, the software is able to provide sufficient information and configuration possibilities for experts as well as ready-to-start templates for propaedeutic examples to show and teach the major features. Most of the time, even experts use for simulation and optimization mainly well established templates which have been once created or adapted for a certain class of problems.

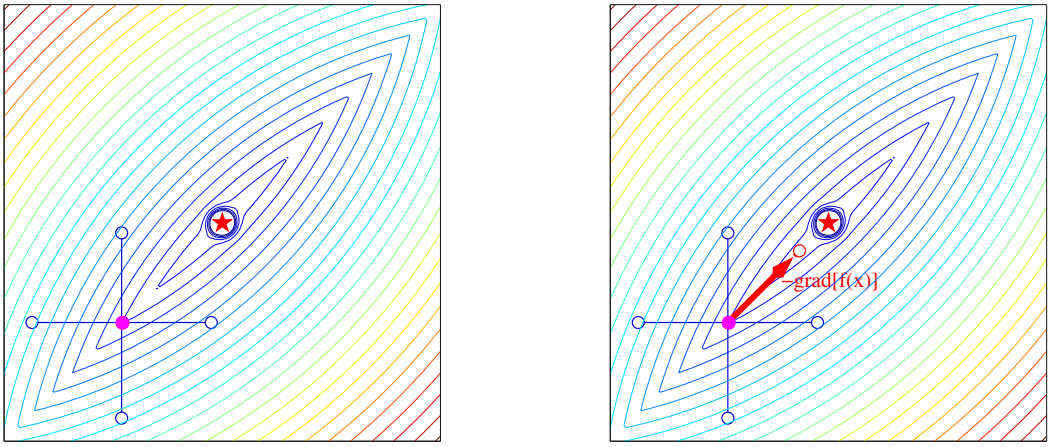
As depicted in Figure 4.2 the optimization loop can be involved at several levels of the design process. [234, 235]. The first and fastest optimization loop is the one depicted on the left hand side of Figure 4.2. It shows the optimization with respect to the parameters of the models used in the process and device simulations. At the end of this optimization loop, the simulation results are compared with the reference data which can be certain figure of merits like guidelines or constraints or can be data obtained from measurements of real devices. The second optimization loop considers the fabrication process as well, where the process specification and the fabrication receipts are changed as the optimization procedure suggests by providing intermediate optimization results. However, design engineers have to verify the intermediate optimization results as well as final results because the obtained data set are numerical the best, but might be physically not reasonable, if for instance certain constraints cannot handle or determine the capabilities of the fabrication processes.

4.2 Optimization Strategies

This section discusses important optimization strategies commonly used with respect to the application to problems in TCAD.

4.2.1 Coordinate Search Algorithm

This algorithm starts with an initial guess in the parameter space, at which the score function is evaluated. Based on this location, a small environmental search is started to obtain a better score than for the initial guess. Within this method, the search algorithm selects one point in each parameter axis to check for a better score value (cf. Figure 4.3a). The corresponding search distance from the current base location can be defined in the configuration section of the optimizer. If a better score value is found at a certain point during the local environmental search, this point is selected as the base location for the next environmental search. If no better results can be found in the given environment around the current base location, the search distance for the algorithm is reduced, by a user-defined factor. A common value for this reduction is one half.



(a) Coordinate Search (b) Gradient-based Search

Figure 4.3: Virtual result space of the score function [236] illustrating the search strategy of the coordinate search mechanism and a gradient-based algorithm.

4.2.2 Gradient-based Optimization

In contrast to the coordinate search algorithm depicted in Section 4.2.1 where the information about the given problem is limited to score values of sample points of the parameter space, gradient-based optimization methods offer the advantage to construct additional information about the shape of the surface for the particular problem. Hence, the gradient of a function provides information about the behavior of a function such as steepness and extrema in the parameter space. With this additional information, the convergence of the search algorithm can be drastically enhanced.

However, information about the gradient is often not available. Therefore, the algorithm has to provide a procedure to ensure the evaluation of a gradient by suggesting additional points for the computation of the gradient in a certain point, as it is shown in Figure 4.3b. In the case that the evaluation of the point resulting from the gradient provides no improvement with respect to the score function, the step length in the direction of the gradient is reduced by a user-defined factor. If this measure also provides no additional improvements, the algorithm has reached its termination criterion and the algorithm stops and provides the best result as the optimization result.

NEWTON Algorithm for Optimization

The NEWTON iteration scheme offers a simple but very powerful root-search algorithm with a quadratic convergence property, if the initial guess is located within a confidence interval (domain) near the root. Hence the NEWTON iteration scheme provides a local root search for a function $g(\mathbf{x})$ and a local optimization strategy for $g(\mathbf{x})$ if this function represents a spatial derivative of an objective function

$$\mathbf{x}_{n+1} - \mathbf{x}_n = \left[\nabla g(\mathbf{x}_n)\right]^{-1} \cdot g(\mathbf{x}_n), \qquad n \geq 0. \tag{4.17}$$

The NEWTON iteration method can be applied for the optimization problem (4.1), where the minimum/maximum search of the score function f_{Score} is equivalent to the root search of the spatial derivative of the score function if the function $g = \nabla f_{\text{Score}}$ substituted to (4.17). Hence, the optimization as a root search reads

$$\mathbf{x}_{n+1} - \mathbf{x}_n = \left[\nabla \left(\nabla f_{\text{Score}}(\mathbf{x}_n)\right)\right]^{-1} \cdot \nabla f_{\text{Score}}(\mathbf{x}_n), \qquad n \geq 0. \tag{4.18}$$

In (4.17) and (4.18) $\mathbf{x}_n|_{n=0}$ represents the used-defined initial value, $\mathbf{x}_n$ is the current value of the NEWTON iteration, and $\mathbf{x}_{n+1}$ is the next value which is improved by the local derivative for the objective function. An additional disadvantage of this method is the requirement of the second derivative of the objective function. If higher dimensional parameter spaces have to be considered, the computational effort can be enormous compared to the nominal number of parameter evaluations. To improve the convergence property the NEWTON optimization algorithm line searches are used. If the curvature of the function is positive, the optimum found is a local minimum, and otherwise a maximum.

Response Surface Method

The response surface method (RSM) has been introduced by Box and Wilson [237] and approximates the original problem (4.1) by a surrogate problem $f_{\text{Surr}}(\mathbf{x})$

$$f_{\text{Score}}(\mathbf{x}) \approx f_{\text{Surr}}(\mathbf{x}), \tag{4.19}$$

$$f_{\text{Surr}}(\mathbf{x}) \longrightarrow \min, \tag{4.20}$$

where the surrogate function f_{Surr} estimates the original function f_{Score}. The surrogate function is often expressed by an analytical function in $\mathbf{x}$ and sample points of the original function f_{Score}

$$f_{\text{Surr}}(\mathbf{x}) = f_{\text{approx}}\big(\mathbf{x}, f_{\text{Score}}(\mathbf{x}_1), f_{\text{Score}}(\mathbf{x}_2), f_{\text{Score}}(\mathbf{x}_3), \ldots, f_{\text{Score}}(\mathbf{x}_n)\big). \tag{4.21}$$

A common method is to use an n-dimensional polynomial function, where the evaluated parameter sets provide base points for interpolation

$$f_{\text{Surr}}(\mathbf{x}) = f_{\text{approx}}\big(\mathbf{x}, \mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \ldots, \mathbf{y}_n\big), \tag{4.22}$$

with the values of the evaluated score functions denoted as $\mathbf{y}_i = f_{\text{Score}}(\mathbf{x}_i)$. This RSM method can be applied if the evaluation of the original function is rather expensive in terms of the computational effort as well as with respect to the computational resources needed for the evaluation of the original problem.

The main goal of this method is to describe the original function with as few sample points as possible to minimize the computational effort for evaluating the original function and to speed up the optimization by optimizing a surrogate function.

As a drawback, the more sample points are required for the optimization problem the more expensive is the optimization of the surrogate problem because each new sample point demands a new calculation of the coefficients of the surrogate problem which entails a high computational burden and a huge memory consumption. Nevertheless, if the computational effort for the evaluation of the original problem is huge and the numerical evaluation is very time consuming, the huge resource demands for the surrogate problem is still economically justifiable and reasonable in terms of the computational and optimization time.

Despite of these drawbacks, this method is commonly used in a wide range of applications [238–241] where large numbers of numerical evaluations of a certain original problem are economically not reasonable within an optimization run.

LEVENBERG-MARQUARDT Algorithm

The LEVENBERG-MARQUARDT algorithm [242] is an efficient method to solve non-linear least squares problems [243]. Thus, it is well suited for complex inverse modeling tasks especially for TCAD applications where the aim of the LEVENBERG-MARQUARDT algorithm is to optimize (minimize) a twice differentiable function

$$f(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}. \tag{4.23}$$

If the original objective function is vector valued, an additional norm has to applied to map the vector to a scalar-valued quantity. The second derivative of the function f is determined by its HESSian[6] matrix. Because the optimization tasks for TCAD problems cannot be described analytically, the derivatives have to be calculated for each single point. Since there is no guarantee that the HESSian $H(f, \mathbf{x})$ is positive definite for non-quadratic forms, the search algorithm might search in the wrong direction. Therefore, a correction term can be introduced to cover this problem by [242]

$$H^k(\mathbf{x}^k) = H(f, \mathbf{x}^k) + \nu^k \tilde{I}. \tag{4.24}$$

If $H^k(\mathbf{x}^k)$ is still not positive definite, the factor ν^k is increased by a certain user-defined factor. Since $H^k(\mathbf{x}^k)$ is now per definitionem positive definite, the next point $\mathbf{x}^{k+1}$ can be calculated by

$$\mathbf{x}^{k+1} = \mathbf{x}^k - H^k(\mathbf{x}^k) \cdot \nabla f(\mathbf{x}^k). \tag{4.25}$$

However, if there is no improvement in the last minimization step ($f(\mathbf{x}^{k+1}) > f(\mathbf{x}^k)$), the factor ν^k has to be modified again and the previously described steps have to be recalculated.

This method is a more robust method than the GAUSS-NEWTON method [244] and provides in general an optimum on less iterations. Nevertheless, if the initial guess of $\mathbf{x}$ is too close to the optimal value, the convergence might be slower than that of the GAUSS-NEWTON method.

[6] Ludwig Otto Hesse (22 April 1811 – 4 August 1874)

Implementations of Gradient-Based Optimization Methods

- The optimizer LMMIN is a local minimizer of vector-valued quantities and is based on the LEVENBERG-MARQUARDT algorithm [242] from the MINPACK project [245, 246].
- The software tool DONOPT is a local optimizer. It minimizes scalar-valued quantities and supports equality and inequality constraints [44, 247, 248]. This optimizer is based on the donlp2 algorithm [249, 250].

Other Gradient-Based Methods

Beside these optimization methods some others have been proven to be well suited for device simulation purposes and have therefore also qualified for optimization in general.

- The "method of steepest descent" [251, 252] is a iterative optimization method and uses the negative gradient of the function as search direction and combine that with a line search algorithm. However, when the condition number of the system matrix is large, the convergence speed is drastically reduced.
- The "conjugate gradient" (CG) algorithm [219, 253] is used to optimize for instance the matrix function $\Phi(\mathbf{x}) = \mathbf{x}^T\mathbf{b} - \frac{1}{2}\mathbf{x}^T A\,\mathbf{x}$, which is equivalent to solving $A\mathbf{x} = \mathbf{b}$, where A is a symmetric, positive definite system matrix. Other CG variants are the"Biconjugate gradient" method [254,255] and "conjugate gradient squared" method [256] were introduced to deal with not symmetric or even with non-positive definite matrices A.

4.2.3 DIRECT Search Optimization

The DIRECT search optimization method [257, 258] uses a primitive search algorithm, where the parameter space is consequently searched by tracking the path of best results with respect to the evaluations of the score function [236]. Thus, this strategy is a global optimization strategy for LIPSCHITZ[7] continuous functions where the existence of derivatives is not required.

4.2.4 Genetic Optimization

The genetic optimization algorithm uses a similar approach to the storage of genetic information in DNA (deoxyribonucleic acid) [259–261]. The retrieval and transfer of genetic information from the parents to the children can be modeled for optimization purposes introducing mathematical operators which an equivalent to the natural mutation of the DNA and are equivalent to the inheritance of properties by constructing a new DNA for a child out of the DNA information from the parents. The DNA consists of a certain amount of chromosomes which are represented by a set of free parameters. The sets of free parameters are also called designs. A population consists of a certain number of individuals (designs). According to their fitness function[8], the individuals remain alive or are discarded in favor of new individuals.

The designs can be altered according to the different operators mutation, inheritance, crossover, and selection. The mutation operator changes a small number of parameters with in one design

[7] Rudolf Otto Sigismund Lipschitz (May 14, 1832 – October 7, 1903)

[8] A fitness function is the negative of the score function. A minimization of a score function is equivalent to the maximization of its fitness function. [219]

to search for an improvement in the near neighborhood in terms of genetic information. Good parent DNA can be inherited and thus the information of a "good" design can be kept for the next iteration. The crossover operator uses the DNA information of two parent designs to form a new design. A crossover can be performed using one-point, two-point, or multiple-point crossover operations. The best individuals of a particular population survive analog to nature. On the basis of these best individuals, the genetic operations are performed. If the number of maximal designs has not been reached, the remaining designs will be initialized by randomly selected parameter sets.

Each of the genetic operations are applied with a certain probability which can be configured at the optimization set-up. According to the different implementations of the genetic algorithms, some operators can be given priority or even switched off. In the worst case the genetic optimizer degenerates to a random generator only.

The termination criteria for the genetic optimization can be classified into three categories: the maximum number of population evaluations has been reached, the maximum of computational time has been exceeded, or an interruption of the user has been initiated. However, the result of a genetic optimization is a certain number (typically 10) of the best designs (individuals) of all evaluated populations.

Extensible Genetic Optimizer (EGO)

The extensible genetic optimizer (EGO) is a state-of-the-art evolutionary computation optimizer [243]. This optimizer is based on a genetic algorithm which has been developed especially for TCAD demands, where computationally expensive score functions have to be evaluated. The optimizer EGO provides a GAUSSian mutation operator, which changes for instance $x \in [a, b]$ to $\min(\max(N(x, \sigma), a), b)$, where $N(x, \sigma)$ is a GAUSSian distribution function and the standard deviation σ depends on the interval length. The crossover operators available in EGO are the linear randomized crossover, the two-point crossover, and the uniform crossover operators. Constraints can be considered as penalty terms in the score function, which usually works not very well due to the reduced convergence property.

Genetic Optimizer (GENOPT)

The optimizer GENOPT is a genetic optimizer which uses a multi-threaded C++ interface of the "Genetic Algorithm Library" (GALIB) [262] with some extensions for multi-threaded operations. GENOPT provides a subset of the capabilities of the GALIB such as the standard selection, mutation and crossover operators as well as scaling and termination methods [243, 261, 263]. However, the convergence of this genetic optimizer is quite slow and should be used only if the other optimization strategies fail.

4.2.5 Evolutionary Optimization

In contrast to genetic algorithms, the evolutionary algorithms describe the social behavior and the change of the behavior of a certain population with respect to different environmental conditions. Hence, the population stays the same but the properties change.

Particle Swarm Approach

Particle swarm optimization (PSO) techniques is a population-based stochastic optimization algorithm which has been developed at Purdue, Indiana in 1995 [264–266]. This optimization technique uses genetic approaches as well as heuristic ones to describe and map social behavior of for instance bird flocking and fish schooling. These kinds of behavior are used for optimization purposes to search for a optimal state. The PSO algorithm uses similar evolutionary techniques for computation as for instance GA. For a PSO run, the system of interest is initialized by a set of initial states (initial guess). As a difference to the pure genetic algorithm, PSO has no genetic operators to build its child-parameter sets as for instance crossover and mutation operators do. Where the genetic algorithm provides the information of all individuals for the genetic operators, PSO provides this type of information only for a selected subset of the best individuals. Nevertheless, the PSO algorithm describes the behavior of an individual according to the best individuals which are determined by a certain fitness function. Contrary to the mapping of the evolutionary algorithm where life and populations are described, the PSO tries to characterize the social behavior of a population and the collective behavior of each individual of the population [267–269]. For instance, a flock or swarm of birds that is searching randomly for food in a certain area organize themselves to obtain a good strategy to find food for each individual as fast as possible. The obvious best answer to that question is to follow that individual which is nearest to food. Thus, with each iteration, the solution of each individual becomes more optimal because each individual is nearer to the food as before [270].

The PSO algorithm can learn from these types of scenaria and use them to find optimal solutions for other tasks. In this case, the individual is called 'bird' and represents a single solution for the optimization problem. Each individual consists of a certain amount of different characteristic parameters (genomes and properties). These individuals can be assumed to be members of a certain population of the optimization algorithm and hence called particle of a swarm. More abstract synonyms are parameter set, design, or solution.

The PSO method learns from the different applied scenaria and uses them to solve the current optimization problem. In PSO, each solution ("bird") is a single point in the parameter space. Each particular bird has a fitness value or score value which is evaluated to optimize the quality of the current solution or how far the bird is away from the food. A characteristic parameter of such a bird could be the position and velocity which directs the trajectory of the bird (particle) towards the food (optimum). Thus, the particles fly through the parameter space by following the particles which have the best solutions.

The PSO algorithm is initialized with a set of randomly chosen solutions. Then, the PSO method searches for optimal parameters by updating the current particle generation by using evolutionary and social algorithms. In every iteration, each particle of the population is updated using two different "best" values: `gbest` and `pbest`. The first "best" value `gbest` is the globally best solution which has been achieved so far during the entire optimization. The second one, `pbest`, is the best solution within a certain environment of the current particle. The environment can be specified by a distance or by a topological description. Hence, `pbest` represents the current local optimum in the current local environment of the particle, where `gbest` is the current global optimum of the optimization task. After determining the two characteristic "best" values of the population, the particle with its current occurrence $\mathbf{x}$ is updated using

$$\mathbf{x} = \mathbf{x} + \mathbf{u} \tag{4.26}$$

$$\mathbf{u} = \mathbf{u} + C_1 X_1 (\texttt{pbest} - \mathbf{x}) + C_2 X_2 (\texttt{gbest} - \mathbf{x}), \tag{4.27}$$

where $\mathbf{u}$ is the update vector of the current individual. The quantities $X_1 \in [0, 1]$ and $X_2 \in [0, 1]$

are two statistically independent random variables which introduce the heuristic part in this optimization method. On the other hand, the learning coefficients C_1 and C_2 are used to specify if the current optimization strategy finds a local rather than a global optimum if $C_1 \gg C_2$. Common values for the learning coefficients C_1 and C_2 are $C_1 = C_2 = 2$ for a balanced learning behavior, where this particular constellation of the learning coefficient preserves on the one hand a rather fast local optimization with some respect to the global optimum and on the other hand, the global optimum will contribute a significant part to the update in order that the final solution is also a global optimizer.

4.2.6 Simulated Annealing Approach

This optimization follows the observed natural law that thermal annealing (rapid heating and slow cooling) yields more regular crystal structures. This method has been first proposed by METROPOLIS[9] in 1953 [271] and has been constantly improved [243, 272–277]. This method is commonly used in semiconductor technology nodes to activate the implanted doping profiles, where the impurities are set to lattice sites of the substrate material. The optimization method uses a similar approach to overcome a trapped situation in local minima. In the hot state, if a local minimum is found, the algorithm searches also in regions different to the local minimizer. This search radius is reduced in each iteration according to the falling temperature, similar to thermal diffusion and BROWNian motion, described by a function proportional to the BOLTZMANN factor $\exp\left(-\mathcal{E}_\mathrm{A}/k_\mathrm{B}T\right)$. At each iteration of the optimizer, the temperature is decreased like a natural cooling process. Hence, the probability that the algorithm searches in different locations goes asymptotically to zero.

Different implementations of simulated annealing such as "adaptive simulated annealing" [275, 276] and "very fast simulated annealing" [243] use different models to described the decreasing temperature. Both implementations use an exponential temperature decay in their models. Advanced versions of the simulated annealing approach also use local optimization techniques and thus optimize the local optima using the exponential temperature decay in the BOLTZMANN factor. Such optimization algorithm start with an initial guess which is locally optimized for instance with a implementation of the LEVENBERG-MARQUARDT optimization algorithm [243, 276]. If a local optimum is found the algorithm jumps to a different point with a certain probability determined by a factor proportional to the BOLTZMANN factor and starts another local optimization with the hope of finding a better local optimum as compared to the existing one.

4.3 State-of-the-art in Optimization

For systems where partially analytical descriptions are available, the determination of local optima often easy to obtain. However, the problem is still hard and may require a huge computational effort and the computational costs to find a global extrema is still NP-hard. Yet, for industrially relevant applications the complexity increases and the number of suitable optimizers, which can be applied to these types of problems decreases rapidly. Because the optimization setup requires knowledge of the optimization process itself as well as an expert knowledge of the methods used in the simulation tools in the optimization process.

Many discussed optimization strategies have prerequisites to guarantee an effective operation of the optimization algorithm. However, most of these requirements cannot be fulfilled from

[9] Nicholas Constantine Metropolis (June 11, 1915 – October 17, 1999)

optimization problems of industrial interest. Therefore, the user has to decide for some approximations to optimize a certain aspect of this particular problem. Nevertheless, this method often yields reasonable intermediate results which can be used for further improvement using different optimization strategies. Hence, the whole optimization task often becomes a stepwise optimization strategy with different applied optimization techniques.

Device Purposes

The optimization of a semiconductor device often requires to perform calibrations and parameter extractions for certain models where reference data is available either through measurements or Monte Carlo (MC) simulations [278, 279]. In general, the optimization for device simulation purposes requires to modify certain model parameters to match the simulated characteristics with reference data. This fact offers the possibility to tune certain local optimizer tools to operate at vectorized target quantities as for instance the LEVENBERG-MARQUARDT algorithms does using an norm to transform the vector-valued quantity to a scalar-valued one. As an alternative the score function can be extended by a p-norm operator to transform a vector-valued quantity into a representative scalar-valued quantity. However, with increasing numbers of model parameter to tune, the optimization task becomes more complex in terms of the parameter dimensions and the number of required parameter evaluations can often demand a tremendous effort if the evaluation of the parameter is computationally expensive. Nevertheless, gradient-based optimization algorithms and heuristic approaches like simulated annealing serve well for a wide range of problem classes and deliver reasonable and accurate results [280, 281].

However, optimization in which the ranges of the parameters differ by orders of magnitude face other challenges. If certain parameters in the range [0,1] have to be considered as well as doping concentration in the range of $[10^{14}, 10^{19}]$, the numerical precision is often not sufficient to compute the parameter correctly. Therefore, the latter parameter can be manually transformed either to $x\,10^{14}$ with $x \in [1, 10^5]$ or to $10^{(14+x)}$ with $x \in [0, 5]$. This particular problem can often be solved by a rigorous scaling of the optimization software. However, each numerical challenge requires a program to solve the problem with respect to the underlying simulation software because the program has to clarify if this simple parameter transformation can be applied to a simulation tool. If not, an additional program is required to produce the appropriate input format for the simulator.

Technology Purposes

Typical tasks for process and technology simulation are to describe an etching process, material deposition, chemo-mechanical polishing, ion implantation, lithography as well as electro-thermal and mechanical investigations of the derived structures. These process steps deal with the patterning and structuring of matter, which requires highly complex models to describe the physical properties as well as the chemical and physical interaction between the materials. However, since the temperature cannot be neglected in many process steps, the model becomes even more complex. Hence, optimization of such problems requires deep knowledge of the important ongoing process inside the current simulation task to estimated the possible bias points of operation. These points can be used as initial values for optimization [282]. If, for instance, an appropriate approximation can be found for a certain bias point, a specially tuned (local) optimization strategy can be applied. For the other cases, heuristic, genetic, or random search methods have to be used to obtain a certain "feeling" of the current problem class.

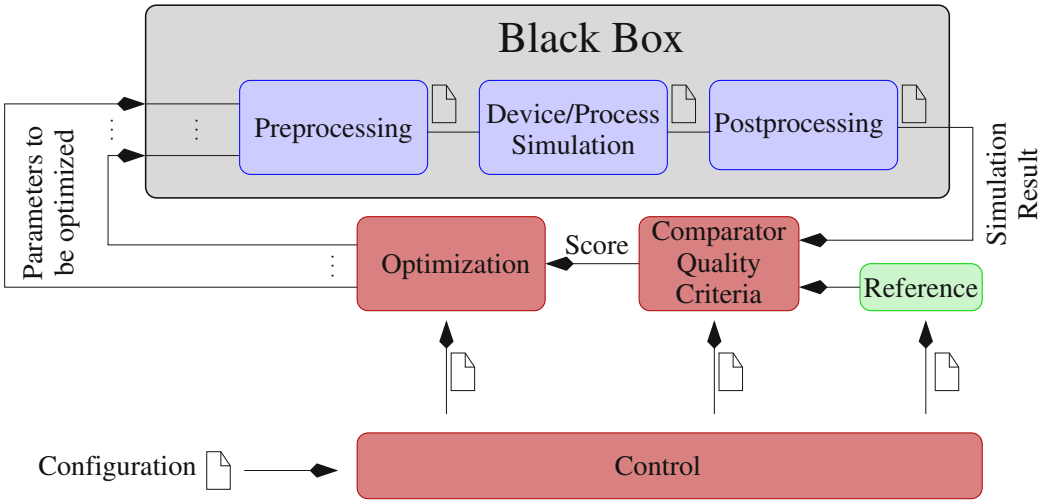


Figure 4.4: Data flow for the optimization mode in SIESTA. The base structure is shown with the external simulation sequence depicted as interchangeable black box.

Black Box Approach

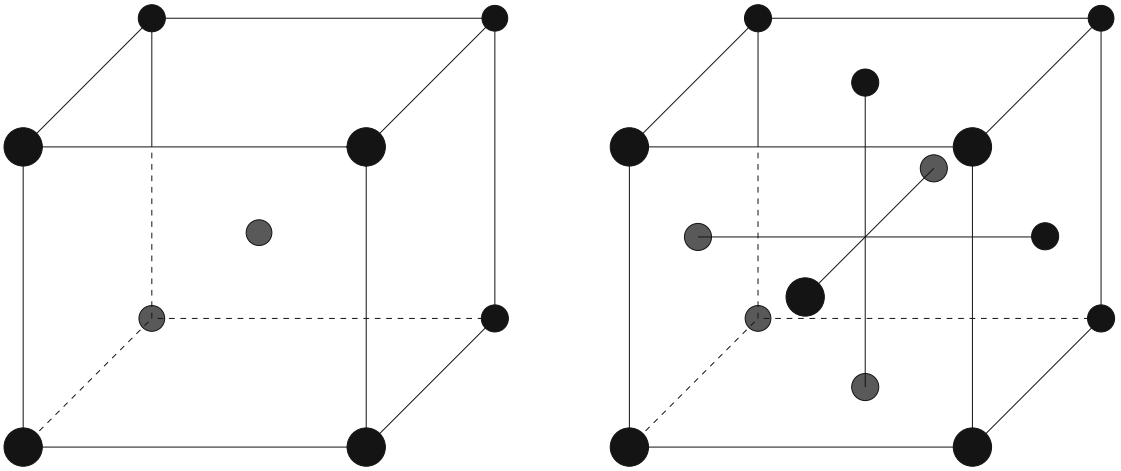
In TCAD applications, most of the problems cannot be formulated using one single set of equations. Due to the steadily increasing complexity of the material models and the physical effects which have to be considered, the behavior might change if a certain quantity has exceeded a certain value. For instance if the temperature has reached the melting point of Al, the conductivity model looses its validity or if the electric field exceeds a certain threshold, a breakdown of the dielectric occurs. Both scenaria requires a switch of the models in the simulator. Therefore, optimization methods which use information about the shape and gradients are very difficult to apply due to the uncertainty of hidden constraints and due to suboptimal discretizations. As an alternative, the objective function, which includes the simulation as well as the score function, can be assumed to be a black box function where only the values of the function at a certain point are known. For this case the standard optimization methods have to calculate the gradient separately which is very time-consuming if the parameter space is of a high dimension. Thus, a gradient-free optimization strategy would be appropriate for this case.

In-situ Approach

Another way to perform optimization tasks is to include the primary optimization loop into or closely to the solving algorithm of the simulator. This is very efficient in terms of computation time [283].

To reduce the time for the lookup of a particular parameter constellation, the optimization can be performed in advance and saved in a database. For each inquiry for different design patterns the database can be consulted instead of many very similar optimization runs (cf. Section 4.3.1).

Within applications where the type of the possible solutions are well known, the optimization can be performed within the iteration loop of the solving algorithm in the simulator by introducing an additional internal optimization loop. However, this is only possible if the source code of the particular software is available and if the source code is well understood to preserve the accuracy and the function of the simulator.



(a) Central Composite Circumscribed (CCC) (b) Central Composite Face-Centered (CFC)

Figure 4.5: Two propaedeutic example designs with three input parameters. The central composite circumscribed (CCC) design is shown in (a) and the central composite face-centered (CFC) design in (b).

A different approach is to extract the problem description from the simulator and reimplement it together with the internal optimization loop as it has been introduced for a doping profile optimization for a diode [283] which has yielded excellent results in terms of accuracy and computation time. Unfortunately, this implementation is only suitable for a certain class of problems. For more general applications different programs have to be developed in order to treat different problem classes.

However, this type of optimization yields very accurate optima and very fast optimization results because the optimization algorithm has all necessary information for its calculation inside the simulator and the optimization can be obtained within one simulation step. Ordinary optimizers would have to iterate several times and therefore to communicate with other software tools via an external interface which is conveniently an API or the file system. These communication procedures would require additional time and would provide only a limited bandwidth to transfer data while the in-situ optimizer does not need to transfer data because everything is pre-calculated by the simulator and the optimizer can access that data because the optimizer is part of the simulator.

4.3.1 Design of Experiments

As a result of a rigorously investigated problem nearly all parameter combinations have to be evaluated for the worst case to determine and verify the global optimum of a particular problem. Since this procedure to obtain the necessary information is very time consuming a very effective design of experiment (DOE) method has been introduced in the industry.

In this method a DOE tool proposes sets of parameters which cover the complete discretized parameters space of a certain problem. These collected results of the parameter sets can be filed, for instance in a data base, and further data manipulation routines can be applied on this data. A common use is to provide this type of data for design environments to check if the desired arrangements of parameters from the user results in reasonable and manufacturable devices.

Table 4.1: Common transformations for designs (cf. for instance [40, 44]).

Parameter	Description
`default`	equidistant
`linear`	$a\,x + b$
`logsimple`	$\log(x)$
`logarithmic`	$a\,\log(b\,x)$
`function1`	$A\,\log\left(\frac{x}{B}\right)$
`function2`	$A\,\frac{x-B}{x}$
`function3`	$B\,\exp\left[A\,\frac{x-C}{C} - 1\right]$
`function4`	$B\,\left(1 - \exp\left[A\,\frac{C-x}{C}\right]\right)$

However, the computation effort cannot be reduced because the parameter evaluations have to be performed to obtain the optimum. Adding additional parameter evaluations and storing them to a database enables the use of the information obtained from the optimization. So, the engineers can use the information obtained from previously run experiments and save a lot of time and computational resources.

There are several methods to distribute parameter sets (designs, experiments) into the parameter space [44, 284]. One possible way is to select the parameters randomly around a certain expected mean value. The full factorial design divides each parameter axis in m sections which results in m^n different designs to evaluate, where n is the dimension of the input parameter domain (hypercube). The sectioning of the hypercube can be chosen according to the featured transformation of the tools used in the DOE. Common transformation are given in Tab. 4.1 where the principle behavior of the parameters can be accounted for. For instance, typical doping concentrations may vary by several orders of magnitudes. Hence, a logarithmic transformation is usually more suitable than a equidistant linear distribution of the designs.

4.4 Challenges in Optimization

As it has been outlined in the last sections, each of the different optimization strategies has its own benefits. Hence, the best way to obtain a good and fast optimization result is to know the problem analytically which would allow to answer the question about its optimum most efficiently.

To account for the complex problem classes in optimization of problems in TCAD, new global optimization strategies have been introduced together with appropriate systems of constraint functions. However, with increasing the number of constraints and constraint functions for the simulation and optimization, the optimization becomes more complex and therefore more computationally expensive.

4.4.1 Constraints

According to chosen constraints, a suitable appropriate optimization strategy has to be chosen, where the desired constraints can be described mathematically. Mostly, input parameters are bound and have thus a lower and an upper bound. These bounds can be even more restrictive if constraint functions are applied which limit the allowed set of parameters in the hyper-cube of the bound parameter space.

A-priori constraints can be checked in advance of the simulation run thus before the time is wasted for a parameter set which is known as infeasible. However, certain constraints can be checked only at the end of the simulation. If for instance a constraint function contains a simulation result, e.g. the electrical conductivity, then an apparent constraint is to demand that this quantity remains greater than zero[10]. But without an application interface to the simulator, the optimizer can only verify this quantity at the end. However, a possible way to overcome a probable long waiting period, the function can be approximated in advance of the simulation. This method is actually a response surface approach for a single quantity. Wrong estimates of such *a-priori* approximations may cause significant loss of the search space and may result in failing to find the global optimal value.

Thus, it is important to determine the lower and the upper bounds for the input parameters carefully because too wide parameter ranges can cause a considerable large number of parameter evaluations and too narrow bounds may oversee certain still feasible parameter constellations which might yield to an optimum. Since not all optimization strategies provide an appropriate treatment of constraint functions, many optimization frameworks offer the possibility to define barrier or penalty functions to provide an optimization set up which is valid for several optimization strategies[11].

4.4.2 Selection of Optimization Strategies and Score Functions

An appropriate selection of the score function decreases the iteration necessary for the optimization and often increases the quality of the optimization result. However, the construction of the score function with the selections of the quality criteria is even more complex than the simulation itself and thus often compared to an art because visually obvious selection criteria have to be described by mathematical equations to select the relevant data out of the simulation results available. The optimal optimization strategy can be described as that one which yields the global optimum within the shortest period of time. However, this assumes that the problem class for the optimization is known, that all constraints can be well formulated, and that the response surface is rather smooth. In realistic applications, these assumptions are often not valid. Most problem descriptions operate at a certain modeling level, where the model is also constructed out of a set of assumptions. In addition, many models check if the operating bias point is still in the region for which the assumption has been set. If not, the model behavior can be switched and the behavior is in general no longer smooth. However, there are certain decision criteria for the selection of optimization strategies according to [285]. If sufficient information about the simulation tools is available, the design engineer may estimate the shape of the response surface with respect to the input parameters. With that information it is possible to select one optimization strategy out of the pool of the available optimizers.

[10] Since optimization is a mathematical method with certain constrains, the result may be physically not feasible if hidden constraints were overseen or neglected due to abstraction or simplifications.

[11] An overview of barrier and penalty functions is given at the beginning of Chapter 4 and in [221–229].

Most optimizers provide at least one termination criterion which is the maximum number of parameter evaluations. However, if the quality cannot be appropriately measured or determined, the best optimization algorithm cannot converge to a good result if the score function produces too much noise, checks or extracts the wrong quantities, or even diverges. The experience over the last decades has shown, that the most commonly used score functions are the EUCLIDian norm (2-norm), the minimum norm, and the maximum norm. All of those score functions converge if the quantity consists of finite real numbers $y = f_{\text{Score}}(\mathbf{x}) \in \mathbb{R}\backslash\{-\infty, +\infty\}$ and if there exists at least one local minimum. However, the choice of the score function determines the speed of convergence.

4.4.3 Convergence

The convergence of the optimization tools are often defined or proven for rather simple assumptions such as that the input parameter space is a convex domain or hyper-cube, that the objective function is a piecewise continuous function, or even that the second derivative of the object function exists[12]. However, these assumptions are rough estimations, which enable the engineer to decide whether a particular optimization algorithm is more suitable than others.

Since most simulation tools cannot provide all the desirable and advantageous properties of an ideal objective function, standard optimization methods are of only limited usability. Moreover, many problems of industrial interest produce no results with certain simulation tools because the simulation tool does not converge or would take too many resources. This might happen when hidden constraints are overseen, for instance if certain input parameters are correctly bound separately, but together they generate constellations that are physically not feasible.

The decision of how well a particular optimization method anticipates the current behavior of the objective score function has to be made by the design engineer who sets up the optimization task for a certain problem and who should know how the simulator behaves.

The convergence speed of standard optimization algorithms is often given by theorems but cannot be guaranteed for a general optimization task in the TCAD environment due to the guarantee that the simulation tool will converge for a finite number of parameter evaluations. However, if the behavior of the simulation tools can be appropriately estimated and a suitable optimization algorithm is applied to this problem, the probability is rather high that the optimization converges within the expected order of the convergence speed.

Optimization frameworks offer the possibility to apply additional limits. For instance, the CPU time for a single parameter evaluation or the total CPU time elapsed during the whole optimization can be limited. Another common method is to limit the number of parameter evaluations and the number of iterations of the optimizer. With these measures, the optimization often yields suboptimal results but it is guaranteed that the optimizer returns either the best result so far, or no result if all parameter evaluations have not been successfully terminated.

A critical point has to be noted here. The change and the manipulation of material boundaries, for instance if geometry parameters are optimized, requires an automatic remeshing. However, many simulation tools are very sensitive to misaligned meshes and require therefore appropriately tuned mesh kernels to provide good results. Therefore, the meshing tools have to be well configured in advance to provide automatic remeshing with the appropriate quality for the needs of the simulation tools [286].

[12] The assumption that the evaluation of every parameter set can be performed in finite time is a hard restrict because the simulation tools cannot guarantee to provide results within a finite number of iterations.

4.4.4 Reasonable Results versus Numerical Optimum

When the optimization is completed, the optimizer returns either the best parameter set (design) or the set of the best designs. Numerically, these designs are the best, however, they have to be checked as to feasibility, or even if they are physically reasonable. Actually, the constraints and barrier function should have guided the optimization algorithm towards finding more reasonable designs.

Numerically reasonable results can often be obtained very quickly. But due to physically constraints, some of these results have to be rejected. In recent times, even physically reasonable designs have to be rejected because they do not fit into the economical and financial constraints. In that cases the constraint functions were not designed appropriately and a redesign of the complete optimization setup is necessary.

4.5 Optimization Framework SIESTA

The simulation framework "Simulation Environment for Semiconductor Technology Analysis" (SIESTA) has been developed at the Institute for Microlelectronics to provide a framework that supports different simulation flows which can be defined and reapplied with different input data. The simulation framework has been extended by optimizers and a load-balancing system [44] to provide distributed computing. The aim of SIESTA is to provide as much flexibility as possible to allow the user to configure and extend the environment for future purposes. The original version was implemented in VLISP[13] [289]. However, due to user requirements, the core of SIESTA was rewritten in PYTHON [290] now known as SEILIB [291]. SIESTA is now based on several modules to provide a better modularity. The base concept of SIESTA is shown in Figure 4.6, where the different modules are listed. SIESTA consists of the SEILIB core optimizer and the design of experiment module. These three modules are able to work together independently of the other SIESTA modules. Possible scenaria for this type of application are optimizations or a sequences of simulations for instance for design of experiments which do not require user interaction.

SIESTA provides with the SEILIB module facilities to communicate with spy-daemons on the remote computers to collect information for distributed computing and load balancing. A typical task for the SEILIB module consists of a configuration file where all data for the automatic and non-interactive mode is stored including a description of the simulation flow.

The optimization framework SIESTA provides facilities to setup an experiment either graphically or with text-based configuration files, to display the simulation results via an external viewer, and to export simulation results to files. For the setup of an experiment, the simulation flow can be constructed from scratch or it can be obtained from a template, which can be edited in the graphical configuration file editor for SIESTA (cf. Figure 4.7). The graphical user interface (GUI) is capable of performing a rudimentary consistency check. It verifies the type of quantities of the connection between input and output ports of the models. If these two types do not match the connection is refused in the GUI.

Figure 4.7 shows an example where a doping concentration (`ChD`) has to be optimized. This quantity is provided to the first simulation flow model which is the tool MKDEV [37, 292], which constructs from a given template file a device geometry for the device simulator MINIMOS-NT [37] in the next simulation flow model. MINIMOS-NT performs a simulation and provides as result a quantity called `IdIoff`, which is the ratio of "on" and "off" current. This quantity is submitted

[13] VLISP is based on XLISP [287], a dialect of LISP [288] and has been extended by several user-defined functions.

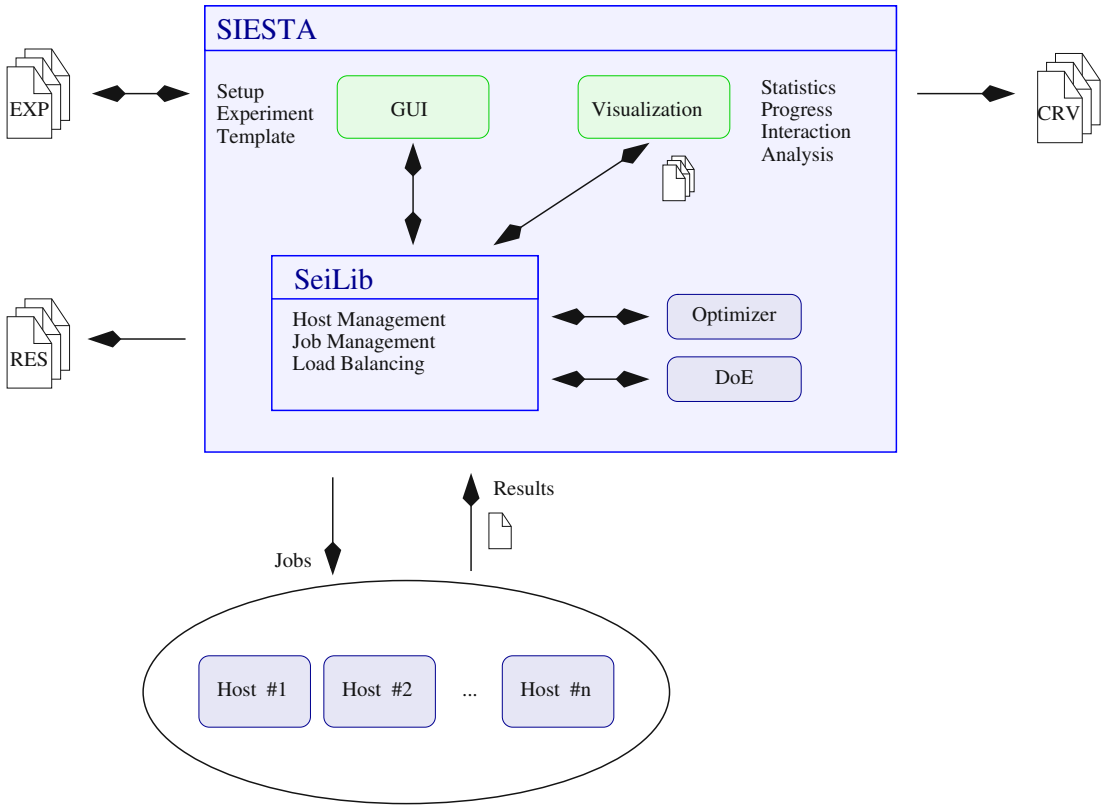


Figure 4.6: Block diagram of SIESTA showing the internal function blocks and its environment.

to the next model which is an auxiliary arithmetic model which performs an arithmetic operation on its input results. However, the output of the auxiliary model is the quantity `error` and is submitted back to the optimizer which proposes according to this value the next `ChD` value to improve (minimize) the `error` value.

The consistency checker is very helpful for introductory examples. However, the consistency check mechanism in the GUI can for instance not verify whether a certain quantity type is a member of a subset of another type. For such specialized setup constructions a text editor provides enough possibilities and is more suitable to build arbitrarily complex optimization setups.

The setup of the simulation flow consists of several template variables in the configuration files which are substituted directly before the particular job is submitted to a computational (remote) node.

Once the simulation job has been submitted to a remote host the job is executed on the remote host and at the end of the simulation job the result files retransmitted to the original directory where the files are either stored or discarded after the required input data has been submitted to the next simulation task.

If the job at the remote host is stalled and not responding, or even if the SEILIB module or SIESTA is terminated, the spies on the remote hosts recognize this fact and terminate the currently running processes which have been started from the current SEILIB instance.

To communicate between the different software components a PYTHON binding of the QT [293] sockets have been chosen which provide fast, simple, and robust communication facilities com-

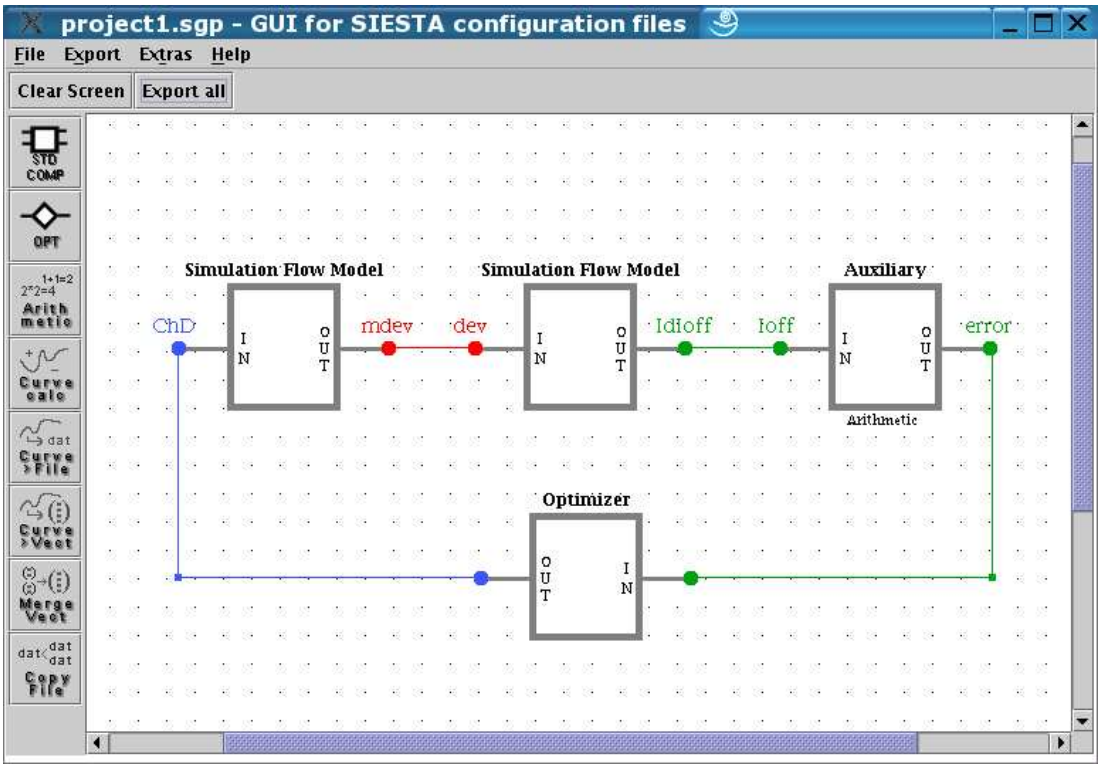


Figure 4.7: Graphical user interface to setup an optimization showing a simple simulation flow which consists of two different simulation tools and a post-processing step.

pared e.g. to CORBA [294, 295]. CORBA would provide direct access to a distributed object which is actually not really necessary. The QT messaging system is more suitable and requires less hardware resources and produces also less network traffic during data transmission. Moreover, the secure shell (SSH) connection provides a tunnel for the QT socket connections from the host on which SEILIB is running to the remote host. Hence, the two software parts communicate with each other by writing into standard I/O streams.

4.5.1 Parallelization Strategy

To shorten the time for a complete optimization run, certain evaluations can be parallelized. For instance, the computation of each single evaluation of a parameter set can be performed within a parallel solver, the evaluations can be executed in parallel, and the optimization algorithm can be parallelized. For simulation tasks which can be executed simultaneously the simulation flow can be split as shown in Figure 4.8. Here, the pre-processing step has to be performed for all following simulation tools. However, the ensuing simulation tools can be started at the same or on different hosts if they have no interdependences as depicted in Figure 4.8. For the case that the simulation flow cannot be split into different, independent parts, the simulation flow represents an individual unit with its own input data. The management of the providing the appropriate input data at the right time and collecting the simulation results from the different hosts are is performed from the SEILIB and other modules from SIESTA. The SEILIB dispatches every job which is ready with its own input data and submits it with the appropriate data to a free (or not fully loaded) remote host.

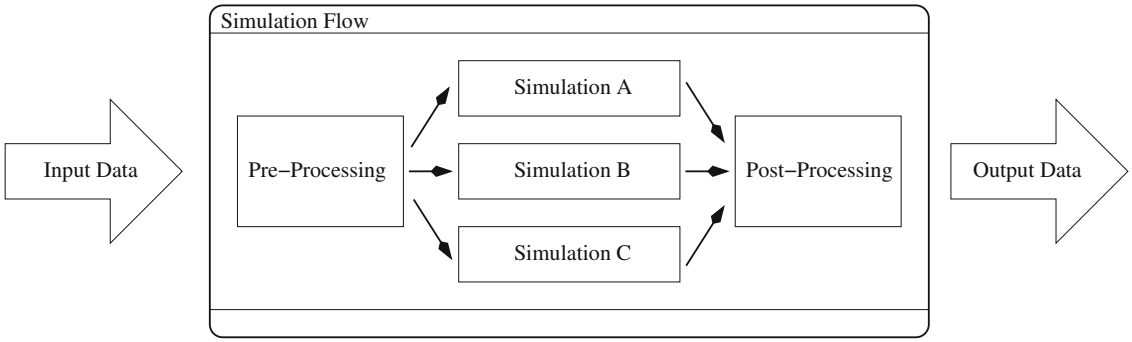
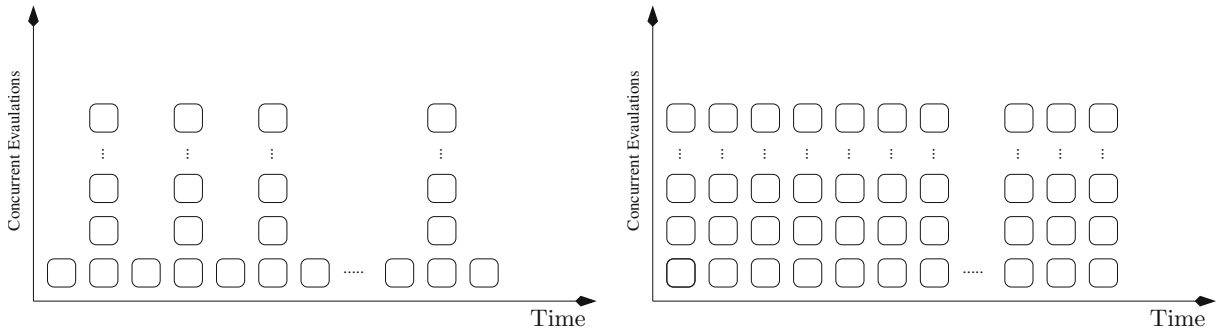


Figure 4.8: Principle block diagram of an arbitrarily structured simulation tool flow.

Figure 4.9 shows two examples of concurrently executed jobs in a particular simulation cluster. Figure 4.9a shows a typical gradient-based optimization strategy where a single parameter evaluation is followed by a certain number concurrent evaluations which can be the results of the numerical calculation of a gradient with respect to the parameter space. Figure 4.9b shows a fully loaded computational cluster. This scenario is typical for genetic and evolutionary optimization algorithms where each population requires the same large amount of parameter evaluations.

4.5.2 Simulation Tools

The simulation tools for TCAD application can be classified into five major categories: process simulation, device simulation, interconnect simulation, geometry manipulation, mesh generation, as well as visualization tools and graphical user interfaces (GUIs). The available simulation tools in the optimization framework SIESTA are listed in Figure 4.10. In this figure also the logical connections between the different simulation tools illustrated. The following will give a brief overview of the available simulation tools.



(a) Parallel execution with sequential test evaluations.

(b) Totally parallel evaluations.

Figure 4.9: Block diagram of parallelization of parameter evaluations.

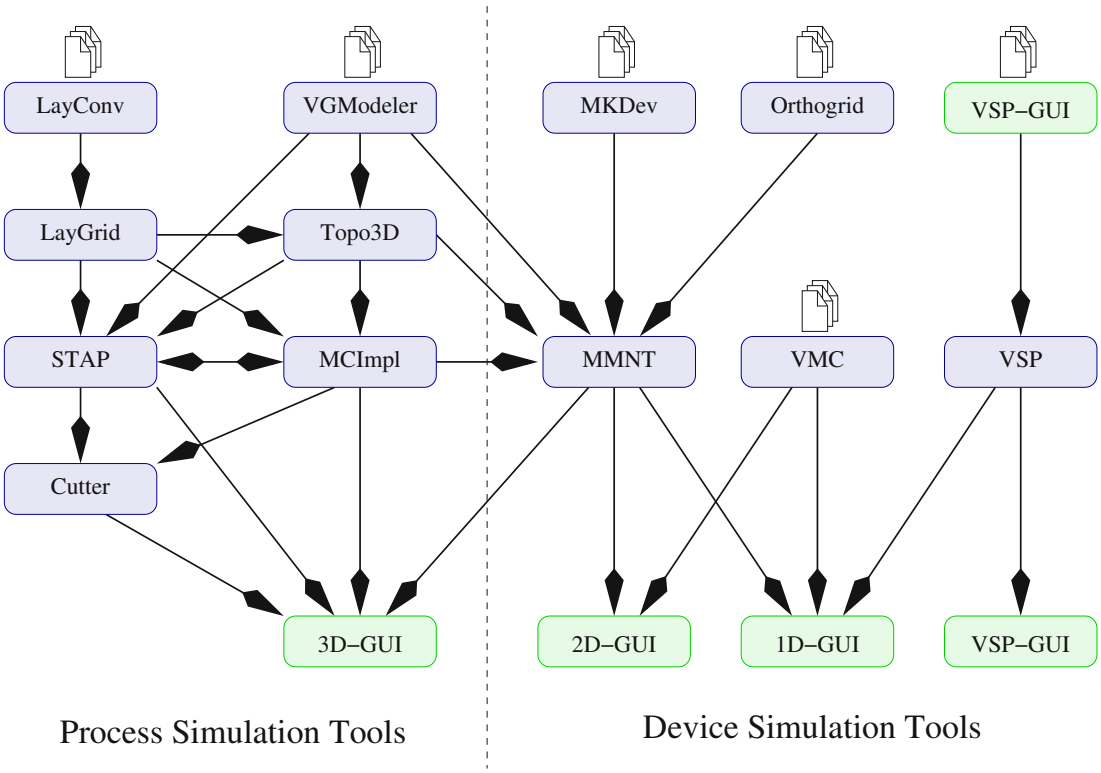


Figure 4.10: Available toolflow showing the variety of TCAD simulation tools provided at the Institute for Microlelectronics.

Process Simulation Tools

Process simulation tools provide the functionality to investigate and depict the different process steps in semiconductor device fabrication. Moreover, this category includes also the extraction of key parameters such as parasitic parameters like resistivity, cross capacitances, inductances, delay time on interconnect lines.

- LayConv [39] has been developed at the Institute for Microlelectronics and provides the capabilities to read several types of mask layers, e.g. CIF and GDS II, and to include them into the process simulation flow. This software additionally offers a proximity approximation feature for mask layers to account for the uncertainty of the optical rays during the development of the photo resist layers.
- Topo3D [296] has been developed at the Institute for Microlelectronics and is a three-dimensional topography simulator that mainly uses the level-set approach for the computational expensive description of the moving surface during etching and deposition steps.
- TSupreme4 [297] is a two-dimensional commercial process simulator from SYNOPSYS which is an advancement of the original simulator code developed at Stanford University [298, 299] and at TMA [300]. This software tool mainly simulates the physical behavior of the dopants in a two-dimensional cut layer through a semiconductor device structure during its fabrication.

- LIGAMENT [40] is a commercial process simulation framework from SYNOPSYS that allows to call external process simulation tools and to apply analytically described process steps, e.g. an implantation of an analytical doping profile.
- SPROCESS [40] is a commercial process simulator from SYNOPSYS that offers self-consistent physical multidimensional modeling for silicon and compound semiconductor process technologies.
- MCIMPL-II [301] is three-dimensional Monte Carlo simulation tool, which has been developed at the Institute for Microlelectronics. In this simulation tool the entire implantation process can be described on a detailed microscopic level.
- STAP [39,92] is a three-dimensional interconnect simulator based on finite elements, which has been developed at the Institute for Microlelectronics. This simulation tool provides key parameter extraction: resistance, capacitance, and inductance.

Geometry Manipulation Tools

The structure generation only through process simulation is often not necessary because for certain problems, the exact geometry description is not required if only intrinsic parameters of the devices are considered. Hence, to save CPU time, the device structures can be built by an analytical description of the solid, for examples by a particular sequence of layers with certain thicknesses or with a graphical editor by drawing the device structure.

- LAYGRID [39] has been developed at the Institute for Microlelectronics and is a three-dimensional device creation tool, which offers the possibilities to create three-dimensional device structures by extrusion of layer defined either by polygons or mask informations.
- VGMODELER [302] is a three-dimensional solid modeling tool which has been developed at the Institute for Microlelectronics. This geometry tool uses descriptions of solids and boolean operations to construct device structures. The resulting geometry description conveniently allows to include slanted walls and fully parameterized configuration descriptions.
- SSE [40] is graphical, three-dimensional solid modeler from SYNOPSYS which allows to include mask layer definitions and to extrude certain layers in the third dimension.
- CUTTER represents a suite of several shell scripts and software tools from several software vendors as well as from the Institute for Microlelectronics. These tools are necessary to cut two-dimensional slices from three-dimensional simulation results to allow a better inspection and an automatic data manipulation for simulation frameworks to automatically extract distributed quantities at certain planes or positions. Most of these scripts are provided with the distribution of the certain simulation tools and can be combined in the optimization environment SIESTA.

Device Simulation Tools

- MINIMOS-NT [37] is a three-dimensional general purpose device simulator developed at the Institute for Microlelectronics for arbitrary device structures using various sophisticated charge carrier transport models. It is capable of performing stationary (AC and DC) and transient simulations as well as to investigate self-heating effects.

- SDEVICE [40] is a three-dimensional multi-purpose device simulator from SYNOPSYS for electrical, thermal, and optical investigation of semiconductor devices.
- VMC [103, 303] has been developed at the Institute for Microlelectronics is a Monte Carlo simulator for stressed semiconductor materials assuming multi-valley bands as well as full band structures. VMC can be applied to binary, ternary, and quaternary semiconductor alloys.
- VSP [304, 305] is a gate stack analyzing program developed at the Institute for Microlelectronics. This tool is based on a coupled SCHRÖDINGER[14]-POISSON equation system for arbitrary geometries with emphasis on reliability.

Geometry and Mesh Generators

- The program MKDEV [37, 292] has been developed at the Institute for Microlelectronics and generates the geometry and the mesh for the device simulator MINIMOS-NT. MKDEV uses template and configuration files to create the geometry including an orthogonal mesh with the specified distributed quantities like doping concentrations using analytical doping profiles.
- CREATEORTHO [37] is a script that creates orthogonal meshes for the device simulator MINIMOS-NT. This tools allows a parameterized mesh generation to create Manhattan structures for device simulation.
- TRIANGLE [306] is a two-dimensional mesh generator for unstructured meshes which uses a DELAUNAY[15] triangulation and a VORONOI[16] tessellation [307–310].
- DELINK [311–313] has been developed at the Institute for Microlelectronics and is a three-dimensional mesh generator for unstructured meshes. This tool requires as input the specified geometry and provides as a result a mesh, which fulfills the DELAUNAY condition [309, 310] and certain other constraints which can be additionally specified.
- SNMESH [40] is suite from SYNOPSYS that provides different meshing algorithms, which are capable of producing tensor product meshes and unstructured tetrahedron meshes according to user-defined constraints. For instance, constraints can be applied with respect to certain parameter values and geometric issues such as certain ranges for angles, aspect ratios, or differences of areas between two or more adjacent elements.

Graphical User Interfaces

Graphical tools offer a convenient method to visualize distributed quantities obtained from simulations. Due to the different vendors and different functionalities of the tools, a various number of graphical user interfaces has been developed.

- 1D-GUIS:

  The VSP-GUI [314] is a graphical user interface for the gate stack analyzer VSP [304]. This GUI has been developed to guide the user through the complete simulation procedure from

[14] Erwin Rudolf Josef Alexander Schrödinger (August 12, 1887 – January 4, 1961)
[15] Boris Nikolajewitsch Delone (French: Delaunay, March 15, 1890 – July 17, 1980)
[16] Georgy Voronoy (April 28, 1868 – November 20, 1908)

the graphical simulation set-up of the device geometry, the configuration of the simulation, and the visualization and data manipulation of the simulation results to extract and export certain distributed quantities.

XCRV [37] is a front-end script for the two-dimensional plotting tool XMGRACE[17]. XCRV has been designed to access the most common plotting features by command line using XMGRACE via scripts.

- 2D-GUIs:

  XPIF2D [37] is a graphical tool to visualize simulation results of the device simulator MINIMOS-NT. It supports two-dimensional visualization and offers the capability to draw and export also cuts along user-defined lines.

- 3D-GUIs:

  SMARTV [315] is a visualization tool for the three-dimensional data formats from process and device simulation tools. SMARTV is capable of visualizing the graphical file formats FEM [39], FLD [39], and WSS [39, 316].

  As an alternative, the data explorer (OPENDX)[18] can be used as it provides a multi-purpose scientific visualization tool with the capability of individual configuration of the inputs and outputs. This includes a feature that allows the format of the data to be specified and adapted for a particular file format and at the user's need.

[17] http://plasma-gate.weizmann.ac.il/Grace/
[18] http://www.opendx.org/

Chapter 5

Applications

"A little inaccuracy sometimes saves tons of explanation."

Hector Hugh Munro[1]

This chapter presents typical applications for optimization in TCAD. Since the field of TCAD includes a wide range of applications, this chapter will present a subset of optimization to show the difficulties of these examples and their solutions. The first part of this chapter presents the calibration of certain model for deposition of SiO_2 using a LPCVD process. The second part deals with the extraction of thermal coefficient for the electrical and thermal conductivity. The third part presents some examples where the information obtain from the first two examples are applied in order to provide accurate material models for more advanced structures.

5.1 Calibration of Trenches

This section focuses on the comparison of calibrated models for Tetra-ethoxy-silane (TEOS) $Si(C_2H_5O)_4$ deposition in a CVD process according to SEM images of SiO_2 layers (cf. Figure 5.1 [317]). In the first part, the level-set algorithm is briefly introduced followed by the description of the quality calculation of the simulation results and different models for deposition of SiO_2 layers for trenches with different aspect ratios. At the end of this section, the final parameter calibration for the best model is presented and discussed in detail.

5.1.1 Treatment of Moving Boundaries

Chemical and physical processes are used to pattern and to form the surface of microelectronic device structures. The simulation of such surfaces requires on the one hand a rigorous data management to describe the proper material interactions of the reactants and the surface materials as well as on the other hand a huge amount of resources to map the real structure into data structures that can be handled within finite time and computational power.

The first attempt has been introduced decades ago, where the solid is divided into cell structures which also represents solids which can be either added (deposited) or removed (etched). This

[1] Hector Hugh Munro (December 18, 1870 – November 14, 1916)

cellular-based approach has been discussed in [172, 318]. However, the simulation of realistic structures is rather limited due to the high demand of memory resources which considerably slows down the simulation if virtual memory has to be assigned for computational purposes.

To overcome this type of limitation, SETIAN and OSHER have introduced a different method to describe moving surfaces and boundaries in [319–324]. The level set function $\mathbf{u}_{\mathrm{LS}}$ describes the evolving surface by a certain speed function $\mathbf{F}(t, \mathbf{x})$ as

$$\partial_t \mathbf{u}_{\mathrm{LS}} + \mathbf{F}(t, \mathbf{x}) \, \|\nabla_{\mathbf{x}} \mathbf{u}_{\mathrm{LS}}\| \; = \; 0, \tag{5.1}$$

where the speed function $\mathbf{F}(t, \mathbf{x})$ includes the physical effects and chemical reaction like etching, deposition, and diffusion. This speed function has to be specified or modeled for each material type separately. The moving surface is described by the set of points where the level set function is zero

$$\{\mathbf{x} | \mathbf{u}_{\mathrm{LS}}(t, \mathbf{x}) \; = \; 0\} \tag{5.2}$$

and the initial surface is given as the zero points at $t = 0$:

$$\{\mathbf{x} | \mathbf{u}_{\mathrm{LS}}(0, \mathbf{x}) \; = \; 0\} \, . \tag{5.3}$$

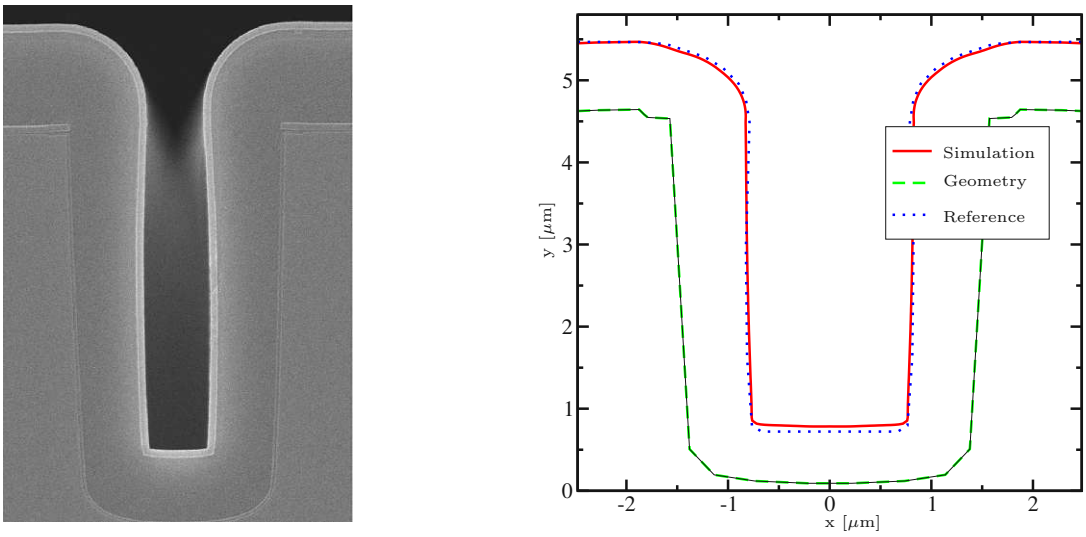
The main idea of SETIAN and OSHER was to describe the movement of the three-dimensional surface with a two-dimensional algorithm which decreases the required memory consumption. As a drawback this algorithm requires much less computational power than the cellular-based algorithm. The computational effort to describe a moving surface in three dimensions is for the level-set algorithm N^2 compared to N^3 for the cellular-based one. Hence, the level set approach provides a good alternative to expensive computation of evolving surfaces [165, 325–328].

5.1.2 Quality Criteria

To determine the quality of the simulation results obtained from the process simulation tool, several approaches have been used which have delivered similarly good results. First, different types of trenches have been investigated, where typical and characteristic distances have been extracted. The quality obtained by this approach is determined by the sum of the differences of these characteristic points of the trenches, where different points can be separately priorized. This procedure yields very fast numerical results. However, the better the simulation results are, the higher the numerical errors in a summation may be. In addition, if the simulated trench differs at other points than the chosen ones, this quality criterion also does not yield optimal results. Therefore, an alternative approach has been developed which uses the area between the simulation result and the measured trench shape. This method uses a numerical integration method, where the surface element between two points is linearly interpolated and the resulting area is meshed by a triangulation method. Hence, this method provides an accurate method to determine the difference between the two trenches. However, this type of calculation requires much more computational effort compared to the method which uses characteristic points only.

5.1.3 TEOS Deposition

The deposition of SiO_2 with TEOS is a complex pyrolytical chemical reaction. In this process TEOS is transported via a carrier gas to the hot surface of the wafer, where TEOS is dissociated [25]. A certain amount of the decomposition products sticks on the surface and build a SiO_2 layer while



(a) TEM picture of a trench series

(b) Simulation of a single trench

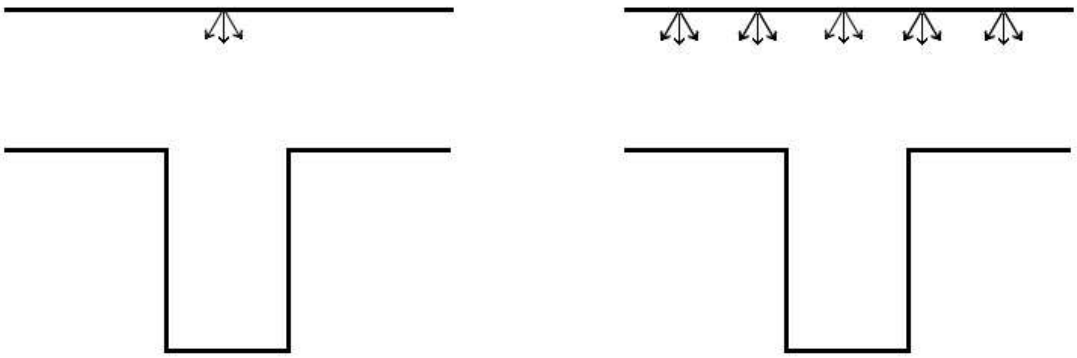
Figure 5.1: A test trench [317] for which different models have been developed to predict the TEOS deposition for a LPCVD process.

the other particles are reflected from the surface. Those are in general highly reactive by-products of TEOS decomposition. In particular, more than 40 secondary reactions have been reported in this complex reaction [328]. A rigorous simulation would cover all possible by-products and their secondary and ternary reactions but it would also require a considerable amount of computational power and memory to calculate and investigate this TEOS reaction.

However, one of the industrial requirements is to provide the engineers with rather fast and sufficiently accurate simulation results. Therefore, this project has focused on developing certain models to predict the TEOS deposition for a certain series of test trenches, where the characteristic aspect ratio (AR) is used to determine or estimate the impact on the chemical reaction behavior.

In order to obtain quantitatively accurate simulation results for deposition processes rather complex chemical models are required to describe the chemical reactions mechanisms. The computational effort is too high to include these rigorous models to software tools for industrial use [329]. Therefore, simplified model have been developed to speed up the simulation time [330] to obtain industrial-ready simulators. However, these simplified models have to be calibrated for each particular deposition process separately. The overall goal for this project was to find an appropriate deposition model and a certain parameter set which can be applied to all trenches of these test series to sufficiently predict the shape of the TEOS deposition.

The final shape of the deposited material can be adjusted by the pressure and the concentration of the reactant gas. In some cases, a conformal material deposition is used to protect the underlying materials. A conformal deposition of TEOS can be achieved using a low pressure chemical vapor deposition process (LPCVD) [331], with a certain constellation of temperature, pressure, and gas compositions. However, the stability is often not sufficient enough with respect to the material growth rate. By increasing the growth rate, the chemical process becomes increasingly unstable, which results in a position-dependent growth rate due to reactant-depletion, for instance, if not enough TEOS is supplied from the material source. In this regime of deposition the aspect ratio



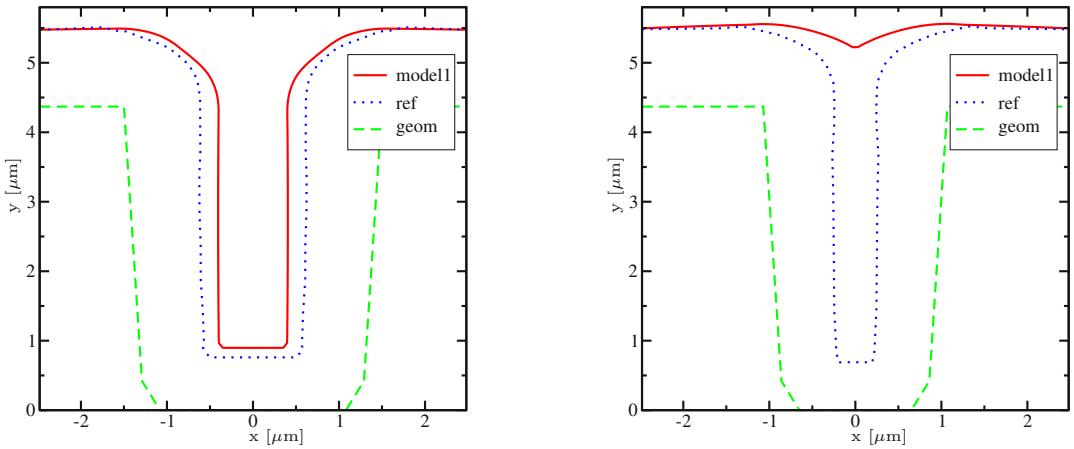
(a) Single source deposition model (b) Line source deposition model

Figure 5.2: Comparison of two deposition models using a point-shaped and a continuous line as material source.

(AR) is an important quantity which determines for some chemical reactions whether the reaction is mass flux limited or reaction-limited (cf. Section 2.4.1). The AR is originally defined on basis of a rectangular-shaped trench [332]. To determine the AR for non-rectangular trenches, a similar calculation has been chosen as for rectangular trenches (cf. Appendix A.4).

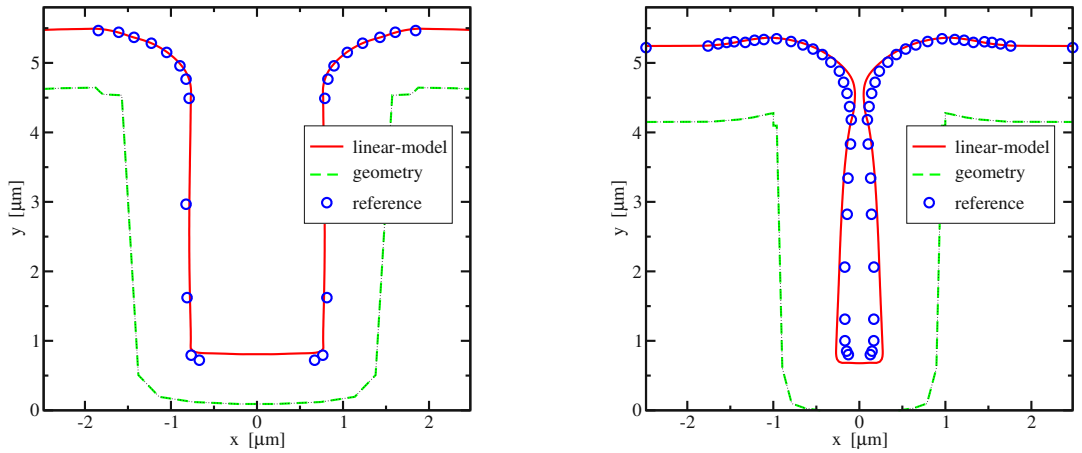
5.1.4 Deposition Models

In this section, the most promising models are described together with the best parameter sets which lead to the most accurate simulation results for the TEOS deposition in a certain AR range. The simulations have been performed using the topography simulator ELSA (Enhanced Level Set Applications) which follows the surface evolution by solving the level set equation [319, 333]. The parameter calibration and the optimizations have been carried out using the simulation and optimization framework SIESTA (cf. [334] and Section 4.5).



Trench with AR= 1.54. Trench with AR= 2.34.

Figure 5.3: Comparison between simulation results and the measurement for the single source model for two different trenches.



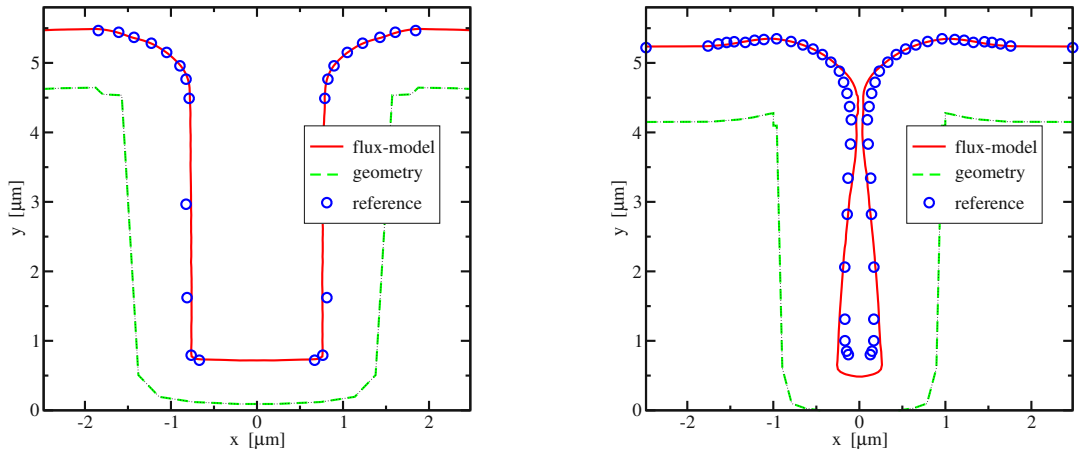
Trench with AR= 1.54. Trench with AR= 2.34.

Figure 5.4: Comparison between simulation results and the measurement for the linear material source model for two different trenches.

In an early attempt a model based on a single point-shaped material source was considered [333]. SIESTA was used to identify the sticking coefficients of the model and to tune the model, the distance between the source and the upper trench surface. The optimization has been performed for a trench with AR = 1.54. In Figure 5.3a a calibration result for this particular trench is presented. However, if this model is applied to trenches with a higher AR, the simulation results become worse and do neither quantitatively nor qualitatively compare with the measurement for an increased AR above approximately 1.2 as is presented in Figure 5.3b for trenches with AR of 1.54 and 2.34. Here, the latter trench is completely closed in contrast to the measurements. Nevertheless, for planar surfaces as well as for trenches with very low aspect ratios, this model gives reasonable good results.

Since the previous approach delivers only satisfactory results for a very limited set of trenches, the single source model has been expanded to a continuous line-source model where the material flux depends only on the visible angle between the surface elements and the source. The model describes the SiO_2 deposition by two sticking coefficients, where the first sticking coefficient β_1 describes the sticking probability of the incoming TEOS particles. As already mentioned, when the TEOS molecule hits the hot wafer surface, the TEOS molecule is pyrolytically dissociated. The probability that the SiO_2 will stick at the surface is modeled with the constant sticking coefficient β_1. With a certain probability $(1 - \beta_1)$, the particle is reflected from the first contact and hits the surface a second time. The probability that the SiO_2 is deposited at the second surface hit is described with the second sticking coefficient β_2. Compared to the earlier attempt, the problem that the top trench closes with SiO_2 at larger ARs, is shifted to higher ARs compared to the previous model, but the geometry at the bottom of the trench does still not satisfactorily agree with measurement as depicted in Figure 5.4. The corresponding parameters were calibrated with SIESTA, where the coefficients for the sticking probability β_1 and β_2 were identified as $\beta_1 = 0.248$ and $\beta_2 = 0.267$. Another improvement has been achieved by using a single flux-dependent sticking model [328]. The deposition reaction follows a half order kinetics of a CVD process of TEOS. Hence, the sticking coefficient β is modeled to be proportional to the inverse of the square root of the local mass flux coming from the source as

$$\beta = \beta_0 \, \Gamma(\vec{x})^{-\frac{1}{2}}, \tag{5.4}$$

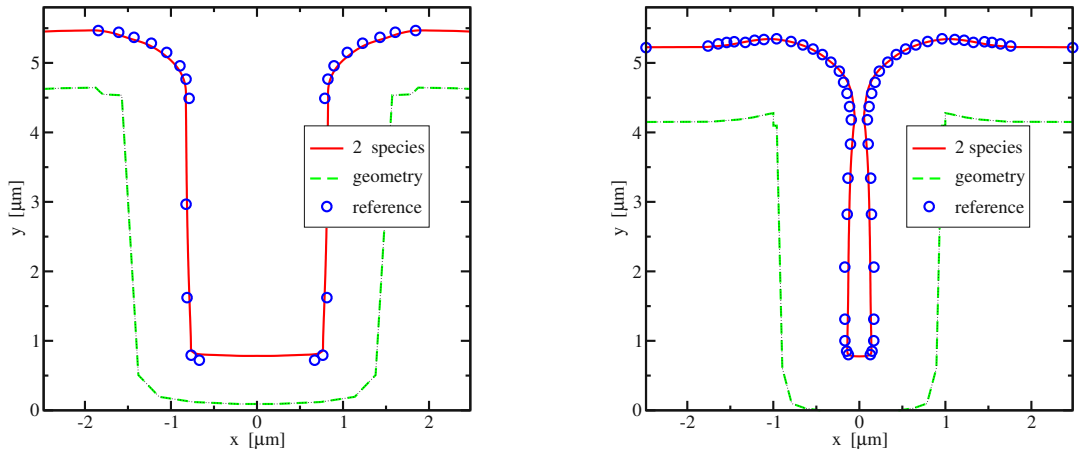


Trench with AR= 1.54. Trench with AR= 2.34.

Figure 5.5: Comparison between simulation results and the measurement for the flux-dependent deposition model for two different trenches.

where $\vec{x}$ is the local position, $\Gamma(\vec{x})$ the local material flux, and β_0 a constant scaling factor for calibration purposes. For the current LPCVD TEOS deposition process, the scaling factor β_0 was identified as $\beta_0 = 0.852$ with good agreement to the measurements for low aspect ratios as shown in Figure 5.5a. However, trenches with higher aspect ratios (cf. Figure 5.5b) show a considerable overestimation of the material thickness on the side walls of the trenches may result in spurious void formations for arbitrarily shaped trenches.

With increasing AR the overestimation of the thickness at the bottom increases as shown in Figure 5.4b and Figure 5.5b. Thus, another deposition model has been developed which considers two statistically independent species taking part in the TEOS deposition reaction. The first species is the original TEOS which is dissociated at the wafer surface. The second species is a sum of all other by-products of this pyrolytic reaction for which an average sticking probability is assigned. Since this model considers as initial condition a uniform distribution of reactants in the reactor, the sticking probability for both types of species can be assumed to be statistically independent. However, there is a dependence that the mass flux of the chemical by-products has to be proportional to the incoming material flux of TEOS due to mass conservation. With the assumption that the TEOS deposition is a LPCVD process, the TEOS particles are in a rather low concentration in the reactor and have a large mean free path between collisions. Thus, the material source for the reactor has to be sufficiently provide TEOS inside the reactor to obtain a stable deposition process. Since this requires a high exchange rate of the gases inside the reactor according to balance the high consumption rate of TEOS, the ratio between the TEOS concentration and the concentration of the by-products can be assumed to be in average constant. Despite of this fact the second species is modeled with an additional constraint which guarantees that the flux of the second species is proportional to the reflections of the first species because the second species is a product from the dissociation of the TEOS particles. If, for instance, trenches with higher ARs are considered, the material deposition at the bottom of the trench is much less than at the top because the distance form the TEOS source to the bottom of the trench is much wider and the by-products cannot be exhausted as quickly as from the top side of the trench. Hence, the flux of the second species is modeled to be proportional to the flux of the first species but with a statistically independent sticking probability.



Trench with AR= 1.54. Trench with AR= 2.34.

Figure 5.6: Comparison between simulation results and the measurement for the two-species deposition model for two different trenches.

This new two-species model shows excellent agreement with measurements for different geometries as shown in Figure 5.6a and Figure 5.6b. The sticking coefficient β_1 and β_2 were identified with SIESTA as $\beta_1 = 0.581$ and $\beta_2 = 0.732$. In addition to the astonishing agreement, this model provides faster simulation results and requires only 80% of the CPU time compared to the flux-dependent deposition model and overcomes the overestimation at the side walls of the trenches. Thus, the presented method enables efficient and accurate geometry optimizations. The hereby extracted sticking coefficients have also been applied to three-dimensional structures and have shown promising results.

5.1.5 Model Calibration

During the model calibrations and the identification of the new model parameters, the inverse modeling mode of SIESTA has been used. The calibration of this fairly complex TEOS process required a large numerical effort in terms of the number of parameter evaluations which was therefore performed on a heterogeneous network with more than 50 computational nodes. This was possible because the parameter sets were independent from each other.

To verify the different models, each model has to be calibrated for certain characteristic trenches in order to check whether the model fulfills the requirements. This procedure is very time consuming. Hence a critical trench (cf. Figure 5.1b) has been selected for performing all tests trench to compare the different models under the same conditions. This calibration target has been chosen as a reference because the aspect ratio is rather small, the shape of the surface is unusual and shows a certain non-conformity of the thickness of the deposited material, and the rounding at the trench openings are asymmetric. Figure 5.7 presents the result of the calibration of the two-species model in comparison to other trenches with a different aspect ratio. The parameters obtained from this calibration have been applied to completely different trenches created with the same process technology (cf. Figure 5.8). As can be seen, the model with the calibrated parameters also gives excellent agreement for different trenches.

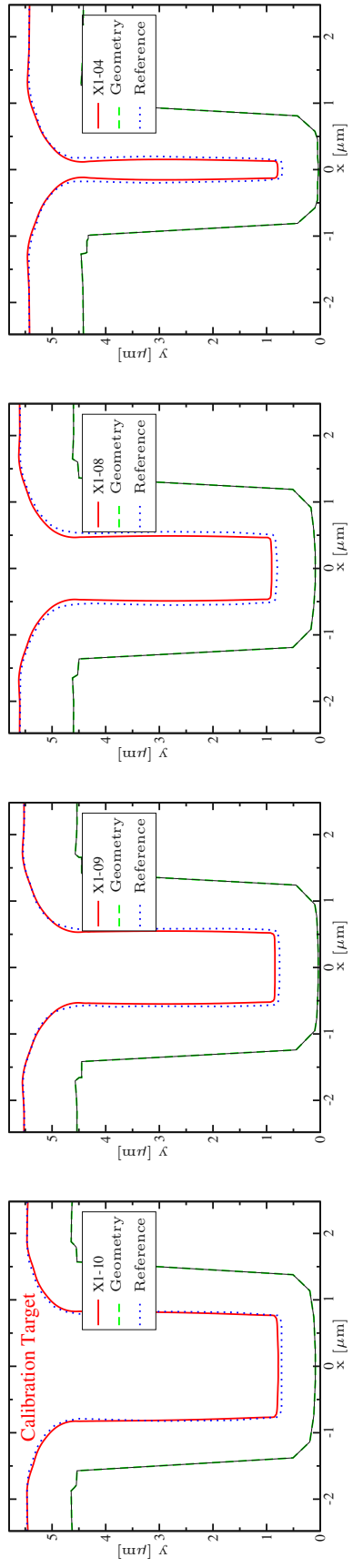


Figure 5.7: Calibration target with some test trenches from the first series using the two-species deposition model.

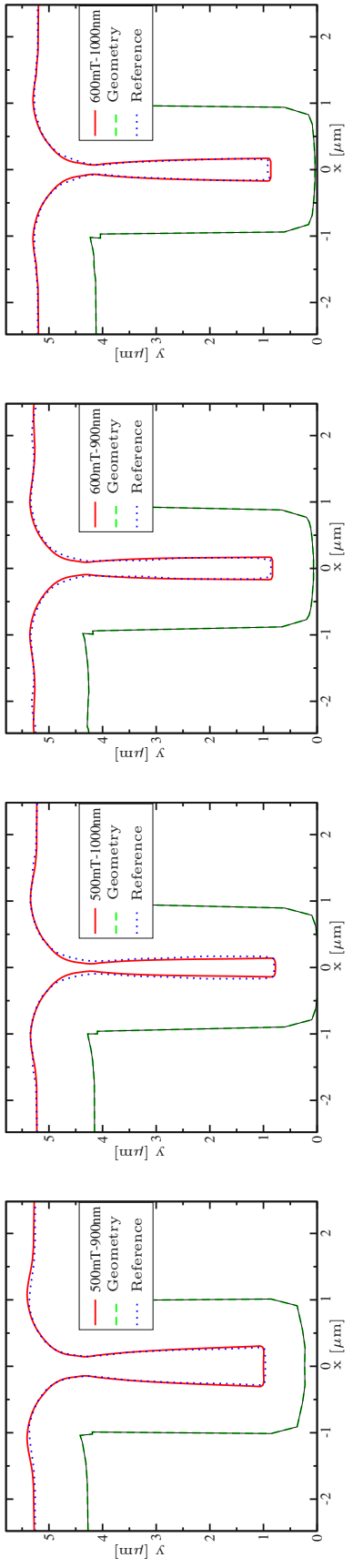


Figure 5.8: Second species test trenches series using the identified parameter parameters from the first trench series.

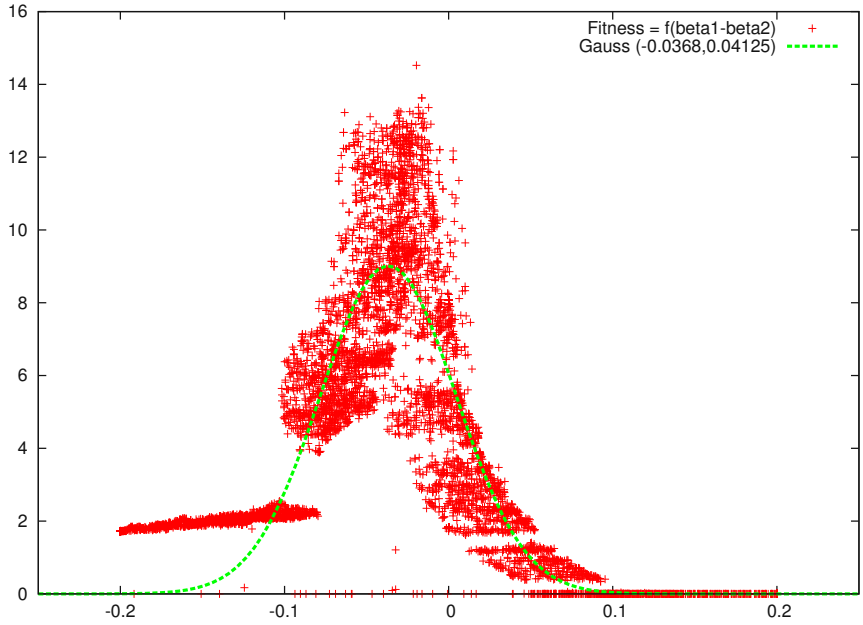


Figure 5.9: Fitness value of the best results obtained by the trench simulation.

In order to demonstrate the computational effort, Figure 5.9 depicts the fitness value of the best 6000 parameter sets of the calibration for the two-species model. The simulation results shown in Figure 5.7 and Figure 5.8 are obtained by applying the extracted sticking coefficients from the calibration of the target trench. However, the best results can be obtained if a calibration is performed for each single trench. The results for some selected test trenches from the first series (cf. Figure 5.7) are shown in Figure 5.10. The tendency shows that with increasing aspect ratios, the sticking probability drops rapidly. Due to the depletion of the active reactants in the trenches an exponential behavior is reported [335].

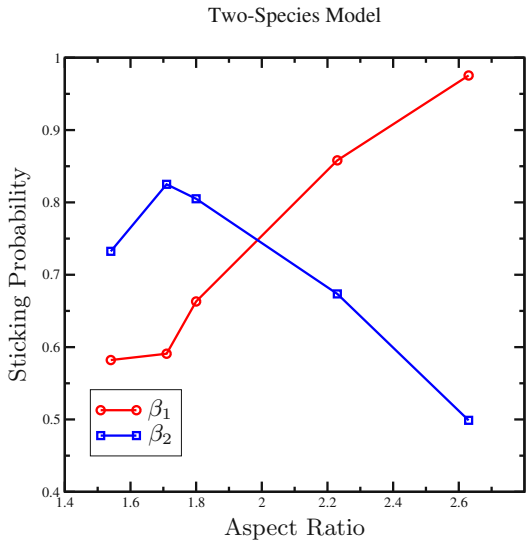


Figure 5.10: Sticking coefficient as function of the aspect ratios of the trenches.

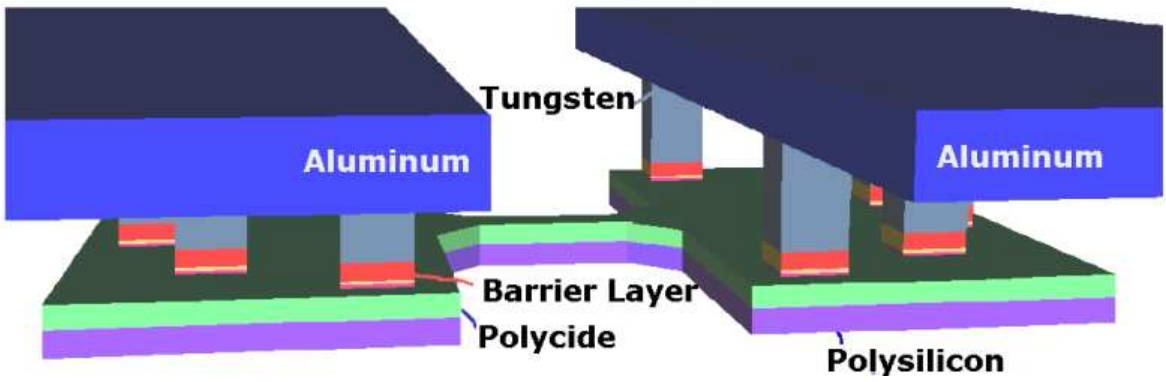


Figure 5.11: Sketch of a fusing device showing the variety of materials included.

5.2 Parameter Extraction of a Fusing Structure

Process technology nodes with 0.35 micron feature size and below offer the opportunity to integrate fuses made of polycrystalline interconnect materials for programmable memory cells. In technology nodes with larger feature sizes, the programming of a fuse can cause severe damage to the passivation layers.

With these fusing devices one can provide silicon circuits with several kilobits of cheap, non-volatile, and area-saving memory cells which can be programmed even during device operation. Rather new developments have been reported which provide tri-state fusing structures [336] as well as special fusing geometries that provide a reliable programming process [88,89,150]. Another important type of application is to use these fusing structures as field programmable gate arrays for trimming circuits [337]. In addition, fuses can be used as elements for programmable arrays for capacitors and resistors [338]. Moreover, fuses can also be used as protective elements for critical components [339].

With the help of sophisticated simulation and optimization tools new models can be developed and calibrated to provide more information to shorten the fusing time and to improve reliability. The main goal of this project was to identify the thermal impact of the electrical and thermal conductivities of the contributing materials.

5.2.1 Fusing Structure

A typical structure of a fuse consisting of various interconnect materials is shown in Figure 5.11, where the complex material composition is presented. To use interconnect materials only is advantageous because additional costs due to extra layers, masks, and process steps can be minimized by using already available materials and process steps. Moreover, in terms of power and area consumption, fuses which are made of already available interconnect materials are economically more attractive compared to hybrid technologies [340] which have to use different materials and thus additional process steps.

The fusing structure consists of two aluminum pads with a dual-layer rod in between, which is forseen to melt during programming. The programming is performed by sending a current pulse through the fuse at an appropriate bias, resulting in an opening of the polycrystalline silicon film

in the dual-layer rod due to thermal second-breakdown. The transition takes place when parts of the polycrystalline silicon layer reach the melting point. The molten silicon is transported from the negatively biased side to the positively biased side through the drift of ions [341]. However, before the polySi is completely molten, an electromigration process starts which accelerates the thinning of the rod and therefore the heating and the melting [89].

Because the downscaling process demands also decreased supply voltages, a careful design is required which includes a rigorous optimization of the fusing structure to ensure the reliability of the programming mechanism [342] and to minimize the power consumption during the programming process of the fuse.

Since the fusing mechanism takes place within a very short time (couple of 10 ns) for an ideal voltage step and several micro seconds for a voltage ramp measurements are hard to obtain [90].

A better insight into the electrical and thermal characteristics is desired for the materials used in the fusing structure as shown in Figure 5.11. In particular, the goal of the parameter extraction is the characteristics of the temperature dependence of the thermal and the electrical conductivity of the key materials polySi and the polycide (WSi_x). In order to obtain reasonable results from the simulation accurate information about of the test circuit for the fusing device is required [47].

Because the measurement of the programming mechanism has to be carried out within a certain amount of nanoseconds, the programming is artificially prolongated to a couple of microseconds for the fusing time by applying a voltage ramp with a rising period of 100 μs. This procedure allows to measure the fusing current with reasonable accuracy. The corresponding measurement set up is shown in Figure 5.12, where a voltage ramp is applied to a buffer amplifier to minimize the influence (impedance) of the function generator and to provide a high slew rate. The resulting measurements are shown in Figure 5.13. These measurements serve as reference data for the parameter identification procedure. At the beginning of the applied voltage ramp (cf. Figure 5.13), the corresponding current shows a non-linear behavior due to the self-heating of the fuse. After a certain time, the structure has been heat up and the current rapidly increases to an externally constraint value which is given by the parasitics of the fuse, the test circuit, and the external measurement equipment. This is the point where the conductivity models used in the simulator will fail. The goal is to predict the thermal evolution of the electrical and thermal conductivity as well as the internal temperature until this point.

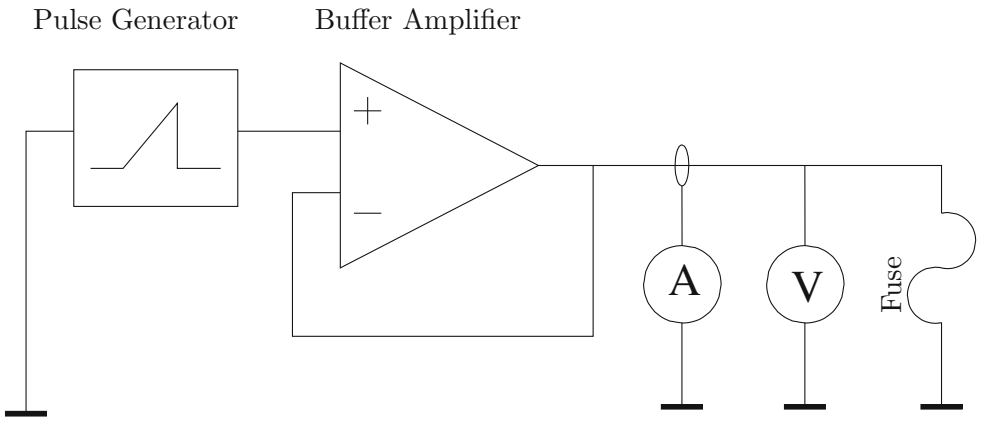


Figure 5.12: Schematic of the test circuit for the poly crystalline fuse.

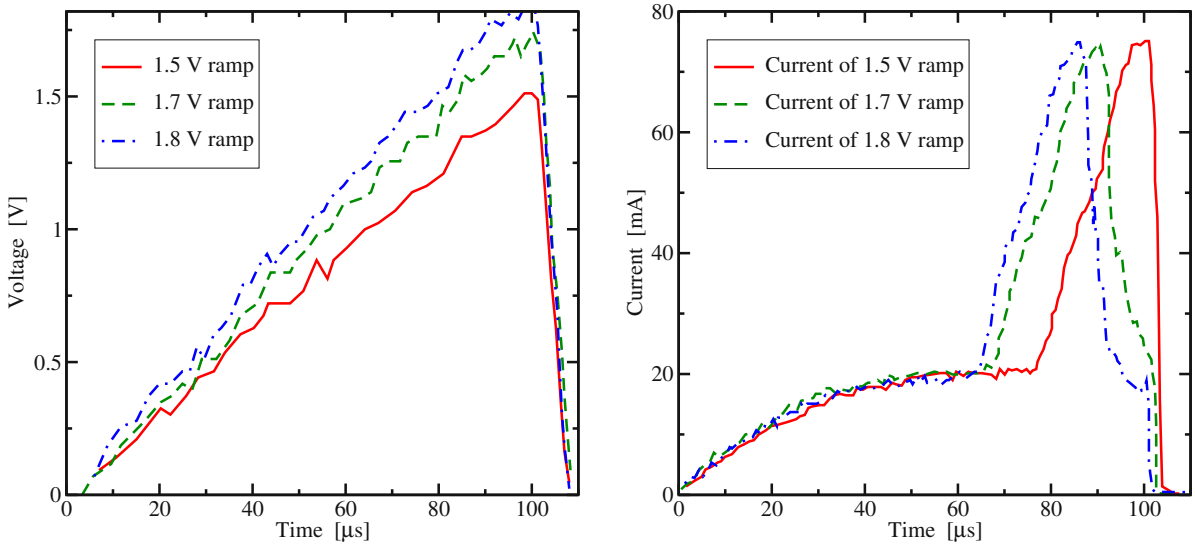


Figure 5.13: The different voltage ramps and the corresponding fusing currents.

5.2.2 Simulation and Parameter Extraction

For the transient simulation of the temperature evolution the three-dimensional interconnect simulator STAP [92] has been used within the simulation and optimization framework SIESTA. STAP calculates for isotropic materials JOULE's self-heating effect by solving the continuity equation (2.10) and the heat conduction equation (2.53), which is coupled to the power loss equation (2.54). For these investigations all material parameters are assumed to be isotropic and for the thermal and the electrical conductivity a temperature-dependence is approximated by a polynomial models

$$\sigma(T) = \frac{\sigma_0}{1 + \alpha_\sigma (T - T_0) + \beta_\sigma (T - T_0)^2} \tag{5.5}$$

$$\lambda(T) = \frac{\lambda_0}{1 + \alpha_\lambda (T - T_0)}, \tag{5.6}$$

where σ_0 and λ_0 are the conductivities at a certain reference temperature T_0 and α_σ, β_σ, and α_λ are the corresponding first and second order temperature coefficients. which represent the target for the parameter extraction.

For the parameter identification, the initial values for the thermal and electrical conductivity of polySi and polycide were obtained from literature [25, 136, 137, 343, 344]. A gradient-based optimization strategy has been used to improve the initial values in order to achieve good agreement with the reference data obtained from the measurements.

5.2.3 Results

This example has shown that with SIESTA it is indeed possible to determine the thermal coefficients of the electrical and thermal conductivities of polysilicon and polycide only from electrical measurements. This offers the possibility to minimize the computational as well as monetary effort for timely expensive caloric measurements.

Table 5.1: Comparison of extracted material parameters with data obtained from [137,343,344].

Quantities		polySi	$\text{polySi}_{\text{Lit}}$	Polycide	$\text{Polycide}_{\text{Lit}}$
σ_0	$[1/\mu\Omega\text{m}]$	0.12	-	1.25	$0.1 - 18.8$
α_σ	[1/K]	9.1×10^{-4}	10^{-3}	8.9×10^{-4}	$5 - 10 \times 10^{-3}$
β_σ	$[1/\text{K}^2]$	7.9×10^{-7}	-	8.1×10^{-7}	3.5×10^{-7}
λ_0	[W/Km]	45.4	40	119.4	$100 - 179$
α_λ	[1/K]	2×10^{-2}	10^{-2}	2.98×10^{-2}	-

With additional consistency checks within the simulator and the optimization framework, intermediate simulation results can be verified whether the obtained data is physically reasonable. The process of the parameter identification took therefore a certain time longer than expected, but has yielded excellent agreement with measurements as shown in Figure 5.14, Figure 5.15, and Tab. 5.1. Identified material parameters are summarized in Tab. 5.1, where the extracted coefficients are compared to data found in the literature [137,343,344]. The wide range of certain parameters shown in Tab. 5.1 is due to the uncertain stoichiometric coefficient of Si in polycide and the possible wide range of the applied doping of the polySi layer, which results in large standard deviations compared to the value for the pure materials.

Another interesting outcome of this investigation is that the temperature at which the resistance falls, is the same for all three different applied voltage ramps. Therefore, one can assume that this particular temperature corresponds to a material-specific phenomenon which is related to thermal run-away. This effect is most likely related to an electromigration process in the polycide layer [88, 89].

As expected the area with the highest local temperature is located at the surface of the bridge in between the two interconnect pads as shown in Figure 5.16. The extracted parameters can be used for further investigation of local temperature distributions and self-heating effects in other interconnect structures where similar materials are used.

Applying the derived model for polySi from Section 2.5.1, the trend of the characteristic resistance evolution can be very well reproduced. Figure 5.17 shows a comparison between the measured resistance, the calibrated polygonal conductivity model, and the polySi model from Section 2.4.1 which consists of various different materials parameters where the material parameters proposed in [179, 180] have been used with the adapted doping concentrations.

However, Figure 5.17 shows the two different parameter sets where only one parameter has been changed by $\pm 2\,\%$. Varying other model parameters offers the possibility to recalibrate the model where their values have no without physical meaning. Thus, without knowledge of the internal materials properties, for instance the distribution of the energy barriers, grain size, trap density at the grain boundary sites, and the thickness of the grain boundary region, the model can be calibrated roughly only. With knowledge of these fundamental material parameters a calibration of this model would yield an excellent match to the experimental data.

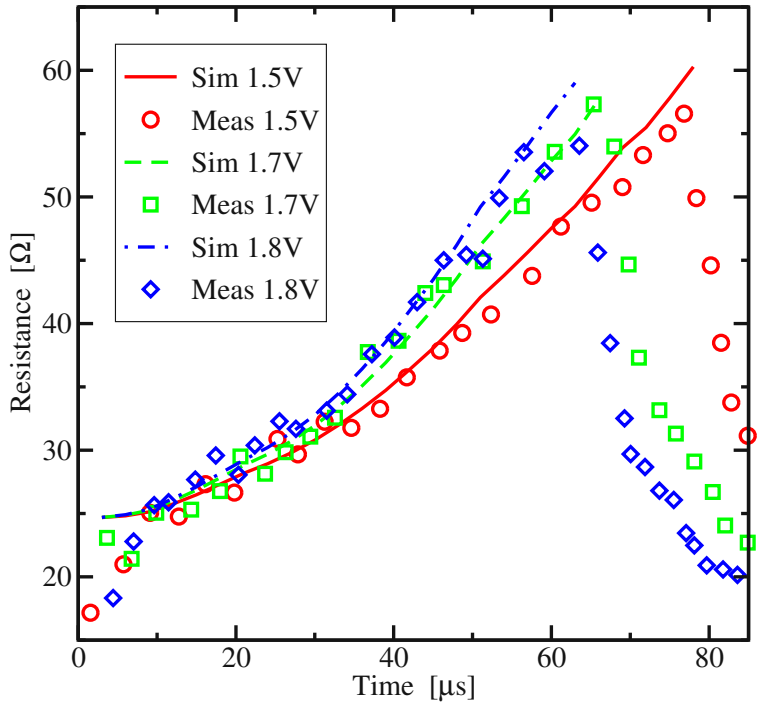


Figure 5.14: Comparison of the simulation results with the resistance measurements showing the evolution of the resistance of the fuse.

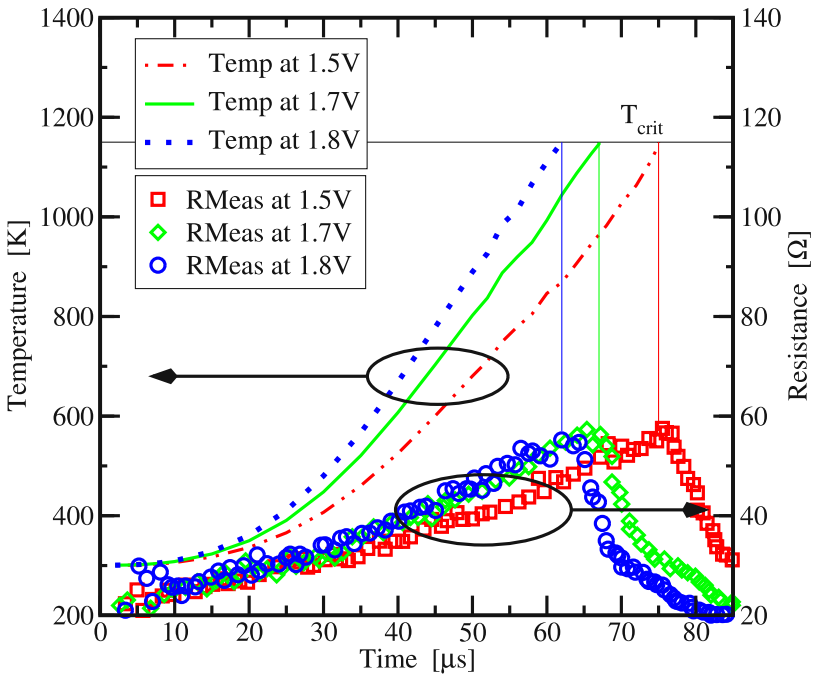


Figure 5.15: Comparison of the simulation results with the resistance measurements confronting the resistance evolution with the maximum temperature in the fusing structure.

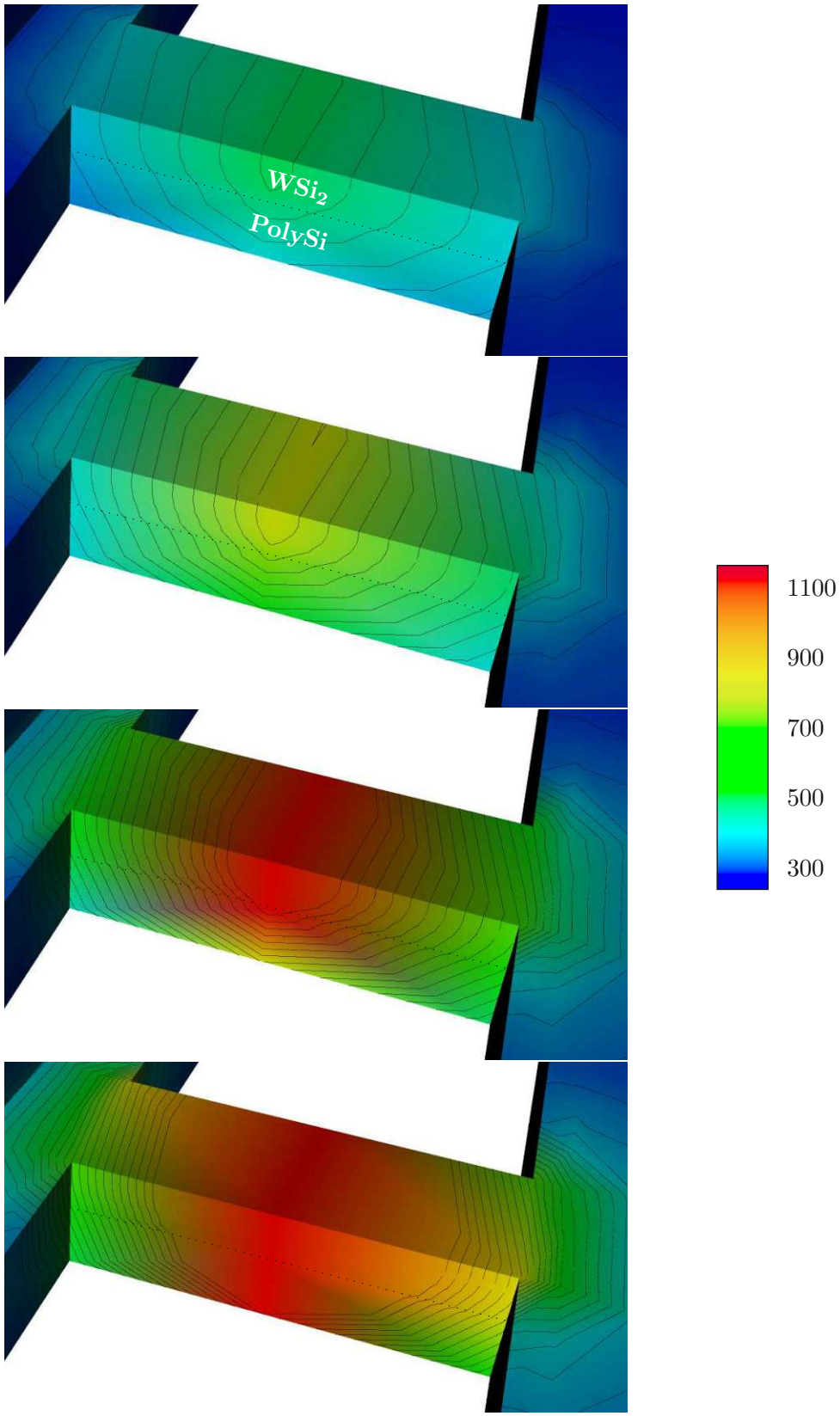


Figure 5.16: The temperature distribution [K] at the hottest spot in the fusing area. It shows that the upper region (WSi_2) reaches the highest temperature during programming at 32 μs, 40 μs, 50 μs, and 60 μs.

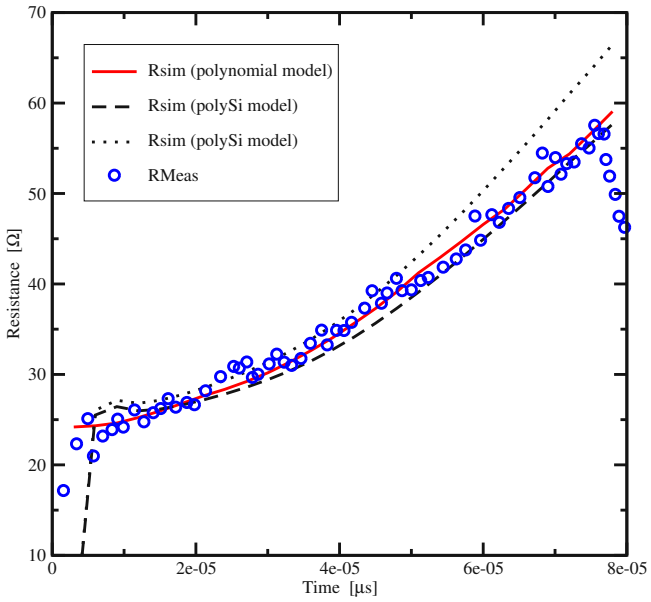


Figure 5.17: Comparison between the polygonal conductivity model and the polySi conductivity model applied to the fusing structure. For the polySi model two completely different parameters sets have been applied which both show rather a good agreement with the measurements.

5.3 Applied Optimized Parameters

The parameters which have been extracted in the previously shown examples have been applied to more complex structures and provide a better insight into the electro-thermal behavior of fairly complex interconnect structures. This is the basis for further simulations to obtain an accurate temperature distribution for rigorous investigation of the mechanical stress distribution and consequently also for electromigration analysis.

5.3.1 Thermal Analysis of a Multi-Layered Interconnect Structure

As an example a complex interconnect structures is depicted in Figure 5.18. This particular structure consists of a typical Cu material system with a Ta/TaN coating as presented in Section 2.3, where two Cu lines are connected through a bridge line. These interconnect lines are connected to the each other through Cu vias. The whole structure is embedded in an idealized low-κ material with $\varepsilon_\mathrm{r} = 3.0$.

This structure is biased with a voltage pulse of 0.5 mV and at the bottom of the structure, where an idealized heat sink is attached which stabilizes the temperature of the bottom layer to 350 K. At the other thermal boundaries, homogeneous NEUMANN conditions are applied. The simulation has been carried out using STAP [39]. This tool performs an electro-thermal simulation with temperature-dependent material models for the heat capacitance as well as for the electrical and the thermal conductivities.

Due to the applied bias, the temperature elevates due to self-heating. The resulting temperature distribution in the interconnect lines is shown in Figure 5.19 for 6.6 μs and 95.4 μs, respectively. As expected, the highest power loss density is located at the bottom of the via structure, which

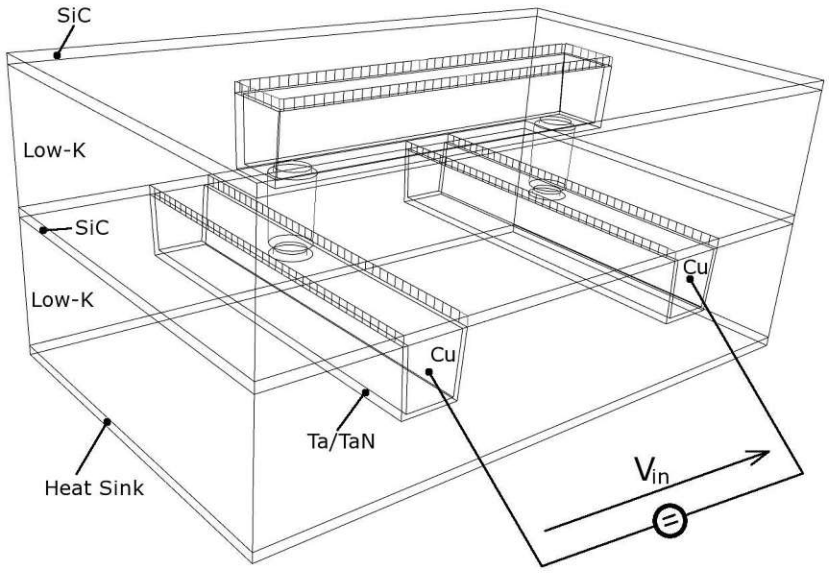
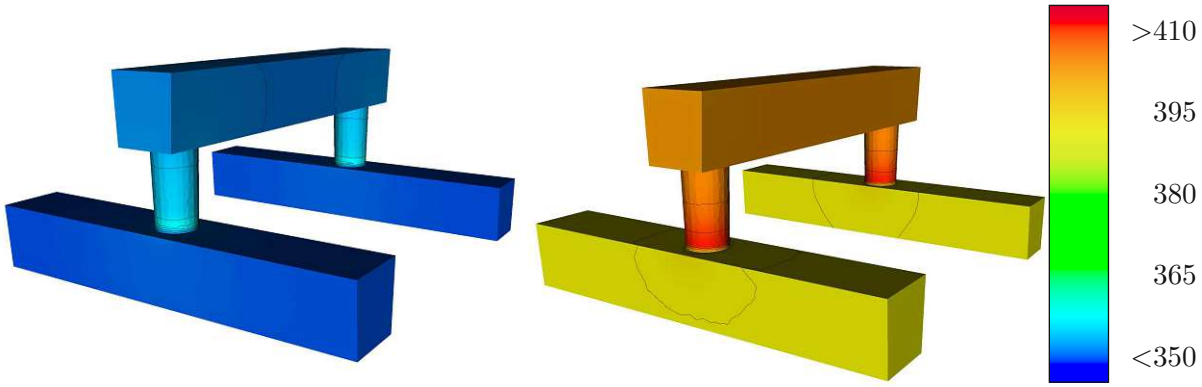


Figure 5.18: A bridge between two interconnect lines is biased with a voltage pulse.

can be seen as the hottest spot in the structure shown in Figure 5.19. Here, the internal temperature distribution of the entire interconnect structure is shown in Figure 5.20. It shows the isothermal surfaces in all materials. The high thermal gradient between the lower Cu lines and the heat sink can be clearly seen by the high density of the isothermal surfaces. At the opposite side — at the top — the heat is accumulated and the structure heats up because there is no appropriate thermal path to a heat sink. The maximum temperature is found at the bottom of each via (cf. Figure 5.19). From this location the heat dissipates rather quickly in all directions because the Cu lines provide a good heat conduction path compared to the surrounding low-κ material.

The material models for thermal investigation of complex interconnect structures are often simplified to reduce the computational effort that is necessary to obtain results with in reasonable time. However, if a rigorous transient thermal analysis is required, the difference between simplified models and temperature-dependent models can be significant.

The current examples structure (cf. Figure 5.18) has been investigated once with temperature-independent and then with temperature-dependent material models. The observed difference is depicted in Figure 5.21. While the power loss density is constant for temperature-independent materials, the material properties such as the electrical and the thermal conductivity of temperature-dependent materials changes. In return, this fact affects the power loss density (heat source) and consequently also the temperature. After a certain time, the system reaches a stationary state and the temperature is saturated. However, the value for the stationary temperature is in general different for these types of electro-thermal investigations. For this particular example, the simulation with deactivated thermal material models overestimates the stationary temperature by approximately 5.5 K (cf. Figure 5.21). It should be noted that the transient temperature evolution has a completely different shape due to the dynamic heating behavior of the interconnect stack.



(a) Temperature distribution [K] at 6.6 μs. (b) Temperature distribution [K] at 95.4 μs.

Figure 5.19: Temperature distribution in a multi-layered interconnect structure at two different points in time showing the global warming in the upper bar that connects the two lower interconnect lines.

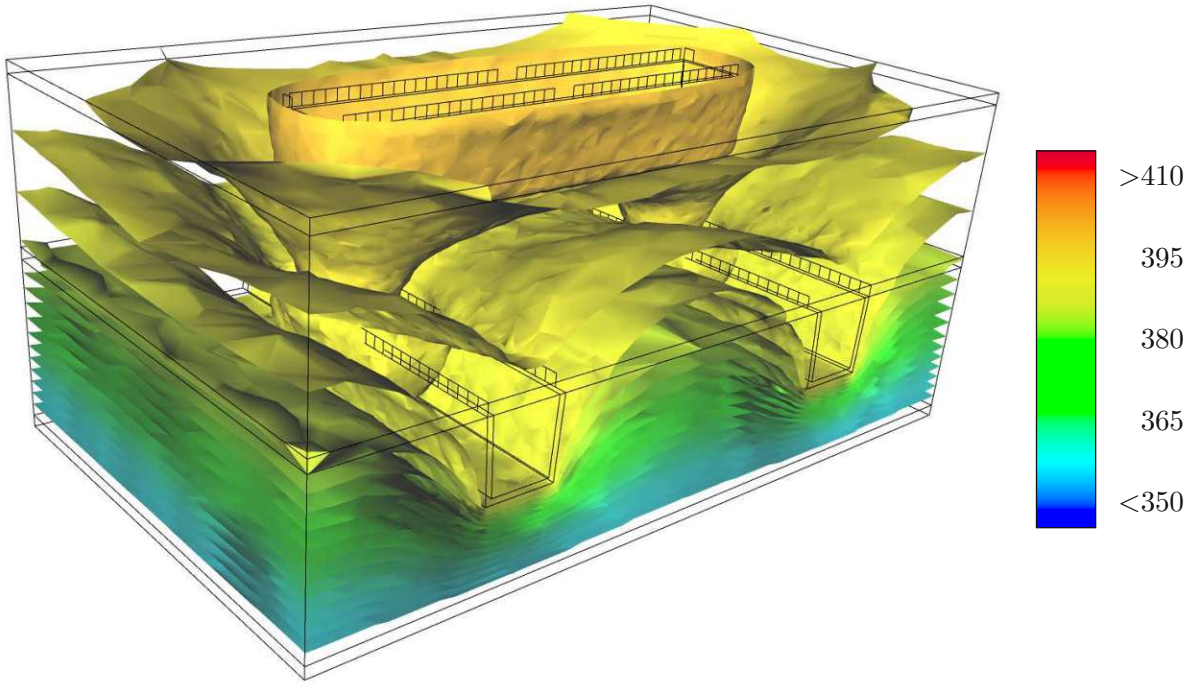


Figure 5.20: Temperature distribution [K] in the entire interconnect structure depicted by isothermal surfaces at 95.4 μs. The maximum temperature is found at the bottom of each via which is indicated by the isothermal surfaces with the highest temperature and by Figure 5.19.

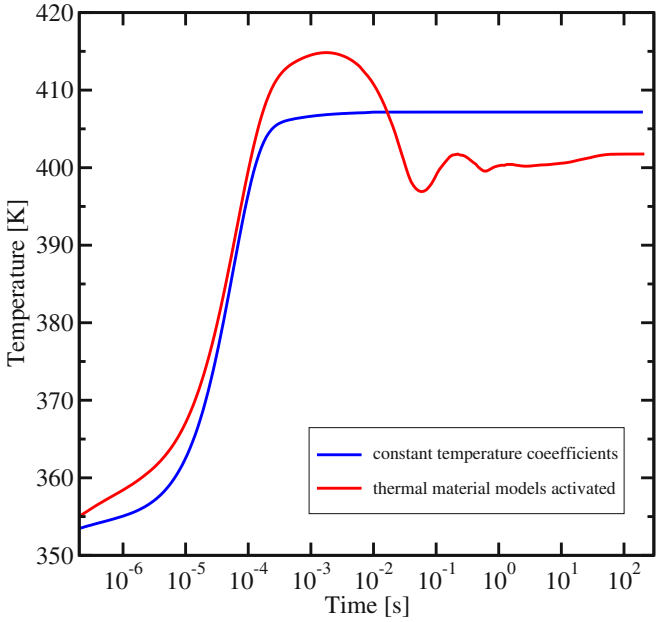


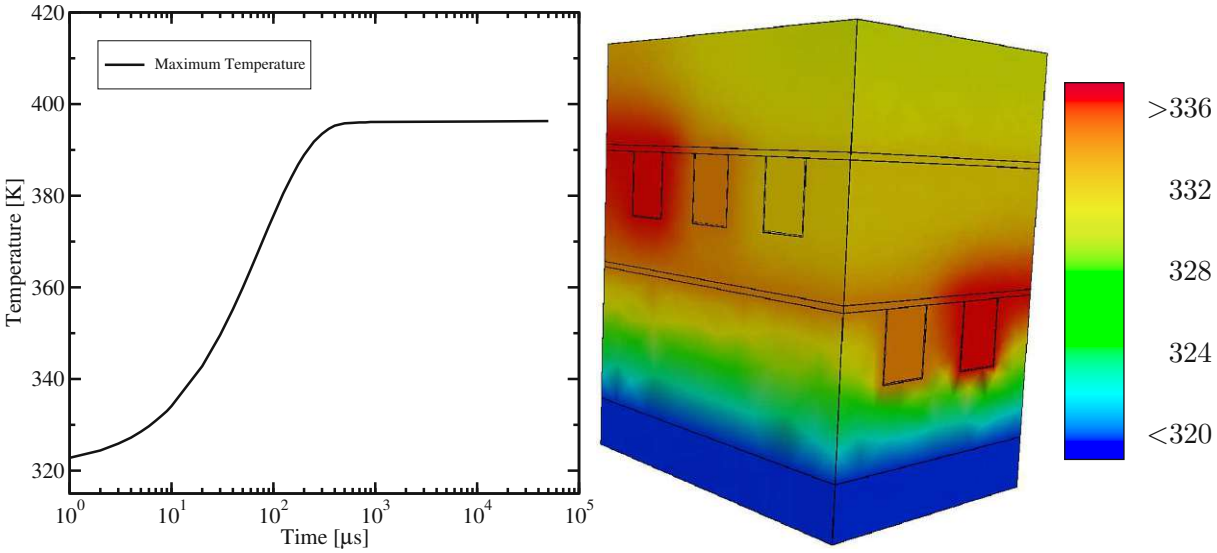
Figure 5.21: The evolution of the maximum temperature in the entire interconnect structure is depicted for activated and deactivated thermal material models.

5.3.2 Interconnect Structure Subjected to Mechanical Stress

If larger and more complex interconnect structures have to be investigated, the required detail level as presented in the previous example might exhaust the available computational capabilities in terms of memory and CPU power. Therefore, appropriately simplified models are used to analyze complex structures as presented for instance in Figure 2.3 to obtain a rough overview. However, the critical regions have to be extracted and recalculated at a higher detail level. Using successfully extracted material parameters provides a good insight into the transient device behavior — even with simplified material models.

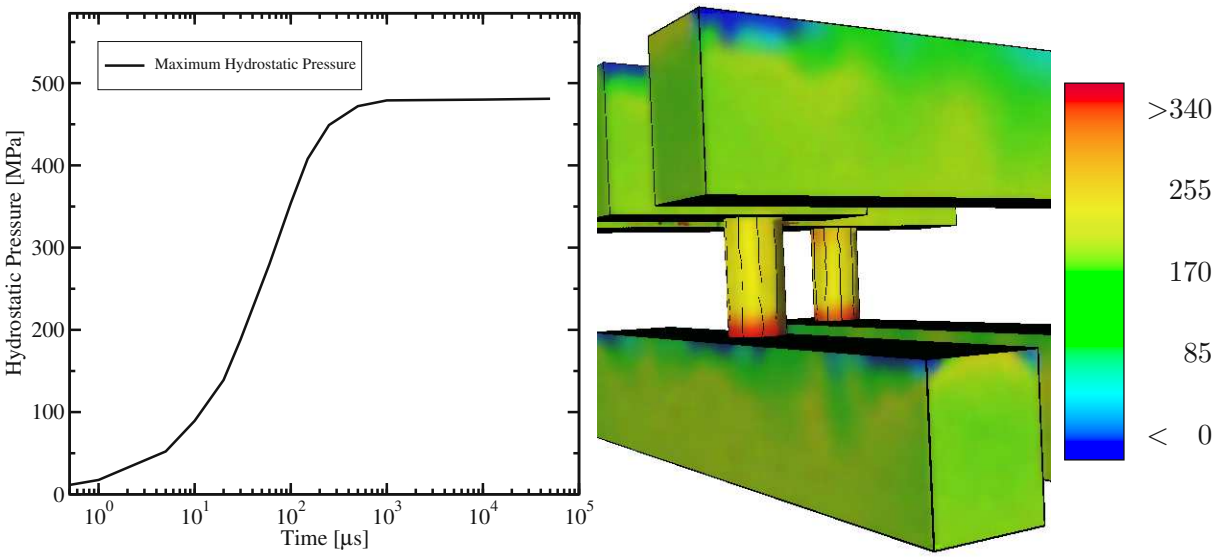
In this example the electro-thermal simulation with STAP [39] is used to produce the appropriate input data necessary for the thermo-mechanical simulation using the finite element simulator FEDOS [36, 345]. The simulation setup is quite similar to the one in the previous example. A bias is applied between two lines, which are connected through a via. A heat sink is attached at the bottom of the interconnect stack to keep the temperature at 320 K. At the other thermal boundaries, homogeneous NEUMANN conditions are applied. The simulations have been carried out with the simplified material model to obtain simulation results within reasonable time because the simulation is separated into two parts: the electro-thermal and the thermo-mechanical simulation.

The result of these simulations is at the first level the temperature distribution (cf. Figure 5.22), which is a required part of the input data of the thermo-mechanical simulator FEDOS. As a final result the hydrostatic pressure (cf. Figure 5.23) is obtained as the trace of the mechanical stress tensor. Figure 5.22a and Figure 5.23a show the evolution of the maximum temperature and pressure, respectively. In addition, typical distributions of the quantities are shown in Figure 5.22b and Figure 5.23b. This type of information can be used as representative quantities for design rule checks and for further optimization of the reliability of interconnect structures.



(a) Maximum temperature. (b) Temperature distribution [K] at 10 μs.

Figure 5.22: A typical multi-layered dual damascene interconnect structure showing the evolution of the maximum temperature (a) and a snap shot of the temperature distribution [K] at 10 μs (b).



(a) Maximum hydrostatic pressure. (b) Pressure distribution [MPa] at 100 μs.

Figure 5.23: The evolution of the maximum hydrostatic pressure in the interconnect structure resulting from an electric pulse (a) and a snap shot the hydrostatic pressure distribution [MPa] at 100 μs (b).

Chapter 6

Conclusions

Both compact modeling and physics-based electro-thermal models [92, 346] have their individual benefits for each application. In order to obtain sufficiently accurate simulation results in reasonable time, physics-based models have to be used at a certain level of abstraction. In contrast, if speed is more important, compact models often provide a good alternative. However, to obtain accurate and reasonably fast simulation results, an appropriate combination of compact models and fundamental physical equations has to be found for each class of simulation problem. There is a very large number of different simulation tools available, which are able to accurately calculate certain effects in different levels of details. But there are also some simulators and frameworks that provide a wide range of different basic models as well as compact models to cover a wide range for industrial demands. For very specific investigations of highly sophisticated devices, however, the standardized models inherently are often not sufficient enough and physic-based models are thus indispensable. Consequently, the user has to decide the level of detail for each simulation, in order to choose the appropriate models in the simulation tools to obtain reasonable results for the current simulation.

6.1 Temperature as a Limiting Factor

As it has been shown in this dissertation, the increasing transistor densities in today's VLSI devices cause more and more serious heating problems because the locally generated heat cannot be appropriately transfered to the corresponding heat sinks fast enough. The contribution of a single device part can often be neglected compared to the overall power loss. However, present devices have interconnect lines with lengths of several kilometers [19] and hundreds of millions of transistors [2] and the small contributions become considerably large and dominate today's microelectronic devices. Thus, the device structures as a whole tend to heat up globally.

The increased global temperature has aggravated several parasitic effects which are normally not as much anticipated as would be required. However, an unintentional temperature increase can be considered as a stochastic global event which affects the whole chip and requires a global heat load management. Other parasitic effects are amplified by the temperature increase as demonstrated by the example of electromigration, where the activation energy is relatively lowered with respect to the maximum allowed electrical load permissible. Mechanical changes, for instance phase changes, changes in the crystal structure, and interface conditions such as adhesion are thermally enhanced as well and can have a major impact on the long life reliability of semiconductor device structures.

Some of the thermally-induced effects are for instance self-heating, which causes raised resistivity and increasing delay times on interconnect lines. This poses two further serious problems for fast electronic circuits: In a circuit with high speed transistors the maximum frequency is rather high. A increase of temperature due to self-heating causes increased parasitic effects like elevated line resistance, while the transition times increase accordingly and thus slow the circuit. The second example is global heating, which is not only limited to on-chip or on-die heating, but affects also the surrounding discrete devices on the circuit board.

The currently chosen thermodynamic treatment builds first faster, smaller, and denser device arrays and obtains therefore a higher power density distribution on the die. If the already considered heat sink provides enough cooling, everything seems to be working properly. But if not, the problem becomes even more critical if, for instance, the power loss due to the leakage currents is considered. In that case, the device generates heat even if it is not operating. For a modern microprocessor with typically 10^8 transistors and an average leakage current of 10 nA per transistor, the total current would be 1 A. Hence, for a typical supply voltage of 1.33 V, the device consumes 1.33 W in idle mode — a significant power consumption that is highly undesirably.

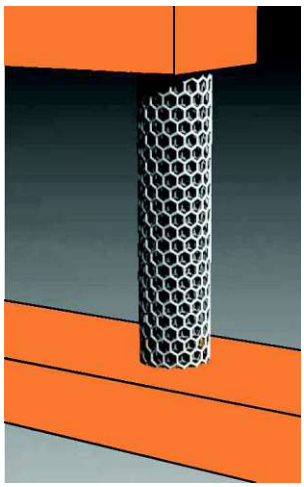
With decreasing feature size of the transistors, the power consumption keeps almost constant, while the power-dissipating area shrinks quadratically and the power-loss density explodes at a square ratio. Possible ways to solve this type of heating problem are to use better materials which show less leakage, to reduce the clock frequency, to introduce highly efficient heat conduction paths through the devices, to minimize the current for instance by the reduction of capacitances, and to decrease the supply voltage.

6.2 Multiple (Thermal) Redundancy

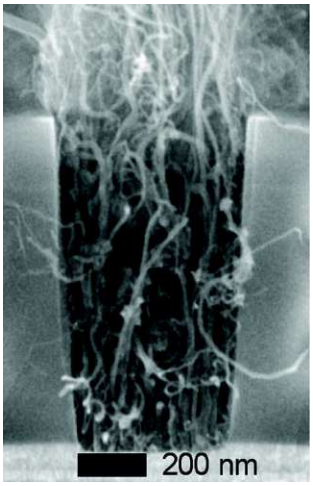
An approach to reduce standby heat resulting from leakage currents is to shut down unused parts of a device. This is done by switching off the supply voltage for the particular device region, so that this part is set to idle unless the supply voltage is switched on again. This system is currently used in highly sophisticated microprocessors, where several arithmetic or floating point units can be set to idle separately. Therefore this feature is highly desirable for high performance devices to cool down certain device regions, but it is also a common method for devices used in battery powered applications.

6.3 New Materials and New Devices Structures

As the last decades have shown, the progress in semiconductor technology seems to continue unabatedly. In the past, despite of the forecast *red brick walls*, the only constant factor and a simultaneously self-fulfilling prophecy was Gordon E. Moore's law, which is likely to remain valid for the next decade. However, there are apprehensions about physical size and other constraints but similar concerns were announced the last decades ago. With the present technologies the industry is facing similar problems as ten or twenty years ago. However, progress through innovation is the only way to succeed in economics as well as in science and technology. Therefore, alternatives are investigated where some of which are looking more promising than others. One among these auspicious technologies is the use of new devices made of or supported by carbo-nano-tubes (CNT). They offer extreme values for the electrical and thermal conductivities but, on the other hand, show major difficulties in fabrication and embedding into existing systems, for instance with respect to contacting the CNTs. However, in nearly every conference on nano



(a) A multi-wall CNT as via between two Cu interconnect lines.



(b) TEM image of CNT grown in a via structure using a CVD process.

Figure 6.1: Comparison of CNTs in interconnect structures [347] providing a promising alternative enhancement of improving the reliability especially in via structures.

technology additional steps towards the usability of CNTs are presented. Thus, it is only a matter of time that such devices will be available the leaders in the sector of advanced semiconductors.

As proposed in [347], CNTs have already been used in interconnects for prototypes Figure 6.1. This may overcome or even solve critical problems with the conductivity and adhesion of Cu, which is related to electro-migration. By using CNTs in the vias, their advantageous properties can be used to increase reliability rather than with the expensive production of protective barrier coatings.

6.4 Outlook

Despite all advances of technology so far as well as in the future, new challenges will arise and new improvements and enhancements will be required in order to fulfill Gordon E. MOORE's (ITRS) prophecy. A theoretical device performance beyond the ITRS is possible, but mostly with an extremely high effort in terms of additional costs and in terms of developments concerning with respect to reliability of the materials and devices. All this boils down to a theoretical feasibility and the question whether it is worth spending more to build and advance such electronic devices.

Here, the future tasks for optimization can be clearly seen. Future device optimization will have to concentrate on providing additional information for improving device characteristics under various electrical, chemical, mechanical, thermal, and economical constraints. With increasing complexity and shrinking feature sizes, these constraints are becoming more critical and thus require effective and robust optimization and simulation tools to ensure proper device operation.

Therefore, product line optimization in terms of saving money will gain momentum for future microelectronic devices. The temporary sub-targets for optimization will change slightly, but the standardized overall targets *faster*, *cheaper*, and, therefore, also *smaller* and *more reliable* than the competitors will remain as the basic motivations for the upcoming advancements of the future state-of-the-art in electronics for economics as well as for military applications.

Appendix A

Mathematical Notes

A.1 Convexity of Sets and Domains

A subset $S \subset X$ is defined to be *convex* if for arbitrarily chosen points $x, y \in S$ the connection line between these two points

$$[x, y] := \{(1 - \xi)x + \xi y : \xi \in [0, 1]\} \tag{A.1}$$

is completely included in S [219, 348]. Therefore, the implication

$$x, y \in S \quad \Rightarrow \quad [x, y] \subseteq S \tag{A.2}$$

can be applied. Figure A.1 shows two typical parameter domains. The left domain in this figure is convex but the second one has obviously not a convex nature. For unconstraint parameters of

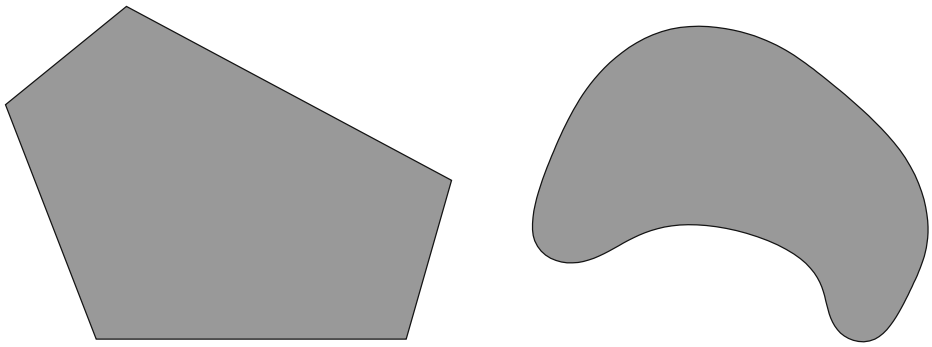


Figure A.1: Convexity of domains

an optimization problem the parameter domain is convex. If non-linear constraint functions are applied to this problem the parameter space might be become non-convex. If penalty or barrier functions are used instead of constraint functions the objective function becomes more complex but the lemmas for the convergence of these optimization problems can be again applied as for unconstraint optimization [219, 220].

A.2 VOIGT Notation

The VOIGT notation is used to exploit the symmetry of condensed matter to transform second-order tensors (cf. e.g. stress tensor) to vectors and kinetic forth-order tensors to square matrices (cf. e.g. elasticity tensor) [98]. This notation is often used in continuum mechanics.

A symmetric tensor, for instance the stress tensor $\tilde{\sigma}^{\text{mech}}$ in three dimensions, can be written as

$$\tilde{\sigma}^{\text{mech}} = \begin{bmatrix} \sigma_{11}^{\text{mech}} & \sigma_{12}^{\text{mech}} & \sigma_{13}^{\text{mech}} \\ \sigma_{21}^{\text{mech}} & \sigma_{22}^{\text{mech}} & \sigma_{23}^{\text{mech}} \\ \sigma_{31}^{\text{mech}} & \sigma_{32}^{\text{mech}} & \sigma_{33}^{\text{mech}} \end{bmatrix} \longrightarrow \begin{pmatrix} \sigma_{11}^{\text{mech}} \\ \sigma_{22}^{\text{mech}} \\ \sigma_{33}^{\text{mech}} \\ \sigma_{23}^{\text{mech}} \\ \sigma_{13}^{\text{mech}} \\ \sigma_{12}^{\text{mech}} \end{pmatrix} = \begin{pmatrix} \widehat{\sigma}_{1}^{\text{mech}} \\ \widehat{\sigma}_{2}^{\text{mech}} \\ \widehat{\sigma}_{3}^{\text{mech}} \\ \widehat{\sigma}_{4}^{\text{mech}} \\ \widehat{\sigma}_{5}^{\text{mech}} \\ \widehat{\sigma}_{6}^{\text{mech}} \end{pmatrix} = \widehat{\sigma}^{\text{mech}}. \tag{A.3}$$

If the shear components $\sigma_{ij}^{\text{mech}}, (i \neq j)$, can be neglected, the VOIGT transform degenerates to the diagonal operator $\text{diag}(\cdot)$. Hence, the later 3 components of the resulting six-dimensional vector vanish ($\sigma_{ij}^{\text{mech}} = 0, i \neq j$) and if only orthotropic materials are considered, the shear components always vanish and the six-dimensional vector can be written as a three-dimensional vector.

This simplification can be often applied if only tensile or compressive stress in the direction of one main axis of an orthotropic material is considered.

For the strain tensor $\tilde{\varepsilon}^{\text{mech}}$ the transformation is very similar. Due to historical conventions, the resulting VOIGT-transform of the strain tensor is the engineering strain $\mathbf{e}^{\text{mech}}$

$$\begin{aligned} \tilde{\mathbf{e}}^{\text{mech}} &= \begin{bmatrix} \tilde{e}_{11}^{\text{mech}} & \tilde{e}_{12}^{\text{mech}} & \tilde{e}_{13}^{\text{mech}} \\ \tilde{e}_{21}^{\text{mech}} & \tilde{e}_{22}^{\text{mech}} & \tilde{e}_{23}^{\text{mech}} \\ \tilde{e}_{31}^{\text{mech}} & \tilde{e}_{32}^{\text{mech}} & \tilde{e}_{33}^{\text{mech}} \end{bmatrix} = \begin{bmatrix} \varepsilon_{11}^{\text{mech}} & 2\,\varepsilon_{12}^{\text{mech}} & 2\,\varepsilon_{13}^{\text{mech}} \\ 2\,\varepsilon_{21}^{\text{mech}} & \varepsilon_{22}^{\text{mech}} & 2\,\varepsilon_{23}^{\text{mech}} \\ 2\,\varepsilon_{31}^{\text{mech}} & 2\,\varepsilon_{32}^{\text{mech}} & \varepsilon_{33}^{\text{mech}} \end{bmatrix} \\ &\longrightarrow \begin{pmatrix} \varepsilon_{11}^{\text{mech}} \\ \varepsilon_{22}^{\text{mech}} \\ \varepsilon_{33}^{\text{mech}} \\ 2\,\varepsilon_{23}^{\text{mech}} \\ 2\,\varepsilon_{13}^{\text{mech}} \\ 2\,\varepsilon_{12}^{\text{mech}} \end{pmatrix} = \begin{pmatrix} \tilde{e}_{1}^{\text{mech}} \\ \tilde{e}_{2}^{\text{mech}} \\ \tilde{e}_{3}^{\text{mech}} \\ \tilde{e}_{4}^{\text{mech}} \\ \tilde{e}_{5}^{\text{mech}} \\ \tilde{e}_{6}^{\text{mech}} \end{pmatrix} = \widehat{\mathbf{e}}^{\text{mech}}, \end{aligned} \tag{A.4}$$

where $\tilde{e}_{\text{ij}}^{\text{mech}}$, are the components of the engineering strain and the shear strain components transform to the CAUCHY strain components as $\tilde{e}_{\text{ij}}^{\text{mech}} = 2\,\varepsilon_{\text{ij}}^{\text{mech}}$, where $i \neq j$.

A.3 Norms

A.3.1 Definition

A norm is a real-valued function $\|\cdot\|$ on a linear space $X \subseteq \mathbb{R}^n$ such that

$$\begin{aligned} \|\mathbf{x}+\mathbf{y}\| &\leq \|\mathbf{x}\| + \|\mathbf{y}\| && \text{(A.5)} \\ \|\alpha\mathbf{x}\| &= |\alpha|\,\|\mathbf{x}\| && \text{(A.6)} \\ \|\mathbf{x}\| = 0 &\iff \mathbf{x} = \mathbf{0} && \text{(A.7)} \end{aligned}$$

where $\mathbf{x}, \mathbf{y} \in X$ and $\alpha \in \mathbb{R}$.

A.3.2 Special Norms

The EUCLIDean[1] norm for a vector $\mathbf{x} \in \mathbb{R}^n$ is defined as

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2}. \tag{A.8}$$

The square of the EUCLIDean norm of a vector $\mathbf{x} \in \mathbb{R}$ can be written as

$$\|\mathbf{x}\|_2^2 = \sum_{i=1}^{n} |x_i|^2 = \mathbf{x}^{\mathcal{T}}\mathbf{x}. \tag{A.9}$$

More generally, the p-norm of the same vector $\mathbf{x}$ is defined as

$$\|\mathbf{x}\|_p = \sqrt[p]{\sum_{i=1}^{n} |x_i|^p}. \tag{A.10}$$

The maximum norm of a vector is defined as

$$\lim_{p\to\infty} \|\mathbf{x}\|_p = \|\mathbf{x}\|_\infty = \max\{|x_i|\}. \tag{A.11}$$

[1] Euclid of Alexandria (*approx.* 325 BC – 265 BC)

A.4 Aspect Ratio

The aspect ratio (AR) of a trench is defined as the ratio between the height H and the width or length L of a trench and is used as a characteristic quantity especially for chemical processes like material etching and depositing. For chemical material deposition, the aspect ratio calculation for rectangular trenches has been introduced in [332] (cf. Figure A.2a) using

$$\mathrm{AR} = \frac{H}{L} \tag{A.12}$$

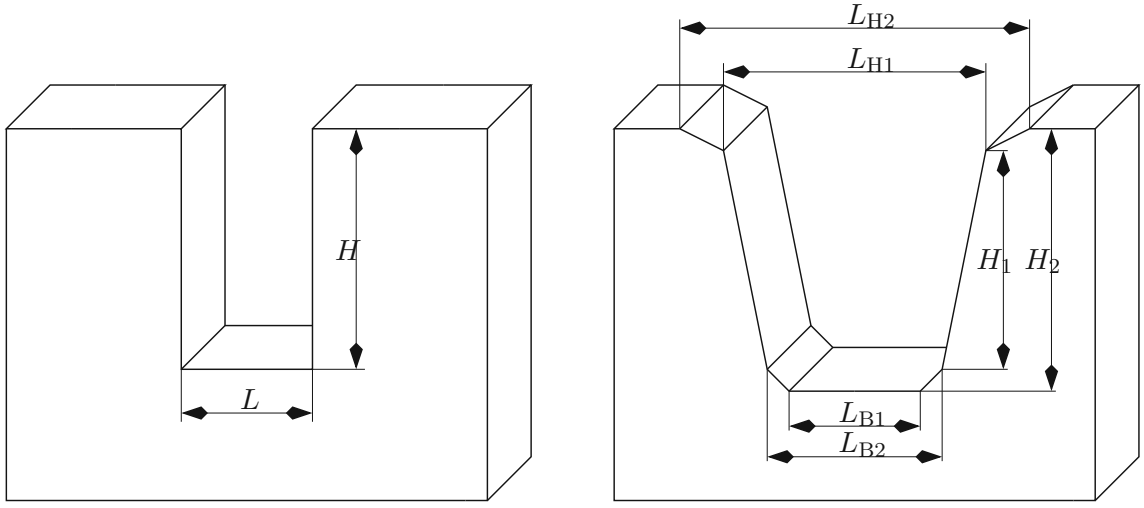
However, fabrication processes in general do not provide rectangular-shaped trenches because etching processes always produce tilted side walls and in the entry section of the trench at the top there is often an additional rounding according to surface diffusion during material deposition. Nevertheless, the impact of the tilted side walls on the width and the height of the particular trench can often be neglected with respect to the overall dimensions of the trench. Then, a rectangular shape can be justified. However, if the trench can not sufficiently be approximated by a rectangle, a different metric has to be applied to provide a characteristic criterion to compare different trenches using a different calculation of AR

$$\mathrm{AR} = \frac{H_\mathrm{m}}{L_\mathrm{m}}, \tag{A.13}$$

where the height and width are approximated by the mean value H_m and L_m, respectively. For instance, the height and the width can be approximated by their mean values as

$$H_\mathrm{m} = \frac{1}{2}\left(H_1 + H_2\right), \tag{A.14}$$

$$L_\mathrm{m} = \frac{1}{2}\left(\frac{1}{2}\left(L_{\mathrm{B1}} + L_{\mathrm{B2}}\right) + \frac{1}{2}\left(L_{\mathrm{H1}} + L_{\mathrm{H2}}\right)\right). \tag{A.15}$$



(a) rectangular-shaped trench (b) non-rectangular-shaped trench

Figure A.2: Differently shaped trenches with their characteristic dimensions which are used to determine the aspect ration (AR).

Appendix B

Timing

B.1 Duty Cycle

To determine the active operation time of a certain system, the duty cycle [349] of this system is used to calculate the power the signal has or the power the system uses. Figure B.1 shows

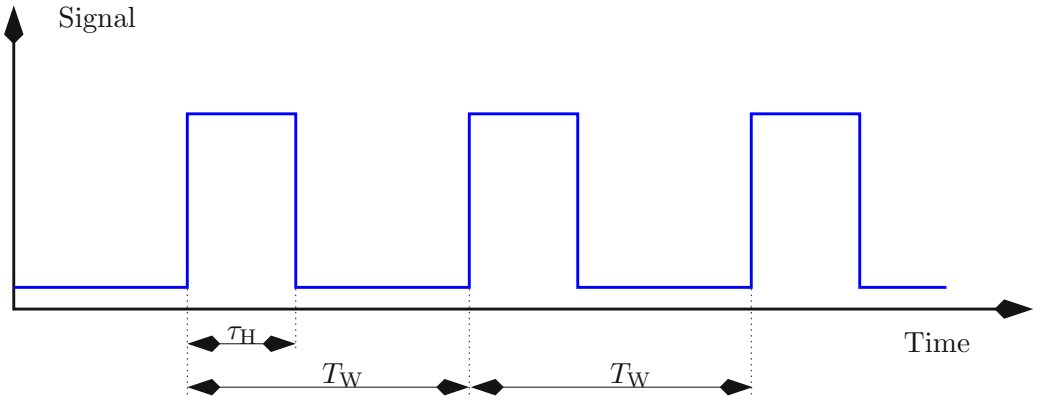


Figure B.1: Duty cycle

a periodic signal with a period duration T_W. During a certain time τ_H the signal is at its high state, hence it is active. Therefore, the duty cycle r_{DC} is defined according to [349] as the ratio between active pulse duration τ_H and the total period T_W:

$$r_{DC} = \frac{\tau_H}{T_W}. \tag{B.1}$$

B.2 Delay Times

An overview of the different types of delay times which occur in a switching process of a typical logic device is given in Figure B.2. Here, the output signal is retarded by a certain amount of

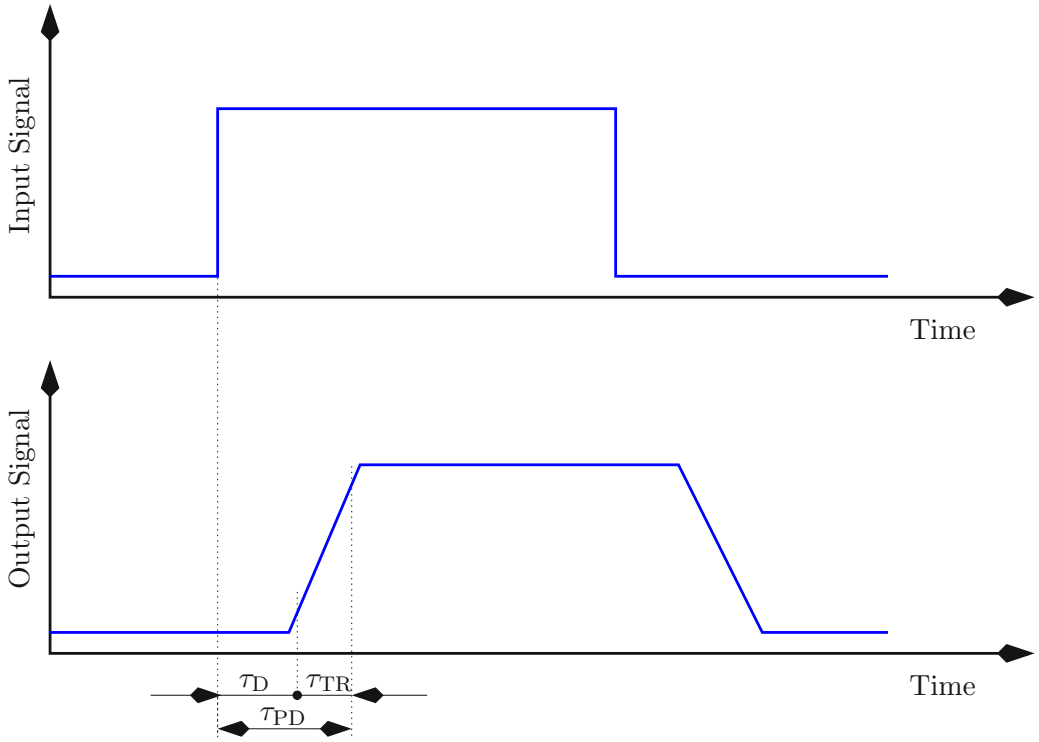


Figure B.2: Delay times

time τ_D. After this retardation, the device changes its state and switches. This time is called transition time τ_{TR} and is counted from the time where the output signal first reaches 10 % of its maximum output signal until the time it reaches 90 % of its stationary maximum signal level. The sum of both time periods plus any decaying times due to signal overshoot is called propagation delay τ_{PD}.

Other authors (cf. [91]) propose to measure the transition time from the 50 % reference voltage level for both the rising and the falling edges.

Bibliography

[1] P.J. Mohr and B.N. Taylor, "CODATA Recommended Values of the Fundamental Physical Constants: 2002", *Rev.Mod.Phys.*, vol. 77, no. 1, pp. 1–107, 2005.

[2] Intel Corporation, "Microprocessor Hall of Fame", 2004, `http://www.intel.com`.

[3] Semiconductor Industry Association, "International Technology Roadmap for Semiconductors - 2005 Edition", 2005, `http://public.itrs.net`.

[4] K. Banarjee, A. Mehrotra, A. Sangiovanni-Vincentelli, and C.M. Hu, "On Thermal Effects in Deep Sub-Micron VLSI Interconnects", in *Proc. 1999 IEEE/ACM International Conference on Computer-Aided Design*, San Jose, CA, 1999, pp. 885–891.

[5] K. Banarjee and A. Mehrotra, "Coupled Analysis of Electromigration Reliability and Performance in ULSI Signal Nets", *Proc. 2001 International Conference on Computer-Aided Design*, pp. 158–164, 2001.

[6] G.E. Moore, "Cramming More Components Onto Integrated Circuits", *Electronics*, pp. 114–117, Apr. 1965.

[7] R.R. Schaller, "Moore's Law: Past, Present, and Future", *IEEE Spectrum*, vol. 34, no. 6, pp. 53–59, June 1997.

[8] G.E. Moore, "Progress in Digital Integrated Electronics", *Proc.IEDM Tech.Dig*, vol. 19, pp. 11–13, 1975.

[9] G.E. Moore, "Lithography and the Future of Moore's Law", in *Optical/Laser Microlithography VIII*, T.A. Brunner, Ed. 1995, vol. 2440, pp. 2–17, SPIE.

[10] Semiconductor Industry Association, "International Technology Roadmap for Semiconductors - 2003 Edition", 2003, `http://public.itrs.net`.

[11] Semiconductor Industry Association, "International Technology Roadmap for Semiconductors - 2004 Update", 2004, `http://public.itrs.net`.

[12] K. Banerjee and A. Mehrotra, "Global (Interconnect) Warming", *IEEE Spectrum*, pp. 16–32, Sept. 2001.

[13] J.S. Kilby, "Miniaturized Self-Contained Cicuit Module and Method of Fabrication", U.S. Patent US03138744, filed May 6, 1959; issued June 23, 1964.

[14] J.S. Kilby, "Semiconductor Device", U.S. Patent US03643138, Filed January 29, 1962; issued February 15, 1972.

[15] L. Smith, V. Moroz, G. Eneman, P. Verheyen, F. Nouri, L. Washington, M. Jurczak, D. Pramanik, and K. De Meyer, "Exploring the Limits of Stress-Enhanced Hole Mobility", *IEEE Electron Device Lett.*, vol. 26, no. 9, pp. 652–654, 2005.

[16] F. Nouri, P. Verheyen, L. Washington, V. Moroz, I. De Wolf, M. Kawaguchi, S. Biesemans, R. Schreutelkamp, Y. Kim, M. Shen, X. Xu, R. Rooyackers, M. Jurcak, G. Eneman, K. De Meyer, L. Smith, D. Pramanik, H. Forstner, S. Thirupapuliyur, and G.S. Higashi, "A Systematic Study of Trade-offs in Engineering a Locally Strained pMOSFET", in *Proc.IEDM Tech.Dig* [107], pp. 1055–1058, 0-7803-8684-1.

[17] A. Gehring and S. Selberherr, "Gate Current Modeling for MOSFETs", *Journal of Computational and Theoretical Nanoscience*, vol. 2, no. 1, pp. 26–44, 2005.

[18] A. Gehring, *Simulation of Tunneling in Semiconductor Devices*, Dissertation, Technische Universität Wien, 2003, `http://www.iue.tuwien.ac.at/phd/gehring`.

[19] T.D. Sullivan, "Reliability Considerations for Copper Metallizations in ULSI Circuits", in *Conf.Proc. American Institute of Physics*, 1999, vol. 491, pp. 39–50.

[20] A. Isobayashi, Y. Enomoto, H. Yamada, S. Takahashi, and S. Kadomura, "Thermally Robust Cu Interconnects with Cu-Ag Alloy for sub 45nm Node", in *Proc.IEDM Tech.Dig* [107], pp. 953–956, 0-7803-8684-1.

[21] D. Ryuzaki, H. Sakurai, K. Abe, K. Takeda, and H. Fukada, "Enhanced Dielectric-Constant Reliability of Low-κ Porous Organosilicate Glass ($\kappa = 2.3$) for 45-nm-Generation Cu Interconnects", in *Proc.IEDM Tech.Dig* [107], pp. 949–952, 0-7803-8684-1.

[22] D.S. Yu, A. Chin, C.C. Laio, C.F. Lee, C.F. Cheng, W.J. Chen, C. Zhu, M.-F. Li, W.J. Yoo, S.P. McAlister, and D.L. Kwong, "3D GOI CMOSFETs with novel IrO_2(Hf) Dual Gates and high-κ Dielectric on 1P6M-18μm-CMOS", in *Proc.IEDM Tech.Dig* [107], pp. 181–184, 0-7803-8684-1.

[23] M. Miyoshi, A. Imanishi, H. Ishikawal, T. Egawal, K. Asai, M. Mouri, T. Shibata, M. Tanaka, and O. Oda, "Growth and Characterization of AlGaN/AlN/GaN HEMTs on 100-mm-diameter Epitaxial AlN/Sapphire Templates", in *Proc.IEDM Tech.Dig* [107], pp. 1031–1034, 0-7803-8684-1.

[24] V. Gutmann and E. Hengge, *Anorganische Chemie*, VCH Verlagsgesellschaft mbH, 1990.

[25] W. Harth, *Halbleitertechnologie*, Teubner, 1981.

[26] J.S. Reid, X. Sun, E. Kolawa, and M.-A. Nicolet, "Ti-Si-N Diffusion Barriers Between Silicon and Copper", *IEEE Electron Device Lett.*, vol. 15, no. 8, pp. 298–300, 1994.

[27] S.S. Ang and W.D. Brown, "Tantalum Oxide Dielectric for Embedded Capacitor Applications", in *Proc. 6th Intl. Conf. on Properties and Application of Dielectric Materials*, June 2000, pp. 841–844.

[28] J.A. Mandarino and M.E. Back, *Fleischer's Glossary of Mineral Species 2004*, Mineralogical Record, ninth edition edition, 2004.

[29] K. Hieda, K. Eguchi, N. Fukushima, T. Aoyama, K. Natori, M. Kiyotoshi, S. Yamazaki, M. Izuha, S. Niwa, Y. Fukuzumi, Y. Ishibashi, Y. Kohyama, T. Arikado, and K. Okumura, "All Perovskite Capacitor (APEC) Technology for (Ba,Sr)TiO_3 Capacitor Scaling Toward 0.10 μm Stacked DRAMs", in *Proc.IEDM Tech.Dig*, San Francisco, CA, Dec. 1998, pp. 807–810, 0-.

[30] U. Tietze and C. Schenk, *Halbleiterschaltungstechnik*, Springer-Verlag, 10 edition, 1993.

[31] S.M. Sze, Ed., *VLSI Technology*, McGraw-Hill, New York, second edition, 1988.

[32] G. Schindler, "Reliability of Narrow Copper Interconnects", in *Semicon 2005*, Munich, Germany, Mar. 2005.

[33] O. Aubel, *Ultrahochbeschleunigte Lebensdauertests an modernen Kupfer-Leiterbahnsystemen in höchstintegrierten Schaltungen*, Dissertation, Universität Hannover, 2004.

[34] J.Y. Park, Y.H. Lee, J.W. Bae, G.Y. Yeom, and J.H. Song, "MoSi(N) As A Diffusion Barrier Between Cu And Si", in *Intl. Microprocesses and Nanotechnology Conference*, July 1998, pp. 248–250.

[35] S.C. Sun, H.K. Yap, C.A. Chen, and P. Lin, "Evaluation of Amorphous (Ta,W,Mo)-Si-N Diffusion Barriers Between Cu and Si", in *Proc. 6th Intl. Conf. on Solid-State and Integrated-Circuit Technology*, Oct. 2001, vol. 1, pp. 397–399.

[36] H. Ceric, *Numerical Techniques in Interconnect and Process Simulation*, Dissertation, Technische Universität Wien, 2005, `http://www.iue.tuwien.ac.at/phd/ceric`.

[37] Institut for Microelectronics, *MINIMOS-NT 2.1 User's Guide*, Technische Universität Wien, Austria, 2004, `http://www.iue.tuwien.ac.at/software`.

[38] ISE Integrated Systems Engineering AG, Zürich, Switzerland, *DESSIS-ISE, ISE TCAD Release 9.0*, Aug. 2003.

[39] Institute for Microelectronics, *The Smart Analysis Programs*, Technische Universität Wien, Austria, 2003, `http://www.iue.tuwien.ac.at/software`.

[40] Synopsys, *TCAD Sentaurus Y-2006.06 User Manual*, Synopsys, Mountain View, 2006.

[41] Synopsys, *Taurus Work Bench User Manual*, Synopsys, Mountain View, 2003.

[42] ISE Integrated Systems Engineering AG, Zürich, Switzerland, *GENESISe, ISE TCAD Release 10.0*, 2004.

[43] ISE Integrated Systems Engineering AG, Zürich, Switzerland, *OptimISE, ISE TCAD Release 10.0*, 2004.

[44] Institute for Microelectronics, *SIESTA – The Simulation Environment for Semiconductor Technology Analysis*, Technische Universität Wien, Austria, 2003, `http://www.iue.tuwien.ac.at/software`.

[45] N. Konofaos, T. Voilas, and G.P. Alexiou, "Design and Simulation of an Embedded DRAM Cell Made Up With MOSFETs Having Alternative Gate Dielectrics", in SPIE 2005 [350], pp. 598–604.

[46] J. Ida, M. Yoshimaru, T. Usami, A. Ohtomo, K. Shimokawa, A. Kita, and M. Ino, "Reduction of Wiring Capacitance With New Low Dielectric SiOF Interlayer Film for High Speed/Low Power Sub-Half Micron CMOS", in *1994 Symposium on VLSI, Tech.Dig.*, 1994, pp. 59–60.

[47] S. Holzer, R. Minixhofer, C. Heitzinger, J. Fellner, T. Grasser, and S. Selberherr, "Extraction of Material Parameters Based on Inverse Modeling of Three-Dimensional Interconnect Fusing Structures", *Microelectronics Journal*, vol. 35, no. 10, pp. 805–810, 2004.

[48] R. Plasun, *Optimization of VLSI Semiconductor Devices*, Dissertation, Technische Universität Wien, 1999, `http://www.iue.tuwien.ac.at/phd/plasun`.

[49] R. Plasun, M. Stockinger, R. Strasser, and S. Selberherr, "Simulation Based Optimization Environment and Its Application to Semiconductor Devices", in *Intl. Conf. on Applied Modelling and Simulation*, Honolulu, HI, Aug. 1998, pp. 313–316.

[50] R. Strasser, R. Plasun, M. Stockinger, and S. Selberherr, "Inverse Modeling of Semiconductor Devices", in *Proc. SIAM Conference on Optimization 1999*, Atlanta, GA, 1999, p. 77.

[51] Ch. Pichler, R. Plasun, R. Strasser, and S. Selberherr, "High-Level TCAD Task Representation and Automation", *IEEE J.Technology Computer Aided Design*, May 1997, `http://www.ieee.org/journal/tcad/accepted/pichler-may97/`.

[52] H. Pimingstorfer, S. Halama, and S. Selberherr, "A TCAD Environment for Process and Device Engineering", in *Proc. Int. Conf. on VLSI and CAD*, Seoul, Korea, Oct. 1991, pp. 280–283.

[53] M. Stockinger and S. Selberherr, "Automatic Device Design Optimization with TCAD Frameworks", in *Proc. 3rd Intl. Conf. on Modeling and Simulation of Microsystems*, San Diego, California, USA, Mar. 2000, pp. 1–6.

[54] H. Sato, K. Tsuneno, K. Aoyama, T. Nakamura, H. Kunitomo, and H. Masuda, "A New Hierarchical RSM for TCAD-based Device Design to Predict CMOS Development", in *Proc. IEEE 1995 Int. Conf. on Microelectronic Test Structures*, Mar. 1995, vol. 8, pp. 299–302.

[55] V. Muntés Mulero, J. Aguilar Saborit, J.L. Larriba Pey, and C. Zuzarte, "A Study of Execution Plan Aware Mutations for Genetic Cyclic Query Optimization", Tech. Rep., Universitat Politècnica de Catalunya (UPC), June 2003, UPC-DAC-2003-37.

[56] Technovest(Pty) Ltd., *The Grail Genetic Optimizer for TradeStation™ (GGO)*, `http://www.thegrailsystem.com`.

[57] J.W. von Goethe, *Faust. Zweiter Theil*, in *Goethes sämtliche Werke*, Bd. 3, Verlag der J. G. Cotta'schen Buchhandlung, 1875.

[58] H. Hofmann, *Das elektromagnetische Feld*, Springer, Wien, Austia, 3rd edition, 1986.

[59] A. Prechtl, *Vorlesungen über Theoretische Elektrotechnik*, Institut für Grundlagen und Theorie der Elektrotechnik, Technische Universität Wien, 1998.

[60] H.J. Dirschmid, *Einführung in die mathematischen Methoden der theoretischen Physik*, Vieweg, 1976.

[61] K.G. Denbigh, *The Thermodynamics of the Steady State*, Methuen & Co. LTD. London, 1950.

[62] S. Selberherr, *Analysis and Simulation of Semiconductor Devices*, Springer, Wien, New York, 1984.

[63] R. Stratton, "Diffusion of Hot and Cold Electrons in Semiconductor Barriers", *Physical Review B*, vol. 126, no. 6, pp. 2002–2014, 1962.

[64] D. Chen, E. Li, E. Rosenbaum, and S.M. Kang, "Interconnect Thermal Modeling for Accurate Simulation of Circuit Timing and Reliability", *IEEE Trans.Computer-Aided Design of Integrated Circuits and Systems*, vol. 19, pp. 197–205, 2000.

[65] G.K. Wachutka, "Rigorous Thermodynamic Treatment of Heat Generation and Conduction in Semiconductor Device Modeling", *IEEE Trans.Computer-Aided Design*, vol. 9, no. 11, pp. 1141–1149, Nov. 1990.

[66] K. Kells, *General Electrothermal Semiconductor Device Simulation*, Hartung-Gorre Verlag, Konstanz, 1994.

[67] N.W. Ashcroft and N.D. Mermin, *Solid State Physics*, W.B. Saunders, 1976.

[68] L. Boltzmann, *Lectures on Gas Theory*, Dover Publications, 1964.

[69] Ch. Kittel and H. Kroemer, *Physik der Wärme*, Oldenbourg, 1993.

[70] I. Priogine, *The End of Certainty: Time, Chaos and the New Laws of Nature*, The Free Press, 1997.

[71] H.S. Leff and A. Rex, Eds., *Maxwell's Demon: Entropy, Information, Computing*, Princeton University Press, 1990.

[72] P.A. Tipler, *Physik*, Spektrum Akademischer Verlag, 2004.

[73] D. Kondepudi and I. Prigogine, *Modern Thermodynamics - From Heat Engines to Dissipative Structures*, John Wiley & Sons, New York, 1998.

[74] R. Mayer, "Remarks on the Forces of Nature", in *Bioenergetics - the Molecular Basis of Biological Energy Transformations*. 1971, Benjamin Cummings.

[75] M. Planck, *Treatise on Thermodynamics*, Dover Publications, 1897.

[76] K.G. Denbigh, *The Principles of Chemical Equilibrium*, Methuen & Co. LTD. London, reprinted edition, 1957.

[77] Y. Zhang, J. Christofferson, A. Shakouri, D. Li, A. Majumdar, Y. Wu, R. Fan, and P. Yang, "Characterization of Heat Transfer Along Silicon Nanowire Using Thermoreflectance Technique", *IEEE Trans. Nanotechnology*, vol. 5, no. 1, pp. 67–74, 2006.

[78] D. Li, Y. Wu, P. Kim, L. Shi, P. Yang, and A. Majumdar, "Thermal Conductivity of Individual Silicon Nanowires", *J.Appl.Phys.*, vol. 83, pp. 2934–2936, 2003.

[79] A. Nathan and H. Baltes, *Microtransducer CAD*, Springer Verlag, Wien, New York, 1999.

[80] G.A. Slack, "Thermal Conductivity of Pure and Impure Silicon, Silicon Carbide, and Diamond", *J.Appl.Phys.*, vol. 35, no. 12, pp. 3460–3466, 1964.

[81] S.P. Gaur and D.H. Navon, "Two-Dimensional Carrier Flow in a Transistor Structure under Non-isothermal conditions", *IEEE Trans.Electron Devices*, vol. 23, no. 1, pp. 50–57, 1976.

[82] D.K. Sharma and K.V. Ramanathan, "Modeling Thermal Effects On MOS I-V Characteristics", *IEEE Electron Device Lett.*, vol. 4, no. 10, pp. 362–364, 1983.

[83] L. Onsager, "Reciprocal Relations in Irreversible Processes. I.", *Physical Review*, vol. 37, no. 2, pp. 405–426, 1931.

[84] L. Onsager, "Reciprocal Relations in Irreversible Processes. II.", *Physical Review*, vol. 38, no. 12, pp. 2265–2279, 1931.

[85] E. Mendoza, *Reflexions of the Motive Power of Fire by Sadi Carnot and other Papers on the Second Law of Thermodynamics by É. Clapeyron and R. Clausius*, Dover Publications, Inc., Mineola, New York, 2005.

[86] L. Boltzmann, "Zur Theorie der thermoelektrischen Erscheinungen", *Wiener Berichte*, vol. 96, pp. 1258–1297, 1887.

[87] M. Knaipp, *Modellierung von Temperatureinflüssen in Halbleiterbauelementen*, Dissertation, Technische Universität Wien, 1998, `http://www.iue.tuwien.ac.at/phd/knaipp`.

[88] W.R. Tonti, J.A. Fifield, J. Higgins, W.H. Guthrie, W. Berry, and C. Narayan, "Reliability and Design Qualification of A Sub-Micron Tungsten Silicide E-Fuse", in *Proc. 42nd Annual Intl. Reliability Physics Symposium*, Apr. 2004, pp. 152–156.

[89] J. Fellner, P. Boesmueller, and H. Reiter, "Lifetime Study for a Poly Fuse in a 0.35μm Polycide CMOS Process", in *Proc. 43rd Annual Intl. Reliability Physics Symposium*, Apr. 2005, pp. 446–449.

[90] R. Minixhofer, S. Holzer, C. Heitzinger, J. Fellner, T. Grasser, and S. Selberherr, "Optimization of Electrothermal Material Parameters Using Inverse Modeling", in *Proc. 33rd European Solid-State Device Research Conference (ESSDERC 2003)*, José Franca and Paulo Freitas, Eds., Estoril, Portugal, Sept. 2003, pp. 363–366, IEEE.

[91] Y. Taur and T.H. Ning, *Fundamentals of Modern VLSI Devices*, Cambridge University Press, 1998.

[92] R. Sabelka and S. Selberherr, "A Finite Element Simulator for Three-Dimensional Analysis of Interconnect Structures", *Microelectronics Journal*, vol. 32, pp. 163–171, 2001.

[93] R. Sabelka and S. Selberherr, "SAP — A Program Package for Three-Dimensional Interconnect Simulation", in IITC'98 [351], pp. 250–252.

[94] A. Nentchev, R. Sabelka, W. Wessner, and S. Selberherr, "On Chip Interconnect Simulation of Parasitic Capacitances in Periodic Structures", in *The 2005 European Simulation and Modelling Conference Proceedings*, Oct. 2005, pp. 420 – 424.

[95] B. Wunderle, R. Schacht, O. Wittler, B. Michel, and E. Reichl, "Thermal Performance, Mechanical Reliability and Technological Features of Different Cooling Concepts for High Power Chip Modules", in *9th Intl. Workshop an THERMal INvestigations of ICs and Systems* [352], pp. 59–64, 0-7803-7999-3.

[96] F. Ziegler, *Mechanics of Solids and Fluids*, Springer, New York, 1995.

[97] O.C. Zienkiewicz, *The Finite Element Method: Basic Formulation and Linear Problems*, McGraw-Hill, England, 1987.

[98] T. Belytschko, W.K. Liu, and B. Moran, *Finite Elements for Nonlinear Continua and Structures*, Wiley, 1996.

[99] H. Ceric, A. Hoessinger, T. Binder, and S. Selberherr, "Modeling of Segregation on Material Interfaces by Means of the Finite Element Method", in *Proc. MATHMOD*, Wien, Austria, Feb. 2003, pp. 445–452.

[100] C. Hollauer, S. Holzer, H. Ceric, S. Wagner, T. Grasser, and S. Selberherr, "Investigation of Thermo-Mechanical Stress in Modern Interconnect Layouts", in *Sixth International Congress on Thermal Stresses*, Wien, Austria, May 2005, pp. 637–640.

[101] H.S .Yang, R. Malik, S. Narasimha, Y. Li, R. Divakuruni, P. Agnello, S. Allen, A. Antreasyan, J.C. Arnold, K. Bandy, M. Belyansky, A. Bonnoit, G. Bronner, V. Chan, X. Chen, Z. Chen, D. Chidambarrao, A. Chou, W. Clark, S.W. Crowder, B. Engel, H. Harifuchi, S.F. Huang, R. Jagannathan, F.F. Jamin, Y. Kohyama, H. Kuroda, C.W. Lai, H.K. Lee, W-H. Lee, E.H. Lim, W. Lai, A. Mallikarjunan, K. Matsumoto, A. McKnight, J. Nayak, H.Y. Ng, S. Panda, R. Rengarajan, M. Steigerwalt, S. Subbanna, K. Subramanian, J. Sudijono, G. Sudo, S.-P. Sun, B. Tessier, Y. Toyoshima, P. Tran, R. Wise, R. Wong, I.Y. Yang, C.H. Wann, and L.T. Su, "Dual Stress Liner for High Performance sub-45nm Gate Length SOI CMOS Manufacturing", in *Proc.IEDM Tech.Dig* [107], pp. 1075–1078, 0-7803-8684-1.

[102] S. Dhar, H. Kosina, V. Palankovski, E. Ungersböck, and S. Selberherr, "Electron Mobility Model for Strained-Si Devices", *IEEE Trans.Electron Devices*, vol. 52, no. 4, pp. 527–533, 2005.

[103] S. Smirnov, *Physical Modeling of Electron Transport in Strained Silicon and Silicon-Germanium*, Dissertation, Technische Universität Wien, 2003, `http://www.iue.tuwien.ac.at/phd/smirnov`.

[104] T. Ayalew, *SiC Semiconductor Devices Technology, Modeling, and Simulation*, Dissertation, Technische Universität Wien, 2004, `http://www.iue.tuwien.ac.at/phd/ayalew`.

[105] B. Witzigmann, V. Laino, M. Luisier, U. T. Schwarz, H. Fischer, G. Feicht, W. Wegscheider, C. Rumbolz, A. Lell, and V. Härle, "Analysis of temperature-dependent optical gain in GaN-InGaN quantum-well structures", *IEEE Photonics Technology Lett.*, vol. 18, no. 15, pp. 1600–1602, 2006.

[106] C.X. Lian, X.Y. Li, and J. Liu, "Optical Anisotropy of Wurtzite GaN on Sapphire Characterized by Spectroscopic Ellipsometry", *Semicond. Sci. Technol.*, vol. 19, pp. 417–420, 2004.

[107] *Proc.IEDM Tech.Dig*, San Francisco, CA, Dec. 2004, 0-7803-8684-1.

[108] *Proc.IEDM Tech.Dig*, San Francisco, CA, Dec. 2005, 0-7803-9269-8.

[109] S.H. Wemple, A. Jayaraman, and M. DiDomenico Jr., "Evidence from Pressure Experiments for Electron Scattering by the Ferroelectric Lattice Mode in ABO_3 Semiconductors", *Phys.Rev.Lett.*, vol. 17, no. 3, pp. 142–145, 1966.

[110] E.A. Avallone and T. Baumeister, *Mark's Standard Handbook for Mechanical Engineers*, McGraw-Hill, New York, 1996.

[111] W.C. Young and R.G. Budynas, *Roark's Formulas for Stress and Strain*, McGraw-Hill, New York, 7th edition, 2002.

[112] N.W. Ashcroft and N.D. Mermin, *Solid State Physics*, Harcourt College Publishers, 1976.

[113] S.M. Sze, Ed., *Semiconductor Devices: Pioneering Papers*, World Scientific, 1991.

[114] E. Liniger, L. Gignac, C.-K. Hu, and S. Kaldor, "In Situ Study of Void Groth Kinetics in Electroplated Cu Lines", *J.Appl.Phys.*, vol. 92, no. 4, pp. 1803–1810, 2002.

[115] L. Arnaud, T. Berger, and G. Reimhold, "Evidence of Grain-Boundary Versus Interface Diffusion in Electromigration Experiments in Copper Damascene Interconnects", *J.Appl.Phys.*, vol. 93, no. 1, pp. 192–204, 2003.

[116] H. Watanabe, "Statistics of Grain Boundaries in Gate poly-Si", in *Proc. Simulation of Semiconductor Processes and Devices*, Tokyo, Japan, Sept. 2005, pp. 39–42.

[117] Electronic Industries Association, "A Procedure for Executing SWEAT", *JEP119*, 1994.

[118] Electronic Industries Association, "Isothermal electromigration Test Procedure", *JESD61*, 1997.

[119] J.O. Weidner, *Charakterisierung von Zuverlässigkeit von subm-µm AlSiCu/TiN/Ti/n-Si-Kontakten bei hochbeschnleungiten Lebensdauertests*, Dissertation, Universität Hannover, 1999.

[120] K. Banarjee, S.J. Souri, P. Kapur, and K.C. Saraswat, "3-D ICs: A Novel Chip Design for Improving Deep-Submicrometer Interconnect Performance and Systems-On-Chip Integration", *Proceedings of the IEEE*, vol. 89, pp. 602–633, 2001.

[121] W. Wu, S.H. Brongersma, M. Van Hove, and K. Maex, "Influence of Surface and Grain-Boundary Scattering on the Resistivity of Copper in Reduced Dimensions", *Appl.Phys.Lett.*, vol. 84, no. 15, pp. 2838–2840, 2004.

[122] K. Fuchs, "The Conductivity of Thin Metallic Films According to the Electron Theory of Metals", *Proc. Cambridge Philosophical Society*, vol. 34, pp. 100–108, 1938.

[123] G.B. Alers, J. Sukamto, S. Park, G. Harm, and J. Reid, "Containing the Finite Size Effect in Copper Lines", *Semiconductor Intl.*, vol. 5, pp. 38–42, 2006.

[124] E. Lee, N. Truong, B. Prater, and J. Kardokus, "Copper Alloys and Alternative Barriers for Sub-45 nm Nodes", *Semiconductor Intl.*, vol. 7, pp. 1–8, 2006.

[125] E.H. Sondheimer, "The Influence of a Transverse Magnetic Field on the Conductivity of Thin Metallic Films", *Phys. Rev.*, vol. 80, pp. 401–406, 1950.

[126] J.C. Anderson, *The Use of Thin Films in Physical Investigations*, Academic Press, 1966.

[127] R.B. Fair, *Rapid Thermal Processing: Science and Technology*, Academic Press, 1993.

[128] G. Schindler, "Mechanical Weakness of Barrier Layers With Respect to Electro-Migration in Copper", Priv.Comm., 2005.

[129] J. Koike and M. Wada, "Self-Forming Diffusion Barrier Layer in Cu-Mn Alloy Metallization", *Appl.Phys.Lett.*, vol. 87, no. 4, pp. 041911, 2005.

[130] T. Usui, H. Nasu, J. Koike, M. Wada, S. Takahashi, N. Shimizu, T. Nishikawa, A. Yoshimaru, and H. Shibata, "Low Resistive and Highly Reliable Cu Dual-Damascene Interconnect Technology Using Self-Formed $MnSi_xO_y$ Barrier Layer", in *2005. Proceedings of the IEEE 2005 International Interconnect Technology Conference*, San Francisco, CA, June 2005, pp. 188–190, 0-7803-8752-X.

[131] L.H. van Vlack, *Elements of Materials Science and Engineering*, Addison-Wesley, 1985.

[132] G.S. Brady, H.R. Clauser, and J.A. Vaccari, *Materials Handbook*, McGraw-Hill, 14$^{\text{th}}$ edition, 1997.

[133] M. Bauccio, *ASM Engineering Materials Reference Book*, ASM International, 2$^{\text{nd}}$ edition, 1994.

[134] E.P. DeGarmo, J.T. Black, and R.A. Kohser, *Materials and Processes in Manufacturing*, Prentice Hall, 8$^{\text{th}}$ edition, 1997.

[135] K.G. Budinski and M.K. Budinski, *Engineering Materials: Properties and Selection*, Prentice Hall, 6$^{\text{th}}$ edition, 1999.

[136] C.J. Glassbrenner and G.A. Slack, "Thermal Conductivity of Silicon and Germanium from 3°K to the Melting Point", *Physical Review*, vol. 134, no. 4A, pp. A1058–A1069, May 1964.

[137] A.D. McConnell, S. Uma, and K.E. Goodson, "Thermal Conductivity of Doped Polysilicon Layers", in *Proc. of the Intl. Conference on Heat Transfer and Transport Phenomena in Microscale Structures*, G.P. Celata *et al.*, Ed., New York, 2000, pp. 413–419, Begell House.

[138] S. Uma, A.D. McConnel, M. Asheghi, K. Kurabayashi, and K.E. Goodson, "Temperature-Dependent Thermal Conductivity of Undoped Polycrystalline Silicon Layers", *International Journal of Thermophysics*, vol. 22, no. 2, pp. 605–616, 2001.

[139] R.P. Gupta, "Theory of Electromigration in Noble and Transition Metals", *Physical Review B*, vol. 25, no. 8, pp. 5188–5196, 1982.

[140] G. Fasching, *Werkstoffe für die Elektrotechnik*, Springer, Wien, 1987, 2. Auflage.

[141] A. Wymyslowski, K. Friedel, J. Felba, and T. Falta, "An Experimental-Numerical Approach to Thermal Contact Resistance", in *9th Intl. Workshop an THERMal INvestigations of ICs and Systems* [352], pp. 161–172, 0-7803-7999-3.

[142] N.F. Mott and H. Jones, *The Theory of the Properties of Metals and Alloys*, Dover Publications, Inc, New York, 1958.

[143] S.P. Murarka, *Silicides for VLSI Applications*, Academic Press, 1983.

[144] Ed. R.A. Levy, *Reduced Thermal Processing for VLSI*, Plemum Press, New York, 1989.

[145] F.Y. Huang, J.X. Lu, D.M. Jiang, X.C. Wang, and N. Jiang, "A Novel Analytical Approach to Parameter Extraction for On-Chip Spiral Inductors Taking Into Account High-Order Parasitic Effects", *Solid-State Electronics*, vol. 50, no. 9-10, pp. 1557–1562, 2006.

[146] Y. Cao, X. Huang, D. Sylvester, T.-J. King, and C. Hu, "Impact of On-Chip Interconnect Frequency-Dependent R(f)L(f) on Digital and RF Design", *IEEE Trans. VLSI Systems*, vol. 13, no. 1, pp. 158–162, 2005.

[147] U.S. Ghoshal and L.N. Smith, "Skin Effects in narrow Copper Microstrips at 77K", *IEEE Trans. Mircowave Theory and Techniques*, vol. 35, no. 12, pp. 1788–1795, 1988.

[148] J. Noguchi, T. Oshima, U. Tanaka, H. Aoki, K. Sato, K. Ishikawa, T. Saito, N. Konishi, S. Hotta, S. Uno, and K. Kikushima, "Integration and Reliability Issues of Cu/SiOC Interconnect for ArF/90nm Node SoC Manufacturing", in *Proc.IEDM Tech.Dig*, San Francisco, CA, Dec. 2004, pp. 527–530, 0-7803-7873-3.

[149] W.F.A. Besling, V. Arnal, J.F. Guillaumond, C. Guedj, M. Broekaart, L. L. Chapelon, A. Farcy, L. Arnaud, and J. Torres, "Integration of ALD TaN Barriers in Porous Low-k Interconnect for the 45 nm Node and Beyond; Solution to Relax Electron Scattering Effect", in *Proc.IEDM Tech.Dig* [107], pp. 325–328, 0-7803-8684-1.

[150] R. Fox, O. Hinsinger, E. Richard, E. Sabouret, T. Berger, C. Goldberg, A. Humbert, G. Imbert, P. Brun, E. Ollier, C. Maurice, M. Guillermet, C. Monget, V. Plantier, H. Bono, M. Zaleski, M. Mellier, J.-P. Jacquemin, J. Flake, B.G. Sharma, L. Broussous, A. Farcy, V. Arnal, R. Gonella, S. Maubert, V. Girault, P. Vannier, D. Reber, A. Schussler, J. Mueller, and W. Besling, "High Performance k=2.5 ULK Backend Solution Using an Improved TFHM Architecture, Extendible to the 45 nm Technology Node", in *Proc.IEDM Tech.Dig* [108], pp. 81–84, 0-7803-9269-8.

[151] P. Bai, C. Auth, S. Balakrishnan, M. Bost, R. Brain, V. Chikarmane, R. Heussner, M. Hussein, J. Hwang, D. Ingerly, R. James, J. Jeong, C. Kenyon, E. Lee, S-H. Lee, N. Lindert, M. Liu, Z. Ma, T. Marieb, A. Murthy, R. Nagisetty, S. Natarajan, J. Neirynck, A. Ott, C. Parker, J. Sebastian, R. Shaheed, S. Sivakumar, J. Steigerwald, S. Tyagi, C. Weber, B. Woolery, A. Yeoh, K. Zhang, and M. Bohr, "A 65nm Logic Technology Featuring 35nm Gate Lengths, Enhanced Channel Strain, 8 Interconnect Layers, Low-k ILD and 0.57μm^2 SRAM Cell", in *Proc.IEDM Tech.Dig* [107], pp. 657–660, 0-7803-8684-1.

[152] M. Abe, M. Tada, H. Ohtake, N. Furutake, M. Narihiro, K. Arai, T. Takeuchi1, S. Saito, T. Taiji, K. Motoyama, Y. Kasama, K. Arita, F. Ito, H. Yamamoto, M. Tagami, T. Tonegawa, Y. Tsuchiya, K. Fujii, N. Oda, M. Sekine, and Y. Hayashi, "A Robust 45 nm-node, Dual Damascene Interconnects with High Quality Cu/barrier Interface by a Novel Oxygen Absorption Process", in *Proc.IEDM Tech.Dig* [108], pp. 77–80, 0-7803-9269-8.

[153] H. Geisler, H. Prinz, I. Zienert, J. Rinderknecht, M. Kiene, and E. Zschech, "Temperatur-Dependent Stress Measurements at Inlaid Copper Interconnect Lines", in Zschech et al. [353], pp. 277–287.

[154] S. Bothra, "Air Gap Dielectric in Self-Aligned Via Structures", U.S. Patent US06281585, filed November 11, 1999; issued August 28, 2001.

[155] M. Ben-Tzur, K. Ramkumar, C.A. Seams, and T.J. Rodgers, "Low-k Dielectric Layer With Air Gaps", U.S. Patent US06903002, filed September 11, 2002; issued June 7, 2005.

[156] S. Nitta, S. Purushothaman, S. Smith, M. Krishnan, D. Canapen, T. Dalton, W. Volksen, R.D. Miller, B. Herbst, C.K. Hu, E. Liniger, J. Lloyd, M. Lane, D.L. Rath, M. Colbum, and L. Gignac, "Successful Dual Damascene Integration of Extreme Low k Materials (k<2.0) Using a Novel Gap Fill Based Integration Scheme", in *Proc.IEDM Tech.Dig* [107], pp. 321–324, 0-7803-8684-1.

[157] G. Wiederhirn, T.J. Balk, G. Dehn J. Nucci, G. Richter, and E.Arzt, "Passivation Effects in Copper Thin Films", in Zschech et al. [353], pp. 185–191.

[158] M.A. Meyer, M. Grafe, H.-J. Engelmann, E. Langer, and E. Zschech, "Investigation of the Influence of the Local Microstructure of Copper Interconnects on Void Formation and Evolution during Electromigration Testing", in Zschech et al. [353], pp. 175–184.

[159] M. Nopper, "Copper and More!", *6. Dresdner Sommerschule Mikroelektronik*, pp. 233–247, 2005.

[160] A. Wirth, "Electroless NiMoP Films as Capping Layers for Copper Interconnects", in *Semicon 2005*, Munich, Germany, Mar. 2005.

[161] W. Fan and X.-G. Gong, "Superheated Melting of Grain Boundaries", *Physical Review B*, vol. 72, no. 6, pp. 064121, 2005.

[162] A.D. McConnell, S. Uma, and K.E. Goodson, "Conductivity of Doped Polysilicon Layers", *Micromechanical Systems*, vol. 10, no. 3, pp. 360–369, 2001.

[163] P.G. Shewmon, *Transformations in Metals*, McGraw-Hill, 1969.

[164] V. Shukharev, "Simulation of Microstructure Influence on EM-Induced Degradation in Cu Interconnects", in Zschech et al. [353], pp. 244–253.

[165] A. Sheikholeslami, *Topography Simulation of Deposition and Etching Processes*, Dissertation, Technische Universität Wien, 2006, `http://www.iue.tuwien.ac.at/phd/sheikholeslami`.

[166] H. Puchner, *Advanced Process Modeling for VLSI Technology*, Dissertation, Technische Universität Wien, 1996, `http://www.iue.tuwien.ac.at/phd/puchner`.

[167] A. Hössinger, *Simulation of Ion Implantation for ULSI Technology*, Dissertation, Technische Universität Wien, 2000, `http://www.iue.tuwien.ac.at/phd/hoessinger`.

[168] D. Tamboli, S. Chang, M. Evans, I. Butcher, Q. Arefeen, M. Waddell, and S. Hymes, "Electrochemical Processes in Metal Planarization Technologies", in ECS 2002-201 [354], pp. 114–125.

[169] A.S. Lawing, "Polish Rate, Pad surface Morphology and Pad Conditioning in Oxide Chemical Mechanical Polishing", in ECS 2002-201 [354], pp. 46–60.

[170] S.M. Sze, *Semiconductor Devices Physics and Technology*, John Wiley & Sons, 2002.

[171] T.S. Cale, T.P. Merchant, L.J. Borucki, and A.H. Labun, "Topography Simulation for the Virtual Wafer Fab", *Thin Solid Films*, vol. 365, no. 2, pp. 152–175, Apr. 2000.

[172] W. Pyka, *Feature Scale Modeling for Etching and Deposition Processes in Semiconductor Manufacturing*, Dissertation, Technische Universität Wien, 2000, `http://www.iue.tuwien.ac.at/phd/pyka`.

[173] A. Sherman, *Chemical Vapor Deposition for Microelectronics*, Noyes Publications, Park Ridge, New Jersey, 1987.

[174] S. Osher and J.A. Sethian, "Fronts Propagating with Curvature Dependent Speed: Algorithm Based on Hamilton-Jacobi Formulation", *J.Comput.Phys.*, vol. 28, pp. 907–922, 1991.

[175] H. Liao and T.S. Cale, "Low-Knudsen-Number Transport and Deposition", *J. Vac. Sci. Technol. A*, vol. 12, no. 4, pp. 1020–1026, 1994.

[176] S. Arrhenius, "On the Influence of Carbonic Acid in the Air Upon the Temperature of the Ground", *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science*, vol. 5, pp. 237–276, 1896.

[177] Z. Yu, B. Riccó, and R.W. Dutton, "A Comprehensive Analytical and Numerical Model of Polysilicon Emitter Contacts in Bipolar Transistors", *IEEE Trans.Electron Devices*, vol. 31, no. 6, pp. 773–784, 1984.

[178] M.M. Mandurah, K.C. Saraswat, and C.R. Helms, "Dopant Segregation in Polycrystalline Silicon", *J.Appl.Phys.*, vol. 51, no. 11, pp. 5755–5763, 1980.

[179] M.M. Mandurah, K.C. Saraswat, and T.I. Kamins, "A Model for Conduction in Polycrystalline Silicon–Part I: Theory", *IEEE Trans.Electron Devices*, vol. 28, no. 10, pp. 1163–1171, 1981.

[180] M.M. Mandurah, K.C. Saraswat, and T.I. Kamins, "A Model for Conduction in Polycrystalline Silicon–Part I: Comparison of Theory and Experiments", *IEEE Trans.Electron Devices*, vol. 28, no. 10, pp. 1171–1176, 1981.

[181] J.Y. W. Seto, "The Electrical Properties of Polycrystalline Silicon Thimn Films", *J.Appl.Phys.*, vol. 42, no. 12, pp. 5247–5254, 1975.

[182] C.M. Osburn, I. Kim, S.K. Han, I. De, K.F. Yee, S. Gannavaram, S.J. Lee, C.-H. Lee, Z.J. Luo, W. Zhu, J. R. Hauser, D.-L. Kwong, G. Lucovsky, T.P. Ma, and M.C. Öztürk, "Vertically Scaled MOSFET Gate Stacks and Junctions: How Far are we Likely to Go?", *IBM J.Res.Dev.*, vol. 46, no. 2/3, pp. 299–315, 2002.

[183] H.-S.P. Wong, "Beyond the Conventional Transistor", *IBM J.Res.Dev.*, vol. 46, no. 2/3, pp. 133–168, 2002.

[184] M. LeRoy, E. Lheurette, O. Vanbesien, and D. Lippens, "Wave-Mechanical Calculations of Leakage Current Through Stacked Dielectrics for nanotransistor Metal-Oxide-Semiconductor Design", *J.Appl.Phys.*, vol. 93, no. 5, pp. 2966–2971, 2003.

[185] G.D. Wilk, R.M. Wallace, and J.M. Anthony, "High-k Gate Dielectrics: Current Status and Materials Properties Considerations", *J.Appl.Phys.*, vol. 89, no. 10, pp. 5243–5275, 2001.

[186] J. Robertson, "Band Offsets of Wide-Bandgap Oxides and Implications for Future Electronic Devices", *J.Vac.Sci.Technol.*, vol. 18, no. 3, pp. 1785–1791, 2000.

[187] T. Hori, Y. Naito, H. Iwasaki, and H. Esaki, "Interface States and Fixed Charges in Nanometer-Range Thin Nitrided Oxides Prepared by Rapid Thermal Annealing", *IEEE Electron Device Lett.*, vol. 7, no. 12, pp. 669–671, 1986.

[188] T. Hori, H. Iwasaki, Y. Naito, and H. Esaki, "Electrical Nitrided and Physical Characteristics of Thin Oxides Prepared by Rapid Thermal Nitridation", *IEEE Trans.Electron Devices*, vol. 34, no. 11, pp. 2238–2244, 1987.

[189] J. Zhang, J.S. Yuan, Y. Ma, and A.S. Oates, "Design Optimization of Stacked Layer Dielectrics for Minimum Gate Leakage Currents", *Solid-State Electron.*, vol. 44, no. 12, pp. 2165–2170, 2000.

[190] J.D. Casperson, L.D. Bell, and H.A. Atwater, "Materials Issues for Layered Tunnel Barrier Structures", *J.Appl.Phys.*, vol. 92, no. 1, pp. 261–267, 2002.

[191] M. von Arx, O. Paul, and H. Baltes, "Process-Dependent Thin Film Thermal Conductivities for Thermal CMOS MEMS", *Microelectromechanical Systems*, vol. 9, no. 1, pp. 136–145, Mar. 2000.

[192] F. Völklein and H. Baltes, "A Microstructure for Measurement of Thermal Conductivity of Polycrystalline Thin Films", *Microelectromechanical Systems*, vol. 1, no. 4, pp. 193–196, 1992.

[193] N.R. Swart, *Heat Transport in Thermal-Based Microsensors*, Dissertation, University of Waterloo, 1994.

[194] O. Paul, M. von Arx, and H. Baltes, "Process-Dependent Thermophysical Properties of CMOS IC Thin Films", in *Digest of Technical Papers, vol. 1, Transducers '95,*, Stockholm, Sweden, 1995, pp. 178–181.

[195] Y.C. Tai, C.H. Mastrangelo, and R.S. Müller, "Thermal Conductivity of Heavily Doped LPCVD Polycrystalline Silicon Films", *J.Appl.Phys.*, vol. 63, no. 5, pp. 1442–1447, 1988.

[196] C.H. Shomate, "A Method for Evaluating and Correlating Thermodynamic Data", *J.Phys.Chem*, vol. 58, no. 4, pp. 368–372, 1954.

[197] A. Cezairliyan, *Specific Heat of Solids*, Hemisphere Publishing Corp., 1988.

[198] L.K. Nash, *Elements of Chemical Thermodynamics*, Dover Publications, Inc., Mineola, New York, 2nd edition, 2005.

[199] K. Yamaguchi and K. Itakagi, "Measurement of High Temperature Heat Content of Silicon by Drop Calorimetry", *J.Thermal.Anal.Cal.*, vol. 69, pp. 1059–1066, 2002.

[200] G. Wang, X. Zhang, and P.S. Ho, "Chip-Packaging Interaction and Reliability Impact on Cu/Low k Interconnects", in Zschech et al. [353], pp. 73–82.

[201] C.H. Pan and C.L. Tung, "A Simple Method for Determination of Thermal Conductivity Coefficients of Dielectric Films", in *Proc. 2001 Intl. Symposium Micromechatronics and Human Science*, Nagoya, Japan, 2001, pp. 109–116.

[202] C. Kittel, *Introduction to Solid State Physics*, Wiley, 7th edition, 1996.

[203] J. Hoekstra, A.P. Sutton, T.N. Todorov, and A.P. Horsfield, "Electromigration of Vacancies in Copper", *Physical Review B*, vol. 62, no. 13, pp. 8568–8571, 2000.

[204] M.J. Aziz, P.C. Sabin, and G.-Q. Lu, "The Activation Strain Tensor: Nonhydrostatic Stress Effects on Crystall-Growth Kinetics", *Physical Review A*, vol. 44, no. 18, pp. 9812–9816, 1991.

[205] P. Pichler, *Intrinsic Point Defects, Impurities, and Their Diffusion in Silicon*, Springer, Wien, NewYork, 2004.

[206] H.B. Huntington and A.R. Grone, "Current-Induced Marker Motion in Gold Wires", *J.Phys.Chem.Solids*, vol. 20, no. 1-2, pp. 76–87, 1961.

[207] M. Karpovski A. Gladkikh, A. Palevski, and Y.S. Kaganovski, "Effect of Microstructure on Electromigration Kinetics in Cu Lines", *J.Phys.D: Appl.Phys.*, vol. 31, no. 14, pp. 1626–1629, 1998.

[208] R. Kircheim, "Stress and Electromigration in Al-Lines of Integrated Circuits", *Acta Metallurg. et Mater.*, vol. 40, no. 2, pp. 309–323, 1992.

[209] J.R. Lloyd, C.E. Murray, T.M. Shaw, M.W. Lane, X.-H. Liu, and E.G. Liniger, "Theory for Electromigration Failure in Cu Conductors", in Zschech et al. [353], pp. 23–24.

[210] E.T. Ogawa, K.D. Lee, V.A. Blaschke, and P.S. Ho, "Electromigration Reliability Issues in Dual-Damascene Cu Interconnects", *IEEE Trans.Reliability*, vol. 51, no. 4, pp. 403–419, 2002.

[211] J.R. Lloyd, J. Clemens, and R. Snede, "Copper Metalization Reliability", *Microelectron.Reliab.*, vol. 39, no. 11, pp. 1595–1602, 1999.

[212] O. Pierre-Louis and T.L. Einstein, "Electromigration of Single-Layer Clusters", *Physical Review B*, vol. 62, no. 20, pp. 1829–1838, 2000.

[213] M. Hauschildt, M. Gall, S. Thrasher, P. Justison, L. Michaelson, R. Hernandez, H. Kawasaki, and P.S. Ho, "Statistical Analysis of Electromigration Lifetimes for Cu Interconnects", in Zschech et al. [353], pp. 164–174.

[214] Z.H. Gan, W. Shao, M.Y. Yan, A.V. Vairagar, T. Zaporozhets, M.A. Meyer, A. Krishnamoorthy, K.N. Tu, A. Gusak, E. Zschech, and S.G. Mhaisalkar, "Understanding the Impact of Surface Engineering, Structure, and Design on Electromigration through Monte Carlo Simulation and In-Situ SEM Studies", in Zschech et al. [353], pp. 34–42.

[215] J.R. Black, "Electromigration - A Brief Survey and Some Results", *IEEE Trans.Electron Devices*, vol. 16, no. 4, pp. 338–347, 1969.

[216] J.R. Black, "Electromigration Failure Modes in Aluminum Mateallization for Semiconductor Devices", *Proceeding fo the IEEE*, vol. 57, pp. 1587–1693, 1969.

[217] D. Dalleau and K. Weide-Zaage, "Three-Dimensional Voids Simulation in Chip Metallization Structures: A Contribution to Reliability Evaluation", *Microelectron. Reliab.*, vol. 41, no. 9-10, pp. 1625–1630, 2001.

[218] C.F. Gauß, "Zur Paralellentheorie. Brief an Taurinus, 8. November 1824", *Werke*, Bd. 8: Arithmetik und Algebra: Nachträge zu Band 1-3, pp. 186–188, 1863.

[219] C. Großmann and J. Terno, *Numerik der Optimierung*, B. G. Teubner, Stuttgart, Germany, 1997.

[220] E.G. Birgin, J.S. Martínez, and M. Raydan, "Algorithm 813: SPG—Software for Convex-Constrained Optimization", *ACM Transactions on Mathematical Software*, vol. 27, no. 3, pp. 340–349, 2001.

[221] R. Fletcher, "Am Ideal Penalty Function for Constrained Optimization", *J.Inst.Math.Appl.*, vol. 15, pp. 319–342, 1975.

[222] R. Fletcher, *Practical Optimization Methods*, Wiley, 2nd edition, 1987.

[223] D.P. Bertsekas, "Necessary and Sufficient Conditions for a Penalty Method to be Exact", *Math.Programming*, vol. 9, pp. 87–99, 1975.

[224] A.V. Fiacco and G.P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, Wiley, 1968.

[225] E.J. Beltrani, *An Alogrithmic Approach to Nonlinear Analysis and Optimization*, Academic Press, 1970.

[226] K.F. Frisch, *The Logarithmic Potential Method of Convex Programming*, Memorandum of May 13, University Institute of Economics, Oslo, 1955.

[227] L. Bittner, "Eine Verallgemeinerung des Verfahrens des logarithmischen Potentials von Frisch für nicht lineare Optimierungsaufgaben", in *Colloqu. on applic.math.to econ.*, 1965, pp. 43–53.

[228] F.A. Lootsma, *Boundary Properties of Penalty Functions for Constrained Minimization*, Philips Res.Rept.Suppl., 3, 1970.

[229] C. Großmann and A.A. Kaplan, *Strafmethoden und modifizierte Lagrange-Funktionen in der nichtlinearen Optimierung*, Teubner, 1979.

[230] C.T. Kelley, *Iterative Methods for Optimization*, SIAM, Philadelphia, 1999.

[231] Yu. Ermoliev and R.J.-B. Wets, Eds., *Numerical Techniques for Stochastic Optimization*, vol. 10 of *Springer Series in Computational Mathematics*, Springer, 1988.

[232] Y. Censor and S.A. Zenios, *Parallel Optimization: Theory, Algorithms, and Applications*, Oxford University Press, Oxford, 1997.

[233] P.E. Gill, W. Murray, and M.H. Wright, *Practical Optimization*, Academic Press, 1995.

[234] P. Lloyd, C.C. McAndrew, M.J. McLennan, S. Nassif, K. Singhal, Ku. Singhal, P.M. Zeitzoff, M.N. Darwish, K. Haruta, J.L. Lentz, H. Vuong, M.R. Pinto, C.S. Rafferty, and I.C. Kizilyalli, "Technology CAD at AT&T", in *Technology CAD Systems*, F. Fasching, S. Halama, and S. Selberherr, Eds., Wien, 1993, pp. 1–24, Springer.

[235] R. Minixhofer, *Integration Technology Simulation into the Semiconductor Manufacturing Environment*, Dissertation, Technische Universität Wien, 2006, `http://www.iue.tuwien.ac.at/phd/minixhofer`.

[236] T.G. Kolda, R.M. Lewis, and V. Torczon, "Optimization by Direct Search:New Perspectives on some Classical and Modern Methods", *SIAM Review*, vol. 45, no. 3, pp. 385–482, 2003.

[237] G.E.P. Box and K.B. Wilson, "On the Experimental Attainment of Optimum Conditions", *Journal of the Royal Statistical Society Series B*, vol. 13, no. 1, pp. 1–45, 1951.

[238] D.S. Boning and P.K. Mozumder, "DOE/Opt: A System for Design of Experiments, Response Surface Modeling, and Optimization Using Process and Device Simulation", *IEEE Trans.Semiconductor Manufacturing*, vol. 7, no. 2, pp. 233–244, 1994.

[239] Ch. Pichler, N. Khalil, G. Schrom, and S. Selberherr, "TCAD Optimization Based on Task-Level Framework Services", in *Simulation of Semiconductor Devices and Processes*, H. Ryssel and P. Pichler, Eds., Wien, Austria, 1995, vol. 6, pp. 70–73, Springer.

[240] R. Cartuyvels, R. Booth, S. Kubicek, L. Dupas, and K.M. De Meyer, "A Powerful TCAD System Including Advanced RSM Techniques for Various Engineering Optimization Problems", in *Simulation of Semiconductor Devices and Processes*, S. Selberherr, H. Stippel, and E. Strasser, Eds., Wien, 1993, vol. 5, pp. 29–32, Springer.

[241] V. Senez, T. Hoffmann, and A. Tixier, "Calibration of a Two-Dimensional Numerical Model for the Optimization of LOCOS-Type Isolations by Response Surface Methodology", *IEEE Trans.Semiconductor Manufacturing*, vol. 13, no. 4, pp. 416–426, Nov. 2000.

[242] D.W. Marquardt, "An Algorithm for the Estimation of Nonlinear Parameters", *J. Soc. Ind. Appl. Maths.*, vol. 11, pp. 431–441, 1963.

[243] C. Heitzinger, *Simulation and Inverse Modeling of Semiconductor Manufacturing Processes*, Dissertation, Technische Universität Wien, 2002, `http://www.iue.tuwien.ac.at/phd/heitzinger`.

[244] R. Pfeifer, *Effektive Messauswertung mit der Gauß'schen Fehlerquadratmethode*, Sport und Buch Strauß, 2001.

[245] J.J. Moré, B.S. Garbow, and K.E. Hillstrom, "Users Guide for MINPACK-1", 1980, Argonne National Laboratory Report ANL-80-74, Argone, IL.

[246] J.J. Moré, D.C. Sorensen, K.E. Hillstrom, and B.S. Garbow, *The MINPACK Project*, Sources and Development of Mathematical Software. Prentice-Hall, Englewood Clifs, NJ, 1984.

[247] *2nd Intl. Conf. on Modeling and Simulation of Microsystems*, San Juan, Puerto Rico, USA, Apr. 1999.

[248] R. Strasser, *Rigorous TCAD Investigations on Semiconductor Fabrication Technology*, Dissertation, Technische Universität Wien, 1999, `http://www.iue.tuwien.ac.at/phd/strasser`.

[249] P. Spellucci, *Donlp2 Users Guide*, NetLib Project, 1995.

[250] P. Spellucci, "An SQP Method For General Nonlinear Programs Using Only Equality Constrained Subproblems", *Mathematical Programming*, vol. 82, no. 3, pp. 413–448, 1998.

[251] G.H. Golub and C.F. Van Loan, *Matrix Computations*, John Hopkins University Press, second edition, 1989.

[252] G.H. Golub and D.P. O'Leary, "Some History of the Conjugate Gradient and Lanczos Algorithms: 1948-1976", *SIAM Review*, vol. 31, no. 1, pp. 50–102, 1989.

[253] M.R. Field, "Optimizing a Parallel Conjugate Gradient Solver", *SIAM J.Sci.Comput.*, vol. 19, no. 1, pp. 27–37, 1998, `http://www.siam.org/journals/sisc/19-1/30220.html`.

[254] P. Concus and G.H. Golub, "A Generalized Conjugate Gradient Method for Nonsymmetric Systems of Linear Equations", in *Proc. 2nd Int.Symp. on Computing Methods in Applied Sciences and Engineering*, R. Glowinski and J.L. Lions, Eds., Berlin, 1976, vol. 134 of *Lecture Notes in Economics and Mathematical Systems*, pp. 56–65, Springer.

[255] H.A. van der Vorst, "BI-CGSTAB: A Fast and Smoothly Converging Variant of BI-CG for the Solution of Nonsymmetric Linear Systems", *SIAM J.Sci.Stat.Comput.*, vol. 13, no. 2, pp. 631–644, 1992.

[256] P. Sonneveld, "CGS, A Fast Lanczos-Type Solver for Nonsymmetric Linear Systems", *SIAM J.Sci.Stat.Comput.*, vol. 10, no. 1, pp. 36–52, 1989.

[257] J. Gablonsky, "An Implementation of the DIRECT Algorithm", Tech. Rep., Department of Mathematics, North Carolina State University, 1998, `http://www4.ncsu.edu/eos/users/c/ctkelley/www/optimization_codes.html`.

[258] M. Björkman and K. Holström, "Global Optimization Using the DIRECT Algorithm in Matlab", *AMO - Advanced Modeling and Optimization*, vol. 1, no. 2, pp. 17–37, 1999.

[259] J. Holland, "Adaption in Natural and Artificial Systems", *University of Michigan Press, Ann Arbor, MI*, 1975.

[260] D.E. Goldberg, *Genetic Algorithms in Search and Optimization*, Addison-Wesley, 1989.

[261] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, Springer, Berlin, 1996.

[262] M. Wall, "GAlib A C++ Library of Genetic Algorithm Components", *Massachusetts Institute of Technology*, 2000, `http://lancet.mit.edu/ga`.

[263] T. Binder, *Rigorous Integration of Semiconductor Process and Device Simulators*, Dissertation, Technische Universität Wien, 2002, `http://www.iue.tuwien.ac.at/phd/binder`.

[264] R.C. Eberhardt and J. Kennedy, "A New Optimizer Using Particle Swarm Theory", in *Proceedings of the Sixth International Symposium on Micromachine and Human Science*, Nagoya, Japan, 1995, pp. 39–45.

[265] J. Kennedy and R.C. Eberhardt, "Particle Swarm Optimization", in *Proceedings of IEEE International Conference on Neural Networks*, Piscataway, NJ., 1995, pp. 1942–1948.

[266] M. Clerc, *"Discrete Particle Swarm Optimization", New Optimization Techniques in Engineering*, Springer, 2004.

[267] F. van den Bergh, *An Analysis of Particle Swarm Optimizers*, Phd thesis, Department of Computer Science, University of Pretoria, 2002.

[268] F. van den Bergh and A.P. Engelbrecht, "A Cooperative Approach to Particle Swarm Optimisation", *IEEE Transactions on Evolutionary Computation*, vol. 8, no. 3, pp. 225–239, 2004.

[269] F. Chang and D. Xue, "Optimal Concurrent Design Based Upon Distributed Product Development Life-Cycle Modeling", *Robotics and Computer-Integrated Manufacturing*, vol. 17, no. 6, pp. 469–486, 2001.

[270] M. Clerc and J. Kennedy, "The Particle Swarm-Explosion, Stability, and Convergence in a Multidimensional Complex Space", *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 1, pp. 58–73, 2002.

[271] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, and A. H. Teller, "Equation of State Calculations by Fast Computing Machines", *The Journal of Chemical Physics*, vol. 21, no. 6, June 1953.

[272] S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi, "Optimization by Simulated Annealing", *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[273] S. Kirkpatrick, "Optimization by Simulated Annealing: Quantitative Studies", *J.Stat.Phys.*, vol. 34, pp. 975–986, 1984.

[274] D. Kirkpatrick, "Optimal Search in Planar Subdivisions", *SIAM J.Computing*, vol. 12, no. 1, pp. 28–35, 1983.

[275] L. Ingber, "Very Fast Simulated Re-Annealing", *Mathematical Computer Modelling*, vol. 12, pp. 967–973, 1989, `http://www.ingber.com/asa89_vfsr.ps.gz`.

[276] L. Ingber, "Genetic Algorithms and Very Fast Simulated Re-Annealing: A Comparision", *Mathematical and Computer Modelling*, vol. 16, pp. 87–100, 1992, `http://www.ingber.com/asa92_saga.ps.gz`.

[277] V. Černy, "Thermodynamical Approach to the Traveling Salesman Problem: an Efficient Simulation Algorithm", *J. Opt. Theory Appl.*, vol. 45, pp. 41–45, 1985.

[278] N. Metropolis and S. Ulam, "The Monte Carlo Method", *Journal of the American Statistical Association*, vol. 44, no. 247, pp. 335–341, 1949.

[279] A. Abramo, L. Baudry, R. Brunetti, R. Castagne, M. Charef, F. Dessenne, P. Dollfus, R. Dutton, W.L. Engl, R. Fauquembergue, C. Fiegna, M.V. Fischetti, S. Galdin, N. Goldsman, M. Hackel, C. Hamaguchi, K. Hess, K. Hennacy, P. Hesto, J.M. Higman, T. Iizuka, C. Jungemann, Y. Kamakura, H. Kosina, T. Kunikiyo, S.E. Laux, H. Lin, C. Maziar, H. Mizuno, H.J. Peifer, S. Ramaswamy, N. Sano, P.G. Scrobohaci, S. Selberherr, M. Takenaka, T.-W. Tang, K. Taniguchi, J.L. Thobel, R. Thoma, K. Tomizawa, M. Tomizawa, T. Vogelsang, S.-L. Wang, X. Wang, C.-S. Yao, P.D. Yoder, and A. Yoshii, "A Comparison of Numerical Solutions of the Boltzmann Transport Equation for High-Energy Electron Transport Silicon", *IEEE Trans.Electron Devices*, vol. 41, no. 9, pp. 1646–1654, 1994.

[280] S. Wagner, V. Palankovski, T. Grasser, G. Röhrer, and S. Selberherr, "Direct Extraction Feature for Scattering Parameters of SiGe-HBTs", *Applied Surface Science*, vol. 224, no. 1-4, pp. 365–369, 2004.

[281] M. Wagner, G. Span, S. Holzer, V. Palankovski, O. Triebl, and T. Grasser, "Power Output Improvement of SiGe Thermoelectric Generators", *ECS Transactions: SiGe and Ge: Materials, Processing, and Devices*, vol. 3, no. 7, pp. 1151–1162, 2006.

[282] T. Binder, C. Heitzinger, and S. Selberherr, "A Qualitative Study on Global and Local Optimization Techniques for TCAD Analysis Tasks", in *Proc. Modeling and Simulation of Microsystems (MSM 2001)*, Hilton Head Island, South Carolina, USA, Mar. 2001, pp. 466–469.

[283] M. Burger and R. Pinnau, "Fast Optimal Design of Semiconductor Devices", *SIAM J.Appl.Math.*, vol. 64, no. 1, pp. 108–126, 2003.

[284] T.J. Lorenzen and V.L. Anderson, *Design of Experiments*, Marcel Dekker, New York, 1991.

[285] J.J. Moré and S.J. Wright, *Optimization Software Guide*, vol. 14 of *Frontiers in Applied Mathematics*, SIAM, Philadelphia, 1993.

[286] P. Fleischmann, E. Leitner, and S. Selberherr, "Optimized Geometry Preprocessing for Three-Dimensional Semiconductor Process Simulation", in *Intl. Conf. on Applied Modelling and Simulation* [355], pp. 317–321.

[287] D.M. Betz, *XLISP: An Object-Oriented Lisp, Version 2.1*, Apple, Peterborough, New Hampshire, USA, 1989.

[288] P.H. Winston and B.K.P. Horn, *Lisp*, Addison-Wesley, 1989.

[289] Institut für Mikroelektronik, Technische Universität Wien, Austria, *VISTA Documentation 1.3-1, VLISP Manual*, Jan. 1996.

[290] Python Software Foundation, *Python*, Jan. 2005.

[291] S. Wagner, *Small-Signal Device and Circuit Simulation*, Dissertation, Technische Universität Wien, 2005, `http://www.ue.tuwien.ac.at/phd/wagner`.

[292] M. Stockinger, *Optimization of Ultra-Low-Power CMOS Transistors*, Dissertation, Technische Universität Wien, 1999, `http://www.iue.tuwien.ac.at/phd/stockinger`.

[293] H. Herold, *Das Qt Buch*, SuSE-Press, 2001.

[294] Z. Yang and K. Duddy, "CORBA: A Platform for Distributed Object Computing (A State-of-the-Art Report on OMG/CORBA)", *ACM Operating System Review*, vol. 30, no. 2, pp. 4–31, Apr. 1996.

[295] W. Emmerich and S. Tai, Eds., *Engineering Distributed Objects*, vol. 1999 of *Lecture Notes in Computer Science*, Springer, 2001.

[296] E. Al-Ani, R. Heinzl, P. Schwaha, T. Grasser, and S. Selberherr, "Three-Dimensional State-Of-The-Art Topography Simulation", in *The 2005 European Simulation and Modelling Conference Proceedings*, Porto, Portugal, Oct. 2005, pp. 430–432.

[297] Synopsys, *Taurus-Tsuprem-4 Two-Dimensional Process Simulation Program User Manual*, Synopsys, Mountain View, 2003.

[298] S.E. Hansen and R. Dutton, *SUPREME-III User's Manual*, Stanford University, 1986.

[299] S. E. Hansen and M. D. Deal, Eds., *SUPREME-IV.GS Two Dimensional Process Simulation for Silicon and Gallium Arsenide*, Stanford University, 1993.

[300] Inc. Technology Modeling Associates, *TMA TSUPREME-4, Two-Dimensional Process Simulation Program, Version 6.5 Users's Manual*, Technology Modeling Associates, Inc., Sunnyvale, CA, 1997.

[301] R. Wittmann, A. Hössinger, and S. Selberherr, "Improvement of the Statistical Accuracy for the Three-Dimensional Monte Carlo Simulation of Ion Implantation", in *15th European Simulation Symposium on Simulation in Industry*, Delft, The Netherlands, Oct. 2003, pp. 35–40.

[302] R. Heinzl and T. Grasser, "Generalized Comprehensive Approach for Robust Three-Dimensional Mesh Generation for TCAD", in *Proc. Simulation of Semiconductor Processes and Devices*, Tokyo, Japan, Sept. 2005, pp. 211–214.

[303] Institut für Mikroelektronik, Technische Universität Wien, Austria, *VMC 2.0 User's Guide*, 2006, `http://www.iue.tuwien.ac.at/software`.

[304] M. Karner, S. Holzer, M. Vasicek, W. Goes, M. Wagner, H. Kosina, and S. Selberherr, "Numerical Analysis of Gate Stacks", *ECS Trans.*, vol. 3, no. 3, pp. 299–308, 2006.

[305] M. Karner, E. Ungersboeck, A. Gehring, S. Holzer, H. Kosina, and S. Selberherr, "Proc. Simulation of Semiconductor Processes and Devices", in *Simulation of Semiconductor Processes and Devices*, Monterey, CA, Sept. 2006, pp. 314–317.

[306] J.R. Shewchuk, "Triangle: Engineering a 2D Quality Mesh Generator and Delaunay Triangulator", in *Applied Computational Geometry: Towards Geometric Engineering*, M. C. Lin and D. Manocha, Eds., vol. 1148 of *Lecture Notes in Computer Science*, pp. 203–222. Springer-Verlag, May 1996.

[307] G.L. Dirichlet, "Über die Reduktion der positiven quadratischen Formen mit drei unbestimmten ganzen Zahlen", *Journal für die Reine und Angewandte Mathematik*, vol. 40, pp. 209–227, 1850.

[308] G. Voronoi, "Nouvelles applications des paramètres continus à la théorie des formes quadratiques", *Journal für die Reine und Angewandte Mathematik*, vol. 133, pp. 97–178, 1907.

[309] B.N. Delaunay, "Sur la Sphere Vide", *Izvestia Akademia Nauk SSSR, VII Seria, Otdelenie Matematicheskii i Estestvennyka Nauk*, vol. 7, pp. 793–800, 1934.

[310] D.F. Watson, "Computing the Delaunay Tesselation with Application to Voronoi Polytopes", *The Computer Journal*, vol. 24, no. 2, pp. 167–172, 1981.

[311] W. Pyka, P. Fleischmann, B. Haindl, and S. Selberherr, "Three-Dimensional Simulation of HPCVD—Linking Continuum Transport and Reaction Kinetics with Topography Simulation", *IEEE Trans.Computer-Aided Design of Integrated Circuits and Systems*, vol. 18, no. 12, pp. 1741–1749, 1999.

[312] P. Fleischmann, W. Pyka, and S. Selberherr, "Mesh Generation for Application in Technology CAD", *IEICE Trans.Electron.*, vol. E82-C, no. 6, pp. 937–947, 1999.

[313] P. Fleischmann, *Mesh Generation for Technology CAD in Three Dimensions*, Dissertation, Technische Universität Wien, 2000, `http://www.iue.tuwien.ac.at/phd/fleischmann`.

[314] K. Kernstock, "Ein Java GUI für den VSP", Tech. Rep., Institut für Mikroelektronik, Technische Universität Wien, Nov. 2006, Internal Project Report.

[315] M. Zohlhuber, *Visualisierung von Simulationsdaten*, Diplomarbeit, Technische Universität Wien, 2003.

[316] T. Binder and S. Selberherr, "Object-Oriented Wafer-State Services", in *Proc. European Simulation Multiconference ESM 92*, Ghent, Belgium, 2000, pp. 360–364.

[317] S. Holzer, M. Wagner, A. Sheikholeslami, M. Karner, G. Span, T. Grasser, and S. Selberherr, "An Extendable Multi-Purpose Simulation and Optimization Framework for Thermal Problems in TCAD Applications", in *12th Intl. Workshop an THERMal INvestigations of ICs and Systems*, Nice, France, Sept. 2006, pp. 239–244, TIMA, 2-9161-8704-9.

[318] A. Hössinger, T. Binder, W. Pyka, and S. Selberherr, "Advanced Hybrid Cellular Based Approach for Three-Dimensional Etching and Deposition Simulation", in *Simulation of Semiconductor Processes and Devices*, Athens, Greece, Sept. 2001, pp. 424–427.

[319] J. A. Sethian, *Level Set Methods and Fast Marching Methods*, Cambridge University Press, Cambridge, 1999.

[320] D. Adalsteinsson and J.A. Sethian, "A Level Set Approach to a Unified Model for Etching, Deposition, and Lithography I: Algorithms and Two-Dimensional Simulations", *J.Comp.Phys.*, vol. 120, pp. 128–144, 1995.

[321] D. Adalsteinsson and J.A. Sethian, "A Level Set Approach to a Unified Model for Etching, Deposition, and Lithography II: Three-Dimensional Simulations", *J.Comp.Phys.*, vol. 122, pp. 348–366, 1995.

[322] D. Adalsteinsson and J.A. Sethian, "A Level Set Approach to a Unified Model for Etching, Deposition, and Lithography III: Re-Deposition, Re-Emission, Surface Diffusion, and Complex Simulations", Tech. Rep., Department of Mathematics and Lawrence Berkeley Laboratory, University of California, Berkeley, California 94720, Aug. 1997.

[323] J.A. Sethian and D. Adalsteinsson, "An Overview of Level Set Methods for Etching, Deposition, and Lithography Development", *IEEE Trans.Semiconductor Manufacturing*, vol. 10, no. 1, pp. 167–184, Feb. 1997.

[324] S. Osher and J.A. Sethian, "Fronts Propagating with Curvature Dependent Speed: Algorithms Based on Hamilton-Jacobi Formulations", *J.Comput.Phys.*, vol. 79, pp. 12–49, 1988.

[325] J.C. Rey, J. Li, V. Boksha, D. Adalsteinsson, and J.A. Sethian, "Topography Simulation for Interconnect Deposition", *Solid State Technology*, vol. 2, pp. 77–82, 1998.

[326] T.J. Barth and J.A. Sethian, "Numerical Schemes for the Hamilton-Jacobi and Level Set Equations on Triangulated Domains", Tech. Rep., Information Sciences Directorate, NASA Ames Research Center and Department of Mathematics, University of California, Berkeley, Sept. 1997.

[327] D. Adalsteinsson and J.A. Sethian, "Computational Performance of Level Set Methods for Etching, Deposition, and Lithography Development", in *Simulation of Semiconductor Processes and Devices*, Tokyo, Japan, 1996, pp. 79–80, Business Center for Academic Societies Japan.

[328] T.S. Cale and G.B. Raupp, "A Unified Line-of-Sight Model of Deposition in Rectangular Trenches", *J.Vac.Sci.Technol.B*, vol. 8, no. 6, pp. 1242–1248, 1990.

[329] M.E. Coltrin, P. Ho, H.K. Moffat, and R.J. Buss, "Chemical Kinetics in Chemical Vapor Deposition: Growth of Silicon Dioxide from Tetraethoxysilane (TEOS)", *Thin Solid Films*, vol. 365, no. 2, pp. 251–263, 2000.

[330] D. Adalsteinsson and J.A. Sethian, "Three-Dimensional Profile Evolution under Low Sticking Coefficient", in *Simulation of Semiconductor Processes and Devices*, Cambridge, Massachusetts, 1997, pp. 61–64.

[331] P. McCann, K. Somasundram, S. Byrne, and A. Nevi, "Conformal Deposition of LPCVD TEOS", in *Micromachining and Microfabrication Process Technology VII*, J.M. Karam and J.A. Yasaitis, Eds., Oct. 2001, pp. 329–340.

[332] U.H. Kwon and W.J. Lee, "Three-Dimensional Deposition Topography Simulation Based on New Combination of Flux Distribution and Surface Representaion Algorithms", *Thin Solid Films*, vol. 445, pp. 80–89, 2003.

[333] C. Heitzinger, J. Fugger, O. Häberlen, and S. Selberherr, "Simulation and Inverse Modeling of TEOS Deposition Processes Using a Fast Level Set Method", in *Simulation of Semiconductor Processes and Devices*, Kobe, Japan, Sept. 2002, pp. 191–194.

[334] S. Holzer, A. Sheikoleslami, S. Wagner, C. Heitzinger, T. Grasser, and S. Selberherr, "Optimization and Inverse Modeling for TCAD Applications", in *Symposium on Nano Devices Technology 2004*, Hsinchu, Taiwan, May 2004, pp. 113–116.

[335] B. Sell, A. Sänger, G. Schulze-Icking, K. Pomplun, and W. Krautschneider, "Chemical Vapor Deposition of Tungsten Silicide (WSix) for High Aspect Ratio Applications", *Thin Solid Films*, vol. 443, no. 1-2, pp. 97–107, 2003.

[336] A. Doyle, "A Thick Polysilicon Three-State Fuse", *Motorola Technical Developements 3*, pp. 31–32, 1993.

[337] O. Kim, "CMOS Trimming Circuit Based on Polysilicon Fusing", *Electr.Lett.*, vol. 34, no. 4, pp. 355–356, 1998.

[338] D.J. Nickel, "Element Trimmable Fusible Link", *IBM Technical Disclosure Bulletin*, vol. 26, no. 8, pp. 4415, 1984.

[339] J.R. Lloyd and M.R. Polcari, "Polysilicon Fuse", *IBM Technical Disclosure Bulletin*, vol. 24, no. 7A, pp. 3442, 1981.

[340] Y. Fukada, S. Kohda, K. Masuda, and Y. Kitano, "A New Fusible-Type Programmable Element Composed of Aluminum and Polysilicon", *IEEE Trans.Electron Devices*, vol. 33, no. 2, pp. 250–253, 1986.

[341] D.W. Greve, "Programming Mechanism of Polysilicon Resistor Fuses", *IEEE Trans.Electron Devices*, vol. 29, no. 4, pp. 719–724, 1982.

[342] D.W. Greve, "Programming Mechanism of Polysilicon Fuse Links", *IEEE Trans.Electron Devices*, vol. 17, no. 2, pp. 349–354, 1982.

[343] C. Vahlas, P.-Y. Chevalier, and E. Blanquet, "A Thermodynamic Evaluation of Four Si-M (M = Mo, Ta, Ti, W) Binary Systems", *Computer Coupling of Phase Diagrams and Thermochemistry*, vol. 13, no. 3, pp. 273–292, 1989.

[344] D.C. Katsis and J.D. van Wyk, "Experimental Measurements and Simulation of Thermal Performance Due to Aging in Power Semiconductor Devices", in *Proc. Industry Applications Conference, 2002, 37th IAS Annual Meeting*, Pittsburgh, PA, 2002, pp. 1746–1751.

[345] C. Hollauer, *Modelling of Thermal Oxidation and Stress Effect*, Dissertation, Technische Universität Wien, 2007, `http://www.iue.tuwien.ac.at/phd/hollauer`.

[346] D. Fedasyuk and V. Makar, "Coupled Thermo-Mechanical Modeling and Optimization of Power Semiconductor Devices", in *9th Intl. Workshop an THERMal INvestigations of ICs and Systems* [352], pp. 93–98, 0-7803-7999-3.

[347] F. Kreupl, A.P. Graham, M. Liebau, G.S. Duesberg, R. Seidel, and E. Unger, "Carbon Nanotubes for Interconnect Applications", in *Proc.IEDM Tech.Dig* [107], pp. 333–326, 0-7803-8684-1.

[348] H.J. Dirschmid, *Mathematische Grundlagen der Elektrotechnik*, Vieweg, 1986.

[349] National Communication System Technology & Standards Devision, "Federal Standard 1037C: Telecommunications: Glossary of Telecommunication Terms", 1996, General Services Administration Information Technology Service.

[350] *Proc. Microtechnologies for the New Millennium 2005: VLSI Circuits and Systems*, Sevilla, Spain, May 2005.

[351] *International Interconnect Technology Conference*, Burlingame, California, June 1998.

[352] *Proc. 9th Intl. Workshop an THERMal INvestigations of ICs and Systems*, Aix-en-Provence, France, Sept. 2003. TIMA, 0-7803-7999-3.

[353] E. Zschech, K. Maex, P.S. Ho, H. Kawasaki, and T. Nakamura, Eds., *Stress-Induced Phenomena in Metallization*. AIP Conference Proceedings, 2006.

[354] *Proc. of the 201th Meeting of The Electrochemical Society: Fifth Intl. Symposium On Chemical Mechanical Polishing*, Philadelphia, PA, May 2002.

[355] *Proc. IASTED Intl. Conf. on Applied Modelling and Simulation*, Honolulu, Hawaii, USA, Aug. 1998.

Own Publications

Publications in Scientific Journals and Books

[J1] S. Holzer, R. Minixhofer, C. Heitzinger, J. Fellner, T. Grasser, and S. Selberherr, "Extraction of Material Parameters Based on Inverse Modeling of Three-Dimensional Interconnect Fusing Structures", *Microelectronics Journal*, vol. 35, no. 10, pp. 805–810, 2004.

[J2] M. Karner, A. Gehring, S. Holzer, and H. Kosina, "Efficient Calculation of Quasi-bound States for the Simulation of Direct Tunneling", in *Large-Scale Scientific Computing, LSSC 2005, LNCS 3743*, I. Lirkov, S. Margenov, and J. Waśniewski, Eds., 2006, pp. 572–577, Springer-Verlag Berlin Heidelberg.

[J3] M. Karner, A. Gehring, S. Holzer, H. Kosina, and S. Selberherr, "Efficient Calculation of Lifetime Based Direct Tunneling Through Stacked Dielectrics", *ECS Trans.*, vol. 1, no. 5, 2006, pp. 693–703.

[J4] M. Karner, S. Holzer, M. Vasicek, W. Goes, M. Wagner, H. Kosina, and S. Selberherr, "Numerical Analysis of Gate Stacks", *ECS Trans.*, vol. 3, no. 3, 2006, pp. 299–308.

[J5] M. Wagner, G. Span, S. Holzer, O. Triebl, T. Grasser, and V. Palankovski, "Power Output Improvement of Silicon-Germanium Thermoelectric Generators", *ECS Trans.*, vol. 3, no. 7, 2006, pp. 1151–1162.

[J6] M. Wagner, G. Span, S. Holzer, and T. Grasser, "Thermoelectric Power Generation Using Large-Area Si/SiGe pn-Junctions with Varying Ge Content", *Semiconductor Science and Technology*, vol. 22, 2007, pp. 173–176.

[J7] S. Holzer, A. Sheikholeslami, M. Karner, T. Grasser, and S. Selberherr, "Comparison of Deposition Models for A TEOS LPCVD Process", *Microelectronics Reliability*, vol. 47, no. 4–5, 2007, pp. 623–625.

[J8] M. Wagner, M. Karner, J. Cervenka, M. Vasicek, H. Kosina, S. Holzer, and T. Grasser, " Quantum correction for DG MOSFETs", *Computational Electronics*, vol. 5, 2007, pp. 397–400.

[J9] M. Karner, A. Gehring, S. Holzer, M. Pourfath, M. Wagner, W. Gös, M. Vasicek, O. Baumgartner, C. Kernstock, K. Schnass, G. Zeiler, T. Grasser, H. Kosina, S. Selberherr, "A Multi-Purpose Schrödinger-Poisson Solver for TCAD Applications", *Journal of Computational Electronics*, vol. 6, 2007, pp. 179–182.

Publications in Conference Proceedings

[C1] R. Minixhofer, S. Holzer, C. Heitzinger, J. Fellner, T. Grasser, and S. Selberherr, "Optimization of Electrothermal Material Parameters Using Inverse Modeling", in *Proc. 33rd European Solid-State Device Research Conference (ESSDERC 2003)*, José Franca and Paulo Freitas, Eds., Estoril, Portugal, Sept. 2003, pp. 363–366, IEEE.

[C2] S. Holzer, R. Minixhofer, C. Heitzinger, J. Fellner, T. Grasser, and S. Selberherr, "Extraction of Material Parameters Based on Inverse Modeling of Three-dimensional Interconnect Structures", in *9th Intl. Workshop an THERMal INvestigations of ICs and Systems*, Aix-en-Provence, France, Sept. 2003, pp. 263–268, TIMA, 0-7803-7999-3.

[C3] S. Holzer, A. Sheikoleslami, S. Wagner, C. Heitzinger, T. Grasser, and S. Selberherr, "Optimization and Inverse Modeling for TCAD Applications", in *SNDT 2004, Symposium on Nano Devices Technology 2004*, Hsinchu, Taiwan, May 2004, pp. 113–116.

[C4] H. Ceric, R. Sabelka, S. Holzer, W. Wessner, S. Wagner, T. Grasser, and S. Selberherr, "The Evolution of the Resistance and Current Density During Electromigration", *Proceedings SISPAD Conference*, München, Germany, pp. 331–334, 2004.

[C5] S. Holzer, C. Hollauer, H. Ceric, S. Wagner, R. Entner, E. Langer, T. Grasser, and S. Selberherr, "Three-Dimensional Transient Electro-Thermal Interconnect Simulation for Stress and Electromigration Analysis", in *Eighth International Conference on Modeling and Simulation of Microsystems*, Anaheim, CA, May 2004, pp. 620–623.

[C6] M. Karner, A. Gehring, S. Holzer, and H. Kosina, "On the Efficient Calculation of Quasi-Bound States for the Simulation of Direct Tunneling", in *5th International Conference on Large-Scale Scientific Computations*, Sozopol, Bulgaria, June 2005, pp. 33–34.

[C7] S. Holzer, C. Hollauer, H. Ceric, S. Wagner, E. Langer, T. Grasser, and S. Selberherr, "Transient Electro-Thermal Investigations of Interconnect Structures Exposed to Mechanical Stress", in *Microtechnologies for the New Millennium 2005: VLSI Circuits and Systems*, Sevilla, Spain, May 2005, pp. 380–387.

[C8] C. Hollauer, S. Holzer, H. Ceric, S. Wagner, T. Grasser, and S. Selberherr, "Investigation of Thermo-Mechanical Stress in Modern Interconnect Layouts", in *Sixth International Congress on Thermal Stresses*, Wien, Austria, May 2005, pp. 637–640.

[C9] H. Ceric, C. Hollauer, S. Holzer, T. Grasser, and S. Selberherr, "Comprehensive Analysis of Vacancy Dynamics Due to Electromigration", in *13th European Symposium on Reliability of Electron Devices, Failure Physics and Analysis*, Singapore, June 2005, pp. 100–103.

[C10] M. Karner, A. Gehring, S. Holzer, H. Kosina, and S. Selberherr, "Efficient Calculation of Quasi-Bound State Tunneling through Stacked Dielectrics", in *Meet. Abstr. Electrochem. Soc.*, Los Angeles, CA, Oct. 2005, vol. 502, p. 569.

[C11] A. Sheikholeslami, S. Holzer, C. Heitzinger, M. Leicht, O. Häberlen, J. Fugger, T. Grasser, and S. Selberherr, "Inverse Modeling of Oxide Deposition Using Measurements of a TEOS CVD Process", in *PhD Research in Microelectronics and Electronics*, Lausanne, Switzerland, Jun. 2005, vol. 2, pp. 279–282.

[C12] S. Holzer and S. Selberherr, "Material Parameter Identification for Interconnect Analysis", in *The Physics of Semiconductor Devices (IWPSD – 2005)*, New Delhi, India, Dec. 2005, pp. 635–641, invited.

[C13] S. Holzer and S. Selberherr, "Optimization Issue for Interconnect Analysis", in *Proc. International Conference on Microelectronics (MIEL)*, Belgrade, Serbia, May, 2006, pp. 465–470, invited.

[C14] M. Karner, A. Gehring, S. Holzer, M. Pourfath, M. Wagner, H. Kosina, T. Grasser, and S. Selberherr, "VSP – A Multi-Purpose Schrödinger-Poisson Solver for TCAD Applications", in *Proc. International Workshop on Computational Electronics (IWCE)*, Wien, Austria, May 2006, pp. 255–256.

[C15] A. Sheikholeslami, R. Heinzl, S. Holzer, C. Heitzinger, M. Spevak, M. Leicht, O. Häberlen, J. Fugger, F. Badrieh, F. Parhami, H. Puchner, T. Grasser, and S. Selberherr, "Applications of Two- and Three-Dimensional General Topography Simulator in Semiconductor Manufacturing Processes", in *Proc. 14th Iranian Electrical Engineering Conference* , Tehran, Iran, May 2006, 4 pages.

[C16] S. Holzer, A. Sheikholeslami, M. Karner, T. Grasser, and S. Selberherr "Comparison of Deposition Models for a TEOS CVD Process", in *Proc. 14th Workshop on Dielectrics in Microelectronics*, Catania, Italy, Jun. 2006, pp. 158–159.

[C17] S. Holzer, C. Hollauer, H. Ceric, T. Grasser, and S. Selberherr, "Three-Dimensional Transient Interconnect Analysis With Regard to Mechanical Stress", in *Proc. 13th International Symposium on the Physical and Failure Analysis of Integrated Circuits*, Singapore, Jul. 2006, pp. 154–157.

[C18] G. Meller, L. Li, S. Holzer, and H. Kosina, "Electron Kinetics in Disordered Organic Semiconductors" in *Proc. ACS/IEEE/MRS 2nd Annual Organic Microelectronics Workshop*, Toronto, Canada. Jul. 2006, p. 42.

[C19] S. Holzer, M. Wagner, M. Karner, L. Friembichler, E. Langer, T. Grasser, and S. Selberherr, "A Multi-Purpose Optimization Framework for TCAD Applications", Leuven, Belgium, Jul. 2006, p. 76.

[C20] M. Karner, E. Ungersboeck, A. Gehring, S. Holzer, H. Kosina, and S. Selberherr, "Strain Effects on Quasi-Bound State Tunneling in Advanced SOI CMOS Technologies", in *Proc. International Conference on the Simulation of Semiconductor Processes and Devices (SISPAD) 2006*, Monterey, CA, Sep. 2006, pp. 314–317.

[C21] M. Wagner, G. Span, S. Holzer, and T. Grasser, "Design Optimization of Large Area Si/SiGe Thermoelectric Generators", in *Proc. International Conference on the Simulation of Semiconductor Processes and Devices (SISPAD) 2006*, Monterey, CA, Sep. 2006, pp. 397–400.

[C22] G. Meller, L. Li, S. Holzer, and H. Kosina, "Simulation of Carrier Injection and Propagation in Molecularly Disordered Systems", in *Proc. of the 6th International Conference on Numerical Simulation of Optoelectronic Devices (NUSOD) 2006*, Singapore, Sep. 2006, pp. 1–2.

[C23] S. Holzer, M. Wagner, A. Sheikholeslami, M. Karner, G. Span, T. Grasser, and S. Selberherr, "An Extendable Multi-Purpose Simulation and Optimization Framework for Thermal Problems in TCAD Application", in *Collection of Papers Presented at the 12th International Workshop on Thermal Investigation of ICs and Systems*, Nice, France, Sep. 2006, pp. 239–244, TIMA, 2-9161-8704-9.

[C24] M. Wagner, G. Span, S. Holzer, O. Triebl, and T. Grasser, "Power Output Improvement of SiGe Thermoelectric Generators", in *Meet. Abstr. Electrochem. Soc.*, vol. 602, Cancun, Mexico, Oct. 2006, p. 1119.

[C25] M. Karner, S. Holzer, W. Gös, M. Vasicek, M. Wagner, H. Kosina, and S. Selberherr, "Numerical Analysis of Gate Stacks", in *Meet. Abstr. Electrochem. Soc.*, vol. 602, Cancun, Mexico, Oct. 2006, p. 1516.

[C26] O. Baumgartner, M. Karner, S. Holzer, M. Pourfath, T. Grasser, H. Kosina, "Adaptive Energy Integration of Non-Equilibrium Green's Functions", in *NSTI Nanotech Proceedings*, vol. 3, Santa Clara, May 2007, 1-4200-6184-4, pp. 145–148.

Curriculum Vitae

January 18^{th}, 1976	Born in Vienna, Austria
Jun. 1995	Technical High School Graduation with honors, TGM Wien Field: Electronics – Telecommunications
Oct. 1995	Enrolled in Electrical Engineering at the Technical University of Vienna
Jun. 2001	Passed 1. Diplomprüfung at the Technical University of Vienna
Apr. 2002	Received degree of "Diplomingenieur" (Dipl.-Ing.) with honors, equivalent to Master of Science (MSc), in Electrical Engineering specialization in Computer Technology from the Technical University of Vienna
May 2002 - Jan. 2003	Compulsory Military Service
Jan. 2003	Enrolled in doctoral program at the Institute for Microelectronics, Technical University Vienna
Mar. 2004	Established a business in multi-level marketing
Jul. - Sep. 2006	Held a visiting research position at Cypress Semiconductor Corp. in Mechelen, Belgium and San Jose, USA
Dec. 2006	Received "European Business Comptence License" (EBCL)