



---

**Forschungszentrum Karlsruhe**  
in der Helmholtz-Gemeinschaft

---

**Wissenschaftliche Berichte**  
FZKA 6774

# **Optimierung von Mikro- systemen für Diagnose- und Überwachungsanwendungen**

**(Abschlussseminar des  
BMBF-Verbundprojektes OMID)  
Karlsruhe, 25. und 26. November 2002**

**H. Eggert, A. Reiffer (Hrsg.)**  
Institut für Angewandte Informatik

November 2002



# **Forschungszentrum Karlsruhe**

in der Helmholtz-Gemeinschaft

**Wissenschaftliche Berichte**

**FZKA 6774**

## **Optimierung von Mikrosystemen für Diagnose- und Überwachungsanwendungen**

**(Abschlußseminar des BMBF-Verbundprojektes OMID)**

**Karlsruhe, 25. und 26. November 2002**

**H. Eggert, A. Reiffer<sup>\*)</sup> (Herausgeber)**

**Institut für Angewandte Informatik**

<sup>\*)</sup> InterConnect Software GmbH  
Am Fächerbad 3  
67131 Karlsruhe

Forschungszentrum Karlsruhe GmbH, Karlsruhe

2002

Für diesen Bericht behalten wir uns alle Rechte vor

Forschungszentrum Karlsruhe GmbH  
Postfach 3640, 76021 Karlsruhe

Mitglied der Hermann von Helmholtz-Gemeinschaft  
Deutscher Forschungszentren (HGF)

ISSN 0947-8620

# **Optimierung von Mikrosystemen für Diagnose- und Überwachungsanwendungen**

**(Abschlußseminar des BMBF-Verbundprojektes OMID)**

## **Zusammenfassung**

Im BMBF-Verbundprojekt OMID (Optimierung von **M**ikrosystemen für **D**iagnose- und **Ü**berwachungsanwendungen) wurden Methoden und Werkzeuge zur automatisierten Designoptimierung entwickelt und an fünf innovativen vorwettbewerblichen Entwicklungen aus der Mikrosystemtechnik erfolgreich erprobt.

Modellbildung, Simulation und Optimierung lieferten qualitativ verlässliche Designs und diese ermöglichten wiederum eine gezielte Herstellung von Labormustern (Prototypen). Die meßtechnische Erfassung von Funktionsverläufen der Labormuster führte zur Vergleichsmöglichkeit mit den vorher berechneten Funktionen. Diese Doppelstrategie ermöglichte über die gesamte Projektlaufzeit hinweg eine iterierende Werkzeug- und Prototypenevaluierung.

Zusätzlich wurden die entwickelten Optimierungswerkzeuge zusammen mit kommerziellen Simulatoren an geeignete Internetmechanismen adaptiert, so daß heute eine räumlich verteilte Simulation und Optimierung von Mikrosystementwürfen möglich ist.

In OMID wurde erstmalig dieser gesamtheitliche Ansatz zur Entwurfsmethodik in der Mikrosystemtechnik als Projektziel definiert und auch erfolgreich umgesetzt.

# **Optimization of Microsystems for Diagnosis and Monitoring Applications**

**(Final Seminar of the OMID Joint Project Funded by the BMBF)**

## **Abstract**

Within the framework of the OMID (**O**ptimization of **M**icrosystems for **D**iagnosis and **M**onitoring Applications) joint project funded by the BMBF, methods and tools for automatic design optimization were developed and tested successfully on five innovative pre-competitive developments in microsystems engineering.

Modelling, simulation, and optimization yielded qualitatively reliable designs that allowed for a specific manufacture of laboratory models (prototypes). Based on the functional patterns measured, comparison with the pre-calculated functions was possible. This double strategy provided for an iterative tool and prototype evaluation over the entire project duration.

In addition, the optimization tools developed were adapted to suitable internet mechanisms together with commercial simulators. Today, a spatially distributed simulation and optimization of microsystems designs can be carried out.

In OMID, this holistic approach to designing in microsystems engineering was defined as a project objective and implemented successfully for the first time.

# Inhaltsverzeichnis

## Übersicht

### **OMID – Werkzeug- und Prototypenentwicklung in der Mikrosystemtechnik**

H. Eggert, Forschungszentrum Karlsruhe.....1

## Methoden und Werkzeuge I

### **Simplorer als Remote-Simulator**

L. Zacharias, Ansoft GmbH.....8

### **Internetbasierte Designoptimierung mit HyGLEAM**

W. Jakob, Forschungszentrum Karlsruhe.....14

### **Modellbasierte Designoptimierung von Mikrosystemen mit MODOS**

D. Peters, H. Bolte, O. Nüssen, R. Laur, Universität Bremen.....28

## Anwendungen I

### **Betrachtung des Flüssigkeitsverhaltens in medizinischen Teststreifen aus Kunststoff**

C. Schön, STEAG microParts.....39

### **Modellbasierte Optimierung medizinischer Teststreifen aus Kunststoff**

O. Nüssen, Universität Bremen

S. Halstenberg, Forschungszentrum Karlsruhe.....45

### **Die Aktorplatte – ein universeller fluidischer Mikroaktor**

I. Tahhan, Bartels Mikrotechnik GmbH.....57

### **Modellierung und Optimierung der Aktorplatte**

A. Quinte, Forschungszentrum Karlsruhe.....66

## **Anwendungen II**

### **Eine Lasersteuerung für optische Meßgeräte**

G. Gärtner, VacuTec Meßtechnik GmbH.....78

### **Modellierung, Simulation und Optimierung einer Lasersteuerung**

P. Schwarz, Fraunhofer Gesellschaft IIS.....86

### **IR-Gassensor für Überwachungsaufgaben in Umwelt- und Gebäudetechnik**

K.-H. Suphan, Micro-Hybrid Electronic GmbH.....95

### **Modellbasierte Optimierung eines IR-Gassensors**

I. Sieber, Forschungszentrum Karlsruhe

O. Nüssen, Universität Bremen.....101

### **Ein hochauflösendes radiometrisches Meßsystem zur Qualitätskontrolle schnell laufender Materialbahnen**

D. Richter, G. Gärtner, VacuTec Meßtechnik GmbH.....111

### **Entwurfsunterstützung für ein radiometrisches Meßsystem – Ein physikalisches Modell der Ladungsträgergeneration**

H. Bolte, Universität Bremen.....118

### **Entwurfsunterstützung für ein radiometrisches Meßsystem – Kammerelektronik und Gesamtsystemmodell**

E. Huck, Fraunhofer Gesellschaft IIS.....127

## **Methoden und Werkzeuge II**

### **MOSCITO – Ein modulares internetbasiertes Programmsystem für die Optimierung von Mikrosystemen**

A. Schneider, Fraunhofer Gesellschaft IIS.....140

### **ORBIT – Internetbasierte Plattform für Remote-Anwendungen**

A. Reiffer, InterConnect Software GmbH.....152



# OMID – WERKZEUG- UND PROTOTYPENENTWICKLUNG IN DER MIKROSYSTEMTECHNIK

H. Eggert, Forschungszentrum Karlsruhe, Institut für Angewandte Informatik  
eggert@iai.fzk.de

## Kurzfassung

Das vom BMBF geförderte Verbundprojekt **OMID** (Optimierung von **MI**krosystemen für **DI**agnose- und **Ü**berwachungsanwendungen) verfolgte das Ziel, in der Mikrosystemtechnik einen Durchbruch im Rahmen des Systementwurfs zu erlangen. Während in vergangenen Projekten -durchaus erfolgreich- entweder einzelne Mikrosysteme bzw. Entwurfsmethoden und -werkzeuge entwickelt wurden, verfolgte OMID einen ganzheitlichen Ansatz zur Entwurfsmethodik von Mikrosystemen. Ganzheitlich bedeutet:

- Bereitstellung von vorhanden Modellierungs- und Simulationswerkzeugen.
- Entwicklung und Bereitstellung von Optimierungsmethoden und -werkzeugen.
- Zugang zu den Werkzeugen über das Internet.
- Entwicklung von Simulationsmodellen und deren Optimierung für exemplarische Anwendungen aus unterschiedlichen Domänen der Mikrosystemtechnik.
- Bereitstellung dieser Simulationsmodelle im Internet.
- Fertigung von Labormustern der exemplarischen Anwendungen.
- Meßtechnische Erfassung des Funktionsverhaltens der Labormuster.
- Evaluierung der optimierten Simulationsmodelle.

Alle Ziele wurden ohne Einschränkung erreicht. Das wurde insbesondere durch Nutzung der Ergebnisse früherer Verbundprojekte möglich.

## 1 Einleitung

Der Entwurf von Mikrosystemen war zunächst durch Intuition, Erfahrung der Designer sowie langwierige und kostenintensive Experimente mit Testaufbauten gekennzeichnet. Dieser Aufwand sollte - vor allem aus Gründen des internationalen Wettbewerbs - durch verstärkten Rechneinsatz reduziert werden. In BMBF-Projekten wie z. B. METEOR [1], MIMOSYS [2] und DEMIS [3] wurden Schritte in diese Richtung getan, diese Vorarbeiten und Ergebnisse ließen sich daher in OMID nutzen. Die im Projekt zu entwickelnden Mikrosysteme führten einerseits zu Geräten mit deutlich erhöhtem Gebrauchswert - und damit erhöhten Absatzchancen -, sind andererseits aber sehr komplexe Komponenten, zu deren Entwicklung ein erhöhter wissenschaftlich/technischer Aufwand getrieben werden mußte. Die Besonderheiten sind:

- Die in der Mikrosystemtechnik typischen hohen Querempfindlichkeiten, die im Systementwurf nur durch domänenübergreifende Modelle auf höheren Abstraktionsebenen beherrschbar werden; dafür ist jedoch der Modellierungsaufwand hoch.
- Ein extrem großer Lösungsraum bereits bei wenigen zu optimierenden Systemparametern, so daß Variantenuntersuchungen von Hand nicht mehr durchführbar sind (weder durch Simulation noch gar durch experimentelle Aufbauten).
- Der Einsatz mehrerer hochspezialisierter Tools (Simulatoren, Optimierungsprogramme, ...), deren Anschaffung für kleine Firmen viel zu teuer ist und die darüber hinaus spezifisches Know-How verlangen, das in KMU's häufig nicht vorhanden ist.

- Die Komplexität beim Mikrosystementwurf (Modellbildung, Simulation, Optimierung, Evaluierung) erfordert eine besonders enge Zusammenarbeit mehrerer Bearbeiter in Unternehmen und Forschungseinrichtungen.
- Zu erwartende Probleme bei der rechnergestützten Optimierung bezüglich
  - Konvergenzsicherheit (Maß für die Wahrscheinlichkeit, eine genügend gute Lösung zu finden),
  - Konvergenzgeschwindigkeit (die benötigte Zeit zur Lösung einer Optimierungsaufgabe),
  - Randbedingungen (Bedingungen, die das Verfahren an die Optimierungsaufgabe stellt).

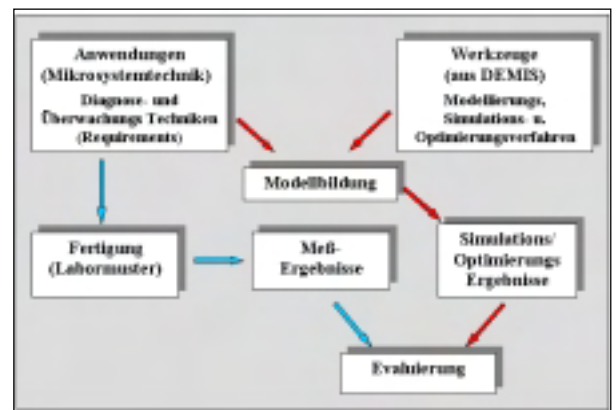
Daraus leiteten sich die *allgemeinen wissenschaftlich/technischen Zielstellungen* für OMID ab:

- Effektive Entwicklung neuer Mikrosysteme durch Einsatz von Simulatoren und Optimierungswerkzeugen.
- Modellbildung der Mikrosysteme (unter Verwendung vorhandener Bibliotheken); dazu gehören FEM-Modelle sowie Netzwerk- und Verhaltensmodelle auf den höheren Abstraktionsebenen.
- Messtechnische Evaluierung der Modelle anhand von Labormustern und mittels spezieller Testaufbauten.
- Entwicklung der Software-Infrastruktur, mit der alle Partner auf die bei den Forschungseinrichtungen vorhandenen und weiterzuentwickelnden Werkzeuge zurückgreifen können (Internet-basierte Arbeitsweise).
- Insbesondere dieser letzte Aspekt, der mit modernster Software unterstützte Dienstleistungscharakter einiger Aufgaben der Forschungspartner, war ein wichtiges Anliegen von OMID. Es sollten prototypisch die Hilfsmittel dafür bereitgestellt und die erfolgreiche Arbeit bei der Entwicklung innovativer Mikrosysteme demonstriert werden.

Die Lösung dieser Aufgaben sollten mit dem folgenden *allgemeinen Lösungsweg* erreicht werden:

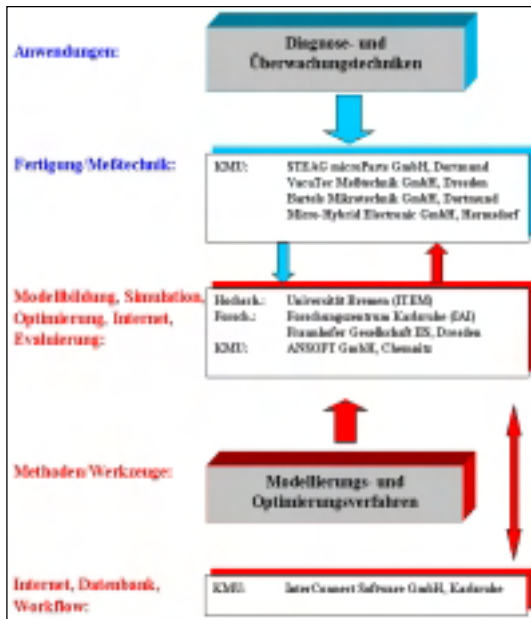
- Die Herstellerfirmen liefern die Requirements für die neuen Mikrosysteme.
- Die Forschungspartner unterstützten die Firmen bei der Modellbildung, Simulation und Optimierung.
- Die Modelle mussten von den Herstellerfirmen an den Labormustern meßtechnisch evaluiert werden.
- Die Modelle der Mikrosysteme wurden mit entsprechenden Verfahren optimiert.
- Die Bereitstellung und Weiterentwicklung von Werkzeugen erfolgte durch die Software-Entwicklungsfirmen und Forschungseinrichtungen. Die Vermarktung von Software-Ergebnissen wird durch die Software-Entwicklungsfirmen übernommen.
- Das Zusammenwirken der Partner wird inhaltlich und organisatorisch durch die verteilte Simulation im Internet realisiert.

Dieser Arbeitsablauf innerhalb des Projektes wird in *Abbildung 1* zusammenfassend beschrieben:



**Abbildung 1: Arbeitsablauf in OMID**

Das Projektkonsortium bestand aus 6 Industriepartnern (KMU's), einer Universität und zwei Forschungseinrichtungen.



**Abbildung 2: Organisationsstruktur von OMID**

## 2 Anwendungen

In diesem Kapitel wird kurz der Stand der Wissenschaft und Technik der einzelnen Anwendungen beschrieben. Ausführliche Beschreibungen befinden sich in den Einzelbeiträgen "Anwendungen".

### 2.1 Medizinische Teststreifen

Mit medizinischen Teststreifen, auch medizinische Schnelltests genannt, können Körperflüssigkeiten einfach und ohne großen Laboraufwand auf das Vorhandensein von bestimmten Komponenten untersucht werden. Das können z. B. Proteine, Viren oder in Folge einer Immunreaktion gebildete Antikörper sein. Häufig müssen die getesteten Inhaltsstoffe der Flüssigkeit mit sogenannten Markern noch sichtbar gemacht werden.

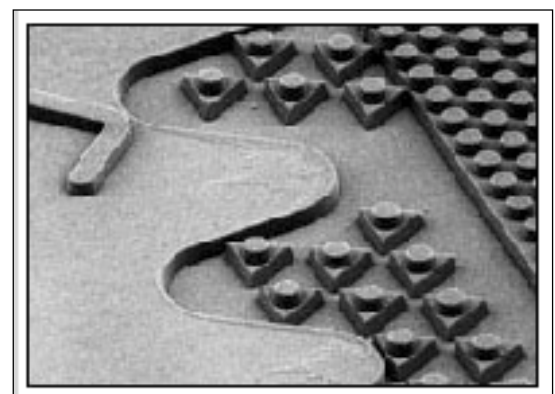
Ein praktisches Beispiel ist die Frühdiagnose und Diagnose von Herzinfarkten. Dabei gibt es verschiedene sogenannte Infarktmarker, die teilweise die dem eigentlichen Herzinfarkt vorausgehenden kleineren Infarkte belegen können, als auch Klarheit darüber schaffen können, ob es sich tatsächlich um einen Herzinfarkt oder nur um eine zeitweilige

Muskelverspannung im Brustkorbbereich handelt.

Bei der Untersuchung von Proben mit solchen Schnelltests wird die Probenflüssigkeit zu einer Reaktionszone transportiert, wo eine chemische Substanz, welche die Infarktmarker enthält, deponiert wurde. Diese Substanz muß für eine bestimmte Reaktionszeit mit der Flüssigkeit in Kontakt kommen. Nur durch die genaue Einhaltung der Reaktionszeit kann auch eine genaue quantitative Aussage über das Vorhandensein der zu testenden Bestandteile der Flüssigkeiten gemacht werden.

Heute beruht der Transport der Körperflüssigkeit zur Reaktionszone meist auf der Kapillarwirkung verschiedener Gewebe (z. B. Vliese) und/oder der Filterwirkung von Membranen. Deren Nachteil ist jedoch, daß die Porengrößen und damit die Kapillarwirkungen statistisch verteilt sind. Somit kann der Test nur eine qualitative Aussage liefern. Andere Verfahren zur Erzeugung von definierten Poren sind sehr teuer (z. B. Kernspur-Verfahren).

Hier besitzt die Mikrostrukturtechnik einen entscheidenden Vorteil, denn mit Hilfe der Mikrostrukturtechnik ist es möglich, Mikrokanäle mit genau definierten Kapillaritäten zu erzeugen. Die Abmessungen dieser Kanäle sind in weiten Grenzen einstellbar (Breite, Länge, Tiefe von wenigen Mikrometern bis zu einigen Zentimetern). Damit wird eine Flexibilität erreicht, die es möglich macht, die mikrostrukturierten Teile in vielen verschiedenen Fällen einzusetzen.



**Abbildung 3: Mikrostrukturen für Medizinische Teststreifen**

Zusätzlich kann durch das Einbringen von Mikrostrukturen (z. B. einige tausend Säulen mit Durchmessern von 10  $\mu\text{m}$  bis 500  $\mu\text{m}$ ) auch das Geschwindigkeitsprofil der Flüssigkeit im Kanal eingestellt werden, um eine genaue Einhaltung der günstigsten Reaktionszeiten gewährleisten zu können.

## 2.2 Aktorplatte

In der Mikrotechnik werden Aktoren seit längerer Zeit durch verschiedenste Verfahren realisiert.

So sind schon vor mehr als 10 Jahren Aktoren auf Siliziumtechnik entworfen und gebaut worden, die sowohl rotatorische als auch translatorisch mechanische Energie zur Verfügung stellen. Ein Hauptnachteil dieser Aktoren ist prinzipbedingt (durch die erreichbaren Größenordnungen als auch die eingesetzten Materialien und Prozesse), daß sie

- eine äußerst geringe Arbeit je Hub bzw. Umdrehung zur Verfügung stellen können,
- nur eine begrenzte Lebensdauer besitzen und
- mit sehr wenigen Ausnahmen nie in ein echtes Industrieprodukt überführt worden sind.

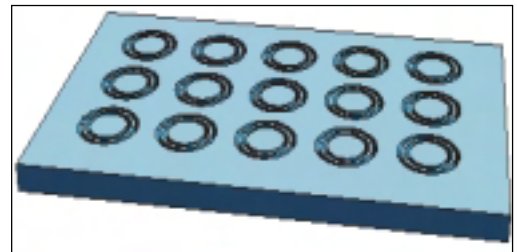
Eine interessante Entwicklung neuerer Zeit besteht darin, bekannte Prinzipien aus der „Makrowelt“ in kleinste Dimensionen zu überführen, wobei es der traditionellen Feinwerktechnik gelang, in neue Dimensionen vorzustoßen, wodurch die Kluft zwischen Mikroaktoren auf Siliziumbasis und traditionell hergestellten deutlich schrumpfte. Elektromotoren mit weniger als 2 mm Durchmesser oder 1 mm Höhe befinden sich bereits im Kleinserienstadium. Hauptnachteile der direkt vor oder bereits in der Markteinführung befindlichen Aktoren sind vor allem der hohe Preis und die für den industriellen Einsatz noch oft ungenügende Robustheit.

Der in OMID entwickelte fluidische Aktor besitzt jedoch entscheidende Vorteile:

- Er ist sehr preisgünstig, weil zum einen die Materialien (Kunststoffe), aber auch die Fertigungsverfahren (Lasertechnik in

der Prototypenphase, Prägetechnik in der Serienproduktion) vergleichsweise geringe Kosten verursachen.

- Er ist deutlich robuster, weil die Materialien nicht brüchig und chemisch resistent sind.
- Er ist stärker, weil die krafterzeugende Quelle sich nicht mehr direkt am Ort des Geschehens, sondern entfernt vom Wirkungsort befindet. So kann auf „makroskopische“ Kraftquellen (Pumpen) ausgewichen werden, wenn dies nötig ist. Andernfalls hält der Baukasten der Mikrotechnik auch eine Mikropumpe bereit, welche direkt in die Mikrosysteme integriert werden kann.
- Er ist flexibel, weil die Aktorplatte alleine, in einer Linie oder auf einer Fläche angeordnet sein kann, und mit denselben Fertigungsschritten unterschiedliche Mikrosysteme realisiert werden können (Braillezeile, Relais, ...). Insbesondere die Kombinierbarkeit des Aktors mit anderen, fluidischen Bauelementen, die praktisch im gleichen Arbeitsgang hergestellt werden, sei hier hervorgehoben.



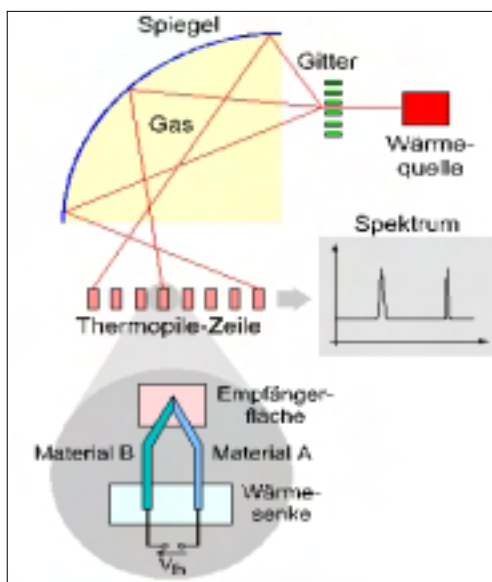
**Abbildung 4: Miniaturisiertes Braille-Display mit Aktorplatten**

## 2.3 Thermopile-Zeilen-Sensor zur Gasanalyse

Thermopile-Einzelsensoren, kurz Thermopiles, sind Infrarot-Strahlungssensoren die sich gegenüber anderen Infrarotsensoren dadurch auszeichnen, daß sie statische Signale verarbeiten können, ihre Empfindlichkeit nahezu konstant im Spektralbereich von ultraviolettem bis infrarotem Licht ist, und ihre Funktion ohne zusätzliche Versorgungsspannung

und zusätzliche Kühlung gewährleistet ist. Das Ausgangssignal eines Thermopiles ist frei von Stromrauschen und proportional zur Temperaturdifferenz zwischen aktivem Element und Referenzelement. Thermopiles werden vorrangig in Pyrometern und in der Infrarotspektroskopie eingesetzt.

In Abhängigkeit vom zu detektierenden Signal liegt die Ausgangsspannung des Sensors ohne zusätzliche Signalaufbereitung im Bereich von einigen  $\mu\text{V}$  bis zu einigen mV. Durch Hermetisierung der Sensorchips können Zuverlässigkeit und Langzeitstabilität gesichert werden. Mit einer geeigneten Wahl des Füllmediums des hermetischen Gehäuses können die anwendungsrelevanten Eigenschaften des Sensors wie Empfindlichkeit und Zeitkonstanten weitgehend angepaßt werden.



**Abbildung 5: Schematische Darstellung der Gasanalyse**

Thermopilezeilen mit 32 Pixel werden derzeit nur von einem amerikanischen Wettbewerber serienmäßig angeboten, allerdings ohne Auswertelektronik.

Im Bereich der Gasspektroskopie sind derzeit Meßsysteme mit linearem Strahlungsverlauf zwischen Infrarotquelle und Infrarotdetektor die auf dem Markt dominierende Lösung. Dieser lineare Strahlungsverlauf stellt die wesentliche Begrenzung für die Anwendung der Infrarotsensoren in der Gassensorik dar, da die nachzuweisenden Gase eine gewisse Mindest-

absorption im Infrarotbereich besitzen müssen, um die Abmessungen der Meßzellen beherrschbar zu halten. Die Erweiterung der Anwendungen wird in der Realisierung von Meßzellen gesehen, die den Beschränkungen des rein linearen Strahlungsverlaufes nicht unterliegen.

## 2.4 Ionisationskammer

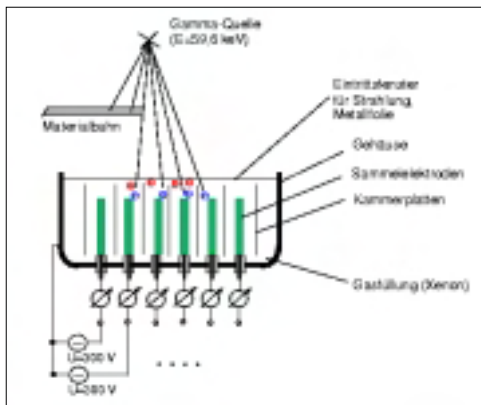
Detektoren für ionisierende Strahlung können in vielfältigen Gebieten eingesetzt werden. Große Anwendungsbereiche liegen im Strahlenschutz, in der Gammaskpektroskopie, in industriellen Anwendungen sowie in der Materialanalyse durch Röntgenfluoreszenzverfahren.

In der Industrie werden Ionisationskammern für die Bestimmung unterschiedlicher Prozessparameter während der Produktion verwendet. So wird zum Beispiel die Materialstärke von Papier-, Kunststoff-, Textil- oder Metallbahnen mit diesen Kammern bei Verwendung von  $\beta$ - oder  $\gamma$ -Strahlern geeigneter Energie festgestellt.

Ein neuer Ionisationskammertyp kann zur Detektion von Materialkanten im Produktionsprozess herangezogen werden. Da die Lage der Materialkante so genau wie möglich festgestellt werden soll, werden umfangreiche Optimierungen und Berechnungen am gesamten System notwendig.

Das System besteht aus einem  $\gamma$ -Strahler und einem ortsempfindlichen Detektor. Absorbierte  $\gamma$ -Quanten erzeugen im Detektor Elektronen-Ionenpaare, die durch Strommessung nachgewiesen werden. Unter der Annahme, daß die Anzahl der erzeugten Elektronen pro absorbiertem Quant im gesamten, aktiven Kammervolumen konstant ist, kann durch Strommessung auf die Materialstärke zurückgeschlossen werden. Die Ortsempfindlichkeit wird durch Aneinanderreihen mehrerer unabhängiger Kammerzellen erreicht („Mehrfachkammer“). Die Materialbahn deckt teilweise den Strahlengang zwischen Strahler und Detektor ab, so daß die Gammaintensität unterhalb der Bahn in bestimmten Detektorbereichen geschwächt ist, und damit in einzelnen Kammerzellen geringere Ströme erzeugt

werden. Typische Ströme liegen im Bereich von  $10^{-12}$  A bis  $10^{-8}$  A. Um eine möglichst schnelle und gute Darstellung der Materialkante zu erreichen, muß die Kurvenform, die sich aus dem Verlauf der Stromsignale als Funktion der Kammerzellen ergibt, optimiert werden (maximaler Anstieg im Bereich der Materialkante, möglichst große Ströme).



**Abbildung 6: Funktionsprinzip der Ionisationskammer**

Weltweit gibt es maximal 25 Hersteller von industriell genutzten Metall-Ionisationskammern. Diese gehen in Meßsysteme ein, die von etwa 12 ... 15 Systemherstellern zusammengestellt und vertrieben werden, wobei die Metall-Ionisationskammern als wichtiges Kernstück der Meßeinrichtung anzusehen ist.

## 2.5 Lasersteuerung

Nach Projektbeginn wurde zusätzlich eine weitere Modellierungs- und Simulationsaufgabe (Lasersteuerung) definiert, welche eine optimale Auslegung eines Messgerätes zur optischen Messung von Oberflächenqualität und Dichte von Stoff- und Papierbahnen zum Ziel hatte. Das Messsystem besteht aus Sender (Laserstrahlenquelle) und Empfänger (Photodioden). Durch die Eigenschaften des zu untersuchenden Materials in der Messstrecke wird die Transmission bzw. Reflektion des Laserlichts beeinflusst, was Rückschlüsse auf die Materialqualität zulässt.

## 3 Methoden

In diesem Kapitel werden nur die

wesentlichen Merkmale der verwendeten Methoden und Verfahren beschrieben. Ausführliche Beschreibungen, insbesondere auch die der verwendeten Werkzeuge, befinden sich in den Einzelbeiträgen "Methoden und Werkzeuge".

### 3.1 Finite Elemente Methode (FEM)

Die Finite-Elemente-Methode (FEM) ist ein numerisches Berechnungsverfahren, das zur Lösung von verschiedenen physikalischen Feldproblemen, bei denen das Verhalten von Kontinua durch partielle, orts- und zeitabhängige Differentialgleichungen beschrieben werden kann.

Dazu gehören Problemstellungen aus den Bereichen der Strukturmechanik, der Strömungsmechanik, der Elektromagnetik und der Wärmeübertragung. Ferner lassen sich auch einige gekoppelt auftretende Feldprobleme wie z. B. thermoelastische oder piezoelektrische Wechselwirkungen behandeln. Mit der FE-Methode ist also eine physiknahe Behandlung von Problemstellungen möglich.

Neueste Entwicklungen in OMID lassen bereits auf der FEM-Ebene eine Modelloptimierung zu. Allerdings ist hierzu eine aufwendige Rechnerinfrastruktur aus Workstations sowie eine flexible Jobverteilungssoftware notwendig, da nur durch Parallelisierung der FEM-Modelle und der Optimierungssoftware vertretbare Rechenzeiten im Minutenbereich erreichbar sind.

### 3.2 Makro- und Verhaltensmodelle

Charakteristisch für technische Systeme bzw. Teilsysteme ist das Auftreten von Effekten unterschiedlicher physikalischer Domänen (z.B. Mechanik, Fluidodynamik, Thermodynamik), die untereinander in Wechselwirkung sein können. Bei einer Modellierung auf physikalischer Ebene (FEM-Modelle) können oftmals die Wechselwirkungen nicht adäquat berücksichtigt werden oder es entstehen sehr große Modelle, die mit nicht mehr vertretbar hohen Rechenzeiten verbunden sind.

Kann das Verhalten des Systems auf der Basis der Energieflußgesetze beschrieben werden, so können Makro- und Verhaltensmodelle eingesetzt werden, die das physikalische Verhalten auf einer abstrakteren Ebene mit gewöhnlichen Differentialgleichungen domänenübergreifend beschreiben. Diese Modelle weisen kürzere Rechenzeiten auf und können dadurch auch sehr gut zur Design-Optimierung eingesetzt werden.

### 3.3 Optimierung

Der Entwurf komplexer technischer Systeme führt zur Notwendigkeit von Modellbildung, Simulation und Optimierung. Das Konzept der modellbasierten Optimierung hat ein Design von hoher Qualität bei möglichst geringem Prototypenbau zum Ziel. Dazu werden unterschiedliche Optimierungsverfahren eingesetzt, die sich ergänzen.

Mit Hilfe evolutionärer Algorithmen wird eine globale umfassende Suche durchgeführt, welche die interessanten Gebiete des Suchraums bestimmt. Die dabei ermittelten Lösungen werden mit konventionellen lokalen Suchverfahren wie z. B. dem Rosenbrock-Verfahren oder dem Complex-Algorithmus weiter verbessert.

### 3.4 Internet basierte Simulation und Optimierung

Eine Marktanalyse ergab, daß zwar derzeit bereits kommerzielle Workflow-, Dokumentations- und Projektmanagementsysteme auf Basis des Internets existieren, jedoch keine kombinierten Systeme, die möglichst alle Bereiche abdecken, angeboten werden. Insbesondere ist die Möglichkeit zur Einbindung von Remote-Tools über das Internet in diesen Systemen derzeit nicht vorhanden, womit die entscheidende und damit wichtigste Fähigkeit einer Mitbenutzung von in der Regel sehr kostspieligen Softwarelizenzen durch die KMU's in diesen Systemen fehlt. Der derzeitige Stand der Technik zeigt, daß bestehende Systeme vor allem im Entwurfsbereich der Mikroelektronik unterstützend eingesetzt werden, jedoch im Umfeld der Mikrosystemtechnik aufgrund des ungleich breiteren Ansatzes

nicht oder nur unzureichend verwendet werden können.

In DEMIS und OMID wurden Internet-Verfahren zum Remote-Tool-Einsatz und zum Remote-Modell/Optimierungs-Einsatz entwickelt. Diese Verfahren wurden evaluiert und haben sich bewährt.

## 4. Schlußbemerkungen und Ausblick

Alle Projektziele wurden ohne Einschränkung erreicht. Im Rahmen der Anwendungen liegen alle Systementwürfe in optimierter Form vor.

Die entwickelten Methoden und Werkzeuge werden, bedingt durch ihren Remote-Einsatz im Internet, im Gegensatz zu früheren Verbundprojekten, auch nach Projektende weiter verfügbar sein.

## Literatur

- [1] Eggert, H. et al. (Hrsg.):  
Methoden- und Werkzeugentwicklung für den Mikrosystementwurf  
Tagungsband 4. Statusseminar zum BMBF-Verbundprojekt METEOR  
18.-19. Nov. 1996, Karlsruhe  
ISSN 0947-1413
- [2] Laur, R. et al. (Hrsg.):  
Methoden und Werkzeuge zum Entwurf von Mikrosystemen  
Tagungsband 2. Statusseminar zum BMBF-Verbundprojekt MIMOSYS  
4.-5. Dez. 1997, Paderborn,  
ISSN 0947-1413
- [3] John, W. et al. (Hrsg.):  
Methoden und Werkzeuge zum Entwurf von Mikrosystemen  
Tagungsband der gemeinsamen Statusseminare zu den BMBF-Verbundprojekten CADWOK, DEMIS, MIMOSYS, SESAM, SIMKOS  
2.-3. Dez. 1999, Berlin  
ISSN 0947-1413

# SIMPLORER ALS REMOTE-SIMULATOR

Dr. Ing. Lutz Zacharias

Ansoft GmbH & Co KG, Chemnitz

[zacharias@simplorer.com](mailto:zacharias@simplorer.com)

## Kurzfassung

Das Simulationssystem SIMPLORER wird von der Firma Ansoft entwickelt und vertrieben. Es wurde speziell für die Belange der Elektronik / Antriebstechnik und Mechatronik entwickelt. Dieser Beitrag stellt die aktuellen Arbeiten zur Remote – Simulation mit dem Simulationswerkzeug SIMPLORER vor.

## 1. Einleitung

Die Firma Ansoft GmbH & Co KG hat den Simulator SIMPLORER entwickelt, der für mechatronische, kontinuierliche und diskrete Systeme geeignet ist. Seit mehreren Jahren bietet die Firma Unterstützung auf dem Gebiet der Designoptimierung an. Im Rahmen der BMBF-Projekte MIMOSYS und DEMIS wurden Optimierungsverfahren entwickelt, implementiert und getestet.

Das System SIMPLORER wird im OMID-Projekt zur Modellierung, Simulation und Optimierung von Anwendungen eingesetzt. Dabei werden die bereits vorhandenen Ergebnisse umgesetzt und anwendbar gemacht. Ziel ist die Verifikation der Tools am konkreten Designprozeß (mikrofluidischer Prozess, Aktorplatte) und die Integration in die Simulationsumgebung SIMPLORER. Diese wird so erfolgen, daß die Unterstützung für die verschiedenen Entwurfsschritte gegeben ist. Besonderer Schwerpunkt ist dabei die Anpassung der Entwicklungsumgebung an die Erfordernisse des Designprozesses. Dazu muss

eine über die reine Simulation hinausgehende Lösung erarbeitet werden, die auch in der Lage ist, Arbeitsabläufe abzubilden und sich in heterogene Arbeitsumgebungen einzuordnen.

Die angestrebte Lösung führt zu einer breit nutzbaren Umgebung mit integrierten Hilfstools, wie sie so derzeit am Markt nicht angeboten wird. Für den Aufbau eines Dienstleistungsangebots über INTERNET ist die Remotefähigkeit der einzelnen Tools Grundvoraussetzung. Durch die Verwendung von **ORBIT** (Internet basierte Simulation und Optimierung) wird die Umgebung remotefähig gemacht werden, womit international neuartige Möglichkeiten des Mikrosystementwurfs entstehen.

## 2. Das Multi-Domain-Simulationswerkzeug SIMPLORER

SIMPLORER, als ein kombinierter Schaltungs- und Systemsimulator adressiert Anwender der Bereiche Mikrosystemtechnik, Mechatronik, Leistungselektronik, Elektrische Antriebs- und Energietechnik sowie MSR-Technik. Einfach und schnell können Anordnungen mit sowohl elektrischen als auch nichtelektrischen Komponenten modelliert und simuliert werden. Dabei werden die nachzubildenden Anordnungen praxisnah mit Hilfe von elektrischen Netzwerken, regelungstechnischen Blockdiagrammen und Zustandsgraphen beschrieben. Stabile Rechenalgorithmen sichern, im Gegensatz zu SPICE-basierten Simulatoren, numerische Konvergenz auch bei schnell schaltenden elektronischen Systemen.



Die umfangreiche Modelldatenbank beinhaltet neben aktiven und passiven elektrischen Grundbauelementen, Operationsverstärkern, regelungstechnischen Übertragungsfunktionen und Nichtlinearitäten auch statische und dynamische Leistungshalbleiter, Sensoren, elektrische Maschinen sowie Mechanik- und Hydraulikkomponenten. Gerade in Gestalt der angeführten Bibliothekserweiterungen stellt SIMPLORER nunmehr noch bessere Funktionalitäten auch außerhalb seiner üblichen Domänen zur Verfügung.

Mit der neu verfügbaren Mechanik-Toolbox kann man translatorisch und rotatorisch bewegte Systeme modellieren. Die für die Arbeit mit der Toolbox relevanten Parameter sind Trägheitsmomente, Steifigkeiten sowie Dämpfungs- und Reibungskoeffizienten. Die in der Bibliothek enthaltenen Basisbausteine sind u.a. Masse-Element (MAS), Steifigkeitselement (STF), Koordinatentransformation (TSF), Getriebeübersetzung, Momentenquelle (SRCF) sowie mechanisches "Bezugspotential" (GND).

So können alle vorhandenen SIMPLORER-Standardmodelle von translatorischen oder rotatorischen Aktoren durch die Ergänzung weiterer mechanischer Modelleigenschaften in ihrer Detailtreue verbessert werden. Das kann zum Beispiel durch das Ankoppeln eines entsprechenden Masselementes geschehen.

Anwendungsvorteile bei Einsatz der SIMPLORER Mechanik-Toolbox ergeben sich insbesondere durch

- Einfacheres und schnelleres Abbilden von (mikro)mechanischen Systemen im Vergleich zur Modellierung mit blockorientierten Signalflußgraphen.
- Erhöhung der numerischen Stabilität und der Rechengeschwindigkeit bei der Simulation mechanischer und mechatronischer Systeme.

Die des weiteren neu verfügbare Hydraulikbibliothek adressiert Entwickler von hydraulischen und elektrohydraulischen Systemen,

der Mikrofluidik, sowie Anwender, die neben der Hydraulikseite auch Leistungselektronik, elektrische Antriebe, Regelungen oder Mechanik betrachten wollen.

Die grafische Repräsentation der hydraulischen Modellkomponenten orientiert sich im Unterschied zu anderen Modellierungsansätzen durchweg an den entsprechenden DIN-ISO Normen. Die Modellbibliothek beschreibt das stationäre Verhalten hydraulischer Systeme sehr genau. Die Parametrierung der vorhandenen Simulationsmodelle geschieht mit Hilfe durchschaubarer und für den Anwender einfach zu beschaffender Parameter. Die Bibliothek enthält u.a. verschiedene Pumpentypen (konstante/variable Fördermenge) und Hydraulikzylinder, eine Vielzahl von Ventilkombinationen sowie Hydromotoren und Druckspeicher. Nützliche Kleinigkeiten, wie z.B. Modelle von End- und Druckschaltern oder Manometern runden die Funktionalität ab.

Durch das Vorhandensein animierter graphischer Symbole kann die Wirkungsweise hydraulischer Systeme praxisnah visualisiert werden, so daß sich der Einsatz der hier vorgestellten Modellbibliothek auch im Ausbildungssektor anbietet.

Verfügbare Erweiterungsmodule, wie Optimierung und Programmierschnittstellen für C, Pascal, Fortran runden die Funktionalität weiter ab. Ein offenes Simulatorinterface gestattet die Kopplung mit Matlab/Simulink, MathCad, speicherprogrammierbaren Steuerungen sowie elektromagnetischer Feldberechnungssoftware auf FEM-Basis (Maxwell-EM-Solver-Produktfamilie).

### 3. Implementation der Remote-Fähigkeit

Ziel des Entwicklungsarbeiten ist die Integration von SIMPLORER in ORBIT.

SIMPLORER in der Version 4.x ist als PC-Programm unter WINDOWS 9.x NT4 WIN 2000 und WIN ME lauffähig. Ein SIMEC eigenes Kommunikationsprotokoll realisiert die Kommunikation der Programme untereinander. Eine Fernsteuerung von SIMPLORER durch Scripte ist zum gegenwärtigen Zeitpunkt nicht möglich.

Es ergaben sich folgende, zu lösende Aufgaben:

Mehrfach Instanzierung des Simulators ermöglichen Script – Steuerbarkeit Schaffung der Remote –Fähigkeit von SIMPLORER Die Arbeitspakete wurden wie folgt untersetzt:

- Mehrfach Instanzierung des Simulators
- Modifikation des Speicher-managements
- Interfacedefinition für Script Access
- Script – Steuerbarkeit
- Starten / Kontrolle der laufenden Simulation
- Modelltransport
- Ergebnistransport
- Mehrfach-Instanzierung des Simulators
- Modifikation des Speicher-managements

Die SIMPLORER-Version 4.x ist standardmäßig nicht in der Lage, mehrfach auf einem PC aufgerufen zu werden. Hauptaufgabe für dieses Arbeitspaket war die Modifikation der Berechnungsmodule und der internen Modellverwaltung. Die Anpassung erfolge in Hinblick auf die voll dynamische Speicherverwaltung verbunden mit einer schnellen Lösung der Simulationsaufgabe. Durch den Einsatz moderner Analysewerkzeuge konnte das

Management so optimiert werden, dass maximale Effektivität erreicht wurde. Ein erster Prototyp ist verfügbar und steht zu Testzwecken bereit.

Bedingt durch die Veränderungen in der Speicherverwaltung musste die Modellverwaltung vollkommen neu konzipiert und implementiert werden. Hierbei kommen objektorientierte Programmierung und ein neuer hierarchischer Ansatz für die Modellverwaltung zum Einsatz. Dieser ist gleichzeitig die Grundlage für die Parameterübernahme, die im Arbeitspaket Modelltransport realisiert werden soll.

#### 3.1. Interfacedefinition für Script Access

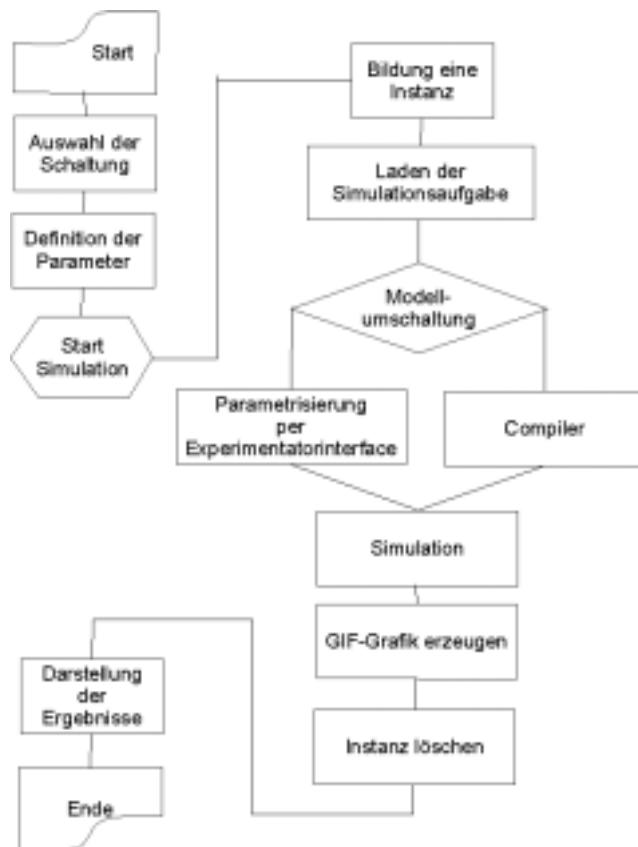
Der Modell – und Parametertransport stellt besondere Anforderungen an das System. Er muss unabhängig von der Hardware und dem Betriebssystem gestaltet sein. Gleichzeitig muss die Multiuser-Funktionalität beachtet werden. D.h. es müssen die Modelle, Parameter, Job-Steuerung und Auswertung separat steuerbar sein.

Für die erste Phase der Realisierung wurden folgende Vereinfachungen getroffen:

- Verwendung von SML – Modellen in Textform
- Vordefinierte Parameterliste
- Keine Fehlerüberwachung der Eingabe
- Keine Modellumschaltung

Durch diese Vereinfachungen können Modelle mit der aktuellen Version von SIMPLORER erzeugt und durch die Remote – Lösung von SIMPLORER simuliert werden. Der Modelltransport wird durch Steuerscripte realisiert.

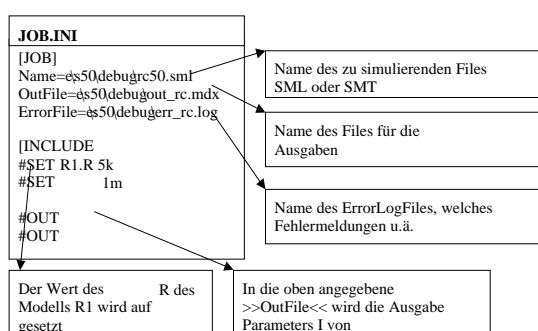
Die Abbildung zeigt den groben Ablauf, der beschrieben wird, wenn ein Anwender die Simulation startet.



Der Transport der Simulationsaufgabe wird durch das JOB – Beschreibungsfile realisiert. Dieses ASCII File beinhaltet alle Informationen, die zur Lösung der gestellten Simulationsaufgabe notwendig ist.

- Modell
- Ziel für die Ausgaben
- Ziel für Fehlermeldungen
- Definition der Modellparameter
- Definition der Ausgaben

Die folgende Abbildung illustriert den Aufbau in der 1. Version.



### 3.2. Script – Steuerbarkeit

Starten / Kontrolle der laufenden Simulation  
Der Startvorgang der Simulation wird durch den Startmanager realisiert. Dieser wurde speziell für die Bedürfnisse von SIMPLORER entwickelt. Durch den Startmanager wird gewährleistet, dass ein einfacher Aufruf von einem VB Script oder PHP - Script möglich ist. Der Manager behält die Kontrolle, bis die Simulation vollständig abgearbeitet oder abgebrochen wurde.

### 3.3. Modelltransport

Für jede Anforderung wird ein eigenes Environment erzeugt, das Modell aufgenommen und die Ausgabe vorbereitet.

### 3.4. Ergebnistransport

Die Ergebnisse speichert der Simulator direkt im vorgegebenen Environment. Die nachfolgende Bearbeitung erzeugt die notwendigen Grafiken, führt definierte Berechnungen aus und stellt alle Ausgaben als HTML Text bereit. Per http Protokoll werden diese Ausgaben zum Anwender übertragen.

## 4. Anbindung an ORBIT

Die Integration von SIMPLORER in das System ORBIT erfordert die Anbindung von SIMPLORER an das von der Firma InterConnect bereitgestellt Tool-Interface. Der ToolAgent basiert auf dem SOAP - Protokoll ( Simple Object Access Protocol ). Weitere Informationen zu SOAP sind unter [www.w3c.org](http://www.w3c.org) erhältlich.

Die Anbindung wurde für die SOAP Version 1.0 erstellt und wird auf die aktuelle SOAP Version 2.0 umgestellt. Per Test – Tool kann die Ansteuerung von SIMPLORER getestet

werden. Der verfügbare Prototyp gibt bereits Aussagen über notwendige Erweiterungen im Modelltransport - Protokoll. So fehlen zum Beispiel in der aktuellen Ausbaustufe Aussagen über die Modellausgaben. Aus SIMPLORER-Sicht ist es sinnvoll, daß der Anwender Ausgaben aus einem Fundus zu- oder abschalten kann. Das reduziert den auf der Festplatte benötigten Speicherplatz und erhöht auch die Simulationsgeschwindigkeit.

Gegenwärtig in der Diskussion ist weiterhin ein Cleaning – Mechanismus. Dieser wird notwendig, wenn man die Größe der Files von Simulationsergebnissen untersucht. Hier gilt es zu untersuchen, ob durch Komprimierung mit Standardverfahren eine Verbesserung zu erreichen ist. Besondere Beachtung gilt dabei den notwendigen Automatismen zur Komprimierung und Dekomprimierung, damit dieser Vorgang als transparent in Bezug auf den Anwender bezeichnet werden kann.

## 5. Remote - SIMPLORER Stand alone Version

Prototyp für die Remote-Simulation mit SIMPLORER. Parametrisierung von Komponenten unter Verwendung des WEB – Interfaces.

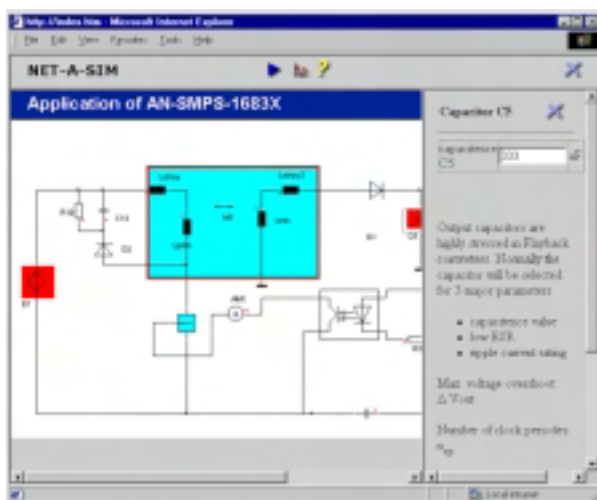


Abbildung 1: Parametereingabe

Voraussetzung für eine derartige Simulation ist das Vorhandensein und die volle Funktionsfähigkeit des Modells. Zusätzlich zum Modell werden noch die Boundaries Benötigt. Der Modellentwickler kann diese für jeden Parameter mit angeben. Aus diesen Informationen erstellen die Exportscripte den notwendigen JavaScript für den WEB-Client

Mit Hilfe diese WEB Interfaces kann der Anwender das Modell in den vorgegebenen Grenze variieren. Automatisch generierte Scripte prüfen auf der Client-Seite sofort die Eingaben auf deren Richtigkeit.

Der Anwender kann die Simulation vom Interface aus direkt starten. Es ist davon auszugehen, dass die Simulationszeit mit entsprechender Komplexität des Modells steigt. Statusinformationen informieren den Anwender automatisch über den Fortschritt der Berechnung.

Schon während der Simulation werden die Ergebnisse pro Zeitschritt zum Client übertragen. Damit erreicht man auch bei geringer Bandbreite akzeptable Simulationsgeschwindigkeiten erreichen, zumal der Hauptanteil der Last auf die Übertragung entfällt. In Kombination mit einer einfachen Online – Komprimierung der Ergebnisdaten erreicht man die geforderten Antwortzeiten des Systems.

Die Darstellung der Simulationsergebnisse erfolgen direkt im WEB- Browser unter Verwendung von JAVA- Applets. Die folgende Abbildung illustriert eine mögliche Präsentation der Simulationsergebnisse.

Als besondere Herausforderung hat sich die Darstellung in den unterschiedlichen Browser-Versionen herausgestellt.

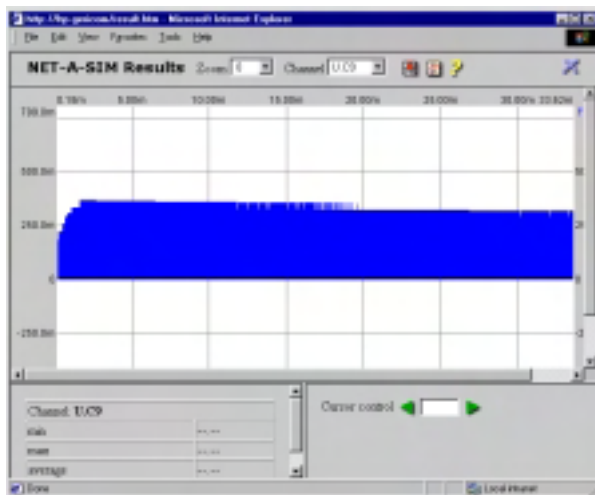


Abbildung 2 : Prototyp der Ergebnisdarstellung

## 6. Ausblick / Weiterentwicklung

Der erste Prototyp der WEB Simulation ist unter [www.net-a-sim.com](http://www.net-a-sim.com) erreichbar und demonstriert die Machbarkeit der Remote-Simulation, unter der Annahme, dass das Modell in einer vorgefertigten und lauffähigen Form vorliegt.

Die folgende Entwicklungsstufe wird sich mit dem Thema der Modellparametereingabe beschäftigen. Hier gilt es automatisierte Lösungen zu entwickeln. Eine interessante Option liegt in der Möglichkeit, dass man direkt während der Modellierung oder beim Modell-export die Parameterbereiche angeben kann. Ein weiterer Entwicklungsschwerpunkt wird die Ergebnisausgabe sein. Hier sollen Lösungen untersucht werden, wie man unter Berücksichtigung von geringen Bandbreiten eine optimale Darstellung der Ergebnisse erreicht.

# INTERNETBASIERTE DESIGNOPTIMIERUNG MIT HYGLEAM

Wilfried Jakob

Institut für Angewandte Informatik, Forschungszentrum Karlsruhe (FZK)  
Email: jakob@iai.fzk.de

Dagmar Peters

Institut für Theoretische Elektrotechnik und Mikroelektronik, Uni Bremen (ITEM)  
Email: peters@item.uni-bremen.de

Andreas Reiffer

InterConnect Software GmbH, Karlsruhe  
Email: reiffer@interconnect.de

## Kurzfassung

Mit der internetbasierten Variante von HyGLEAM können Optimierungen über einen Remoteaufruf auf einem entfernten Rechner eines Application Service Providers ausgeführt werden. HyGLEAM integriert zwei allgemein anwendbare lokale Suchverfahren in den global arbeitenden Evolutionären Algorithmus GLEAM und kombiniert die Vorteile der beteiligten Verfahren unter Vermeidung ihrer jeweiligen Nachteile. Dazu wird eine neue Methode zur Steuerung des Verfahrenseinsatzes benutzt. Die Robustheit und die gesteigerte Performance des resultierenden hybriden Verfahrens werden an Hand mathematischer Benchmarkfunktionen und dreier Designanwendungen nachgewiesen.

## 1 Einleitung

Das Verbundprojekt OMID (Optimierung von Mikrosystemen für Diagnose- und Überwachungsanwendungen) hat unter anderem zum Ziel, die im Vorgängerprojekt DEMIS entwickelten und getesteten Optimierungsverfahren zu integrieren, die dabei entstehenden Hybride zu bewerten und das resultierende Optimierungswerkzeug in einer webbasierten Remote-Version zur Verfügung zu stellen. In DEMIS wurde deutlich, dass alle untersuchten Verfahren ihre Stärken und Schwächen haben [1, 2]. So liefern lokale Suchverfahren wie der Rosenbrock-Algorithmus oder das Complex-Verfahren vergleichsweise schnell ein Ergebnis, das aber bei multimodalen Aufgabenstellungen meist nur ein lokales Optimum dar-

stellt. Globale Suchverfahren wie die Evolutionären Algorithmen (EA) bieten hingegen eine wesentlich größere Chance, das globale Optimum zu finden, benötigen dafür aber zum Teil wesentlich mehr Berechnungen der Qualitätsfunktion. Um die meist aufwendigen Simulationen zur Bestimmung der Qualität einer Designlösung zu verringern ohne den globalen Charakter der Suche aufzugeben, wurde versucht, die beiden Algorithmenklassen so geschickt miteinander zu verbinden, dass der resultierende Hybrid die Vorteile der beteiligten Verfahren ohne deren Nachteile in sich vereint. Erste Tests an Designoptimierungsaufgaben aus dem laufenden Projekt zeigten vielversprechende Ergebnisse [3].

In Kap. 2 werden die am FZK entwickelten Hybridisierungsmethoden [4, 5] vorgestellt. Integriert wurden zwei am ITEM implementierte lokale Suchverfahren [1] und das GLEAM-Verfahren (General Learning Evolutionary Algorithm and Method) [6, 7], das am FZK zu HyGLEAM (Hybrid General Purpose Evolutionary Algorithm and Method) weiterentwickelt und als GADO (Genetischer Algorithmus zur Designoptimierung) und HyGADO zur Designoptimierung eingesetzt wurde. Kap. 3 beschreibt die untersuchten Testfälle bestehend aus mathematischen Benchmarkfunktionen und OMID-Anwendungen. In Kap. 4 sind die Testergebnisse und die sich daraus ergebenden Schlussfolgerungen enthalten. Kap. 5 stellt die auf dem Orbit-System der Firma InterConnect [8] basierende Remote-Lösung von HyGLEAM vor, während Kap. 6 mit einer Zusammenfassung des Beitrags schließt.

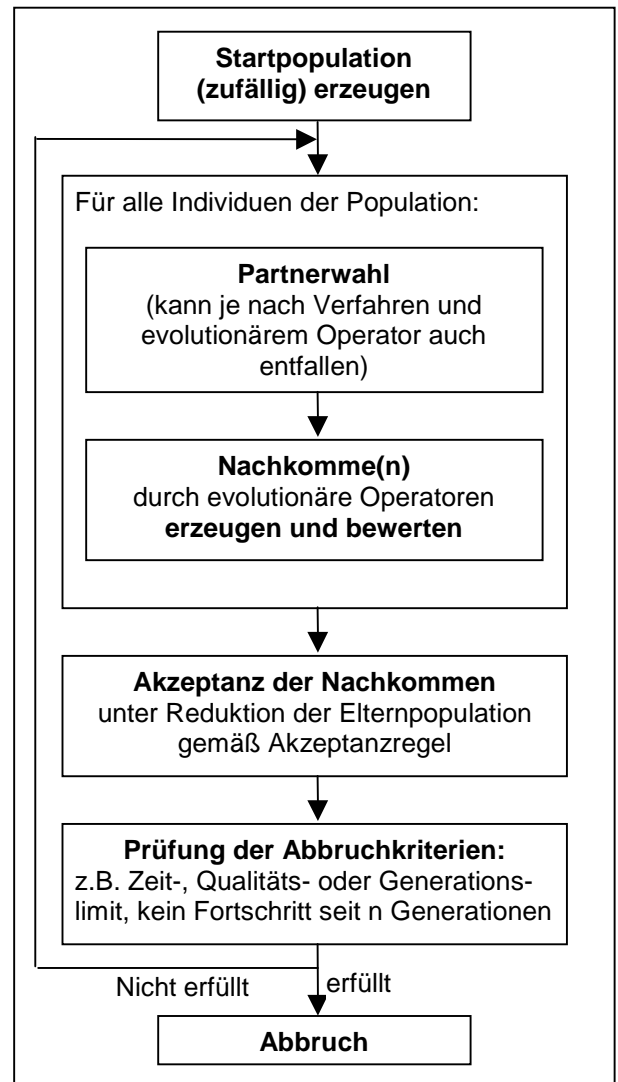
## 2 Das HyGLEAM-Konzept

Das in [4, 5] ausführlicher dargestellte Konzept von HyGLEAM kann hier nur in der gebotenen Kürze erläutert werden. Zu den beiden lokalen Verfahren gibt Schwefel [9] eine ausführlichere Beschreibung mit weiteren Anwendungsfällen.

### 2.1 Benutzte Optimierungsverfahren

#### 2.1.1 GLEAM

GLEAM ist ein eigenständiger evolutionärer Algorithmus, der Elemente der Evolutionsstrategie und der (reellcodierten) Genetischen Algorithmen mit Konzepten der Informatik (abstrakte Datentypen) verbindet. Die Codierung der zu optimierenden Parameter erfolgt in den sogenannten Aktionen. Eine Aktion besteht aus ein oder mehreren Parametern vom Typ Bool, Integer oder Real entsprechend ihrem anwendungsspezifisch definierten Aktionstyp. Eine Aktion drückt damit eine bestimmte Eigenschaft eines Designs aus, z.B. die Anwesenheit einer bestimmten Komponente, die durch die Aktionsparameter näher bestimmt wird oder die Relationen zu anderen Komponenten entsprechend den Aktionsparametern eingeht. Im einfachsten Fall codieren die Aktionen nur gewisse einzelne Parameter eines Designs. Die Aktionen bilden eine Aktionsskette (AK), die den Chromosomen des biologischen Vorbilds entspricht. Es gibt derzeit in GLEAM drei AK-Typen: Bei den ersten beiden ist jeder Aktionstyp mit genau einer Aktion vertreten und sie unterscheiden sich darin, ob die Reihenfolge der Aktionen relevant ist. Die Variante ohne bedeutungstragende Reihenfolge wird bei Designoptimierungsaufgaben, bei denen es um reine Parameteroptimierung geht, verwendet. Die andere Variante spielt bei OMID keine Rolle, sie kommt aber bei anderen Anwendungen wie z.B. Schedulingproblemen zum Einsatz [10]. Schließlich gibt es noch den AK-Typ mit variabler Länge, bei dem jeder Aktionstyp mit keiner, einer oder mehreren Aktionen in beliebiger Reihenfolge vertreten sein kann. Dieser Typ findet bei anspruchsvolleren Designoptimierungsaufgaben wie z.B. der in Kap. 3 beschriebenen Akkorplatte Verwendung.



**Abb. 1: Prinzipieller Ablauf Evolutionärer Algorithmen**

Abb. 1 zeigt den prinzipiellen Ablauf Evolutionärer Algorithmen. Ausgehend von einer meist zufällig erzeugten Startpopulation, die durch manuell oder auf andere Weise erzeugte Lösungen angereichert werden kann, werden alle Individuen generationsweise der nachgebildeten Evolution unterworfen. Dazu werden Nachkommen durch Mutation und/oder durch Rekombination des Genmaterials zweier Eltern erzeugt. Im letzten Fall findet zuvor eine Partnerwahl statt, die meist auf der Fitness der Individuen beruht. In GLEAM wird dafür eine ranking-basierte Selektion benutzt, die die Chancen der besten Individuen, selektiert zu werden, etwas zu Gunsten der schlechteren dämpft. Dadurch soll unter anderem vorzeitiger Konvergenz auf ein Suboptimum vorgebeugt werden. GLEAM enthält einen Satz vordefinierter genetischer Operatoren (Muta-

tionen der Werte oder der Aktionsposition, verschiedene Crossover-Operatoren), die aufgabenspezifisch konfiguriert zur Anwendung gebracht werden können. Die Wertemutationen berücksichtigen die in den Aktionstypdefinitionen mithinterlegten Wertebereichsgrenzen, sodass explizite Restriktionen immer eingehalten werden. Kleine Änderungen sind dabei entsprechend der Evolutionsstrategie wahrscheinlicher als große. Pro Paarung werden in GLEAM mehrere Nachkommen erzeugt, wobei die dabei beteiligten genetischen Operatoren sowie ihre Anwendungswahrscheinlichkeiten problemspezifisch ausgewählt und konfiguriert werden können. Schließlich müssen die Nachkommen der Elterngeneration akzeptiert oder verworfen werden. Die akzeptierten Kinder ersetzen ebensoviel Mitglieder der Elterngeneration, sodass die Populationsgröße konstant bleibt. In GLEAM wird dazu eine elitäre Akzeptanzregel benutzt, bei der das beste Individuum unabhängig von seinem Alter immer überlebt, sodass eine Verschlechterung der pro Generation erreichten Lösungsqualität nicht eintreten kann.

### 2.1.2 Verfahren nach Rosenbrock

Kernprinzip des Verfahrens nach Rosenbrock [11] ist die Suche des Optimums entlang der Richtungen eines im  $n$ -dimensionalen Raum rotierenden Koordinatensystems. Beschränkungen werden mit einer partiellen, internen Straffunktion berücksichtigt.

Das in Abb. 2 dargestellte Verfahren setzt einen zulässigen Startpunkt voraus. Im Anschluss an die Initialisierung wird ein Testschritt in die aktuelle Richtung  $i$  gemacht. Sind alle Bedingungen erfüllt und ist der Wert der partiellen Barrierefunktion ermittelt worden, wird untersucht, ob es sich bei dem aktuellen Punkt um einen besseren handelt als bei dem Vorgänger. Ist dies der Fall, vergrößert sich die Schrittweite um den Faktor drei. Wurde mit dem aktuellen Punkt kein Erfolg erzielt, wird die Suche in die entgegengesetzte Richtung mit halbiertem Schrittweite fortgesetzt. Innerhalb einer Iteration muss in jede Richtung des aktuellen Koordinatensystems mindestens ein Erfolg gefolgt von einem Misserfolg erzielt werden, bevor das Abbruchkriterium überprüft wird.

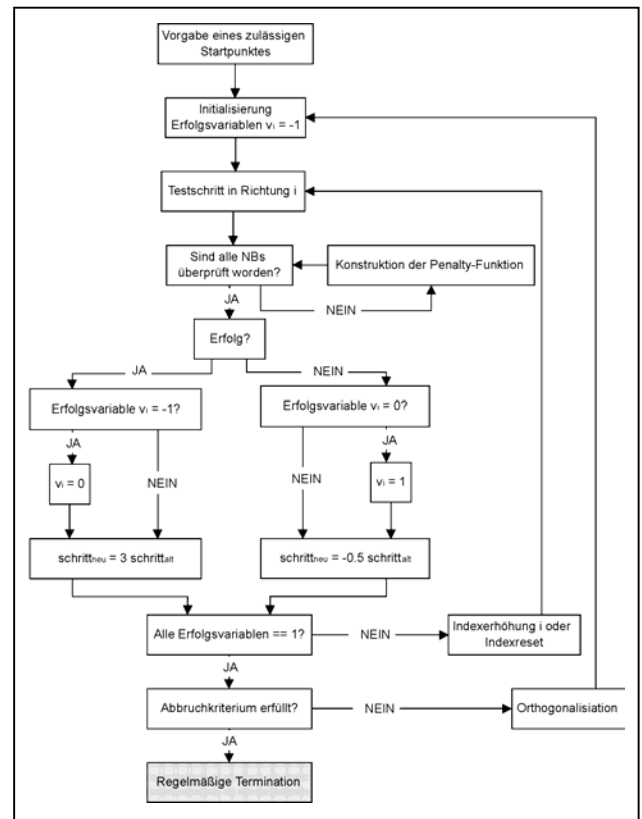


Abb. 2: Ablaufdiagramm des Rosenbrock-Verfahrens

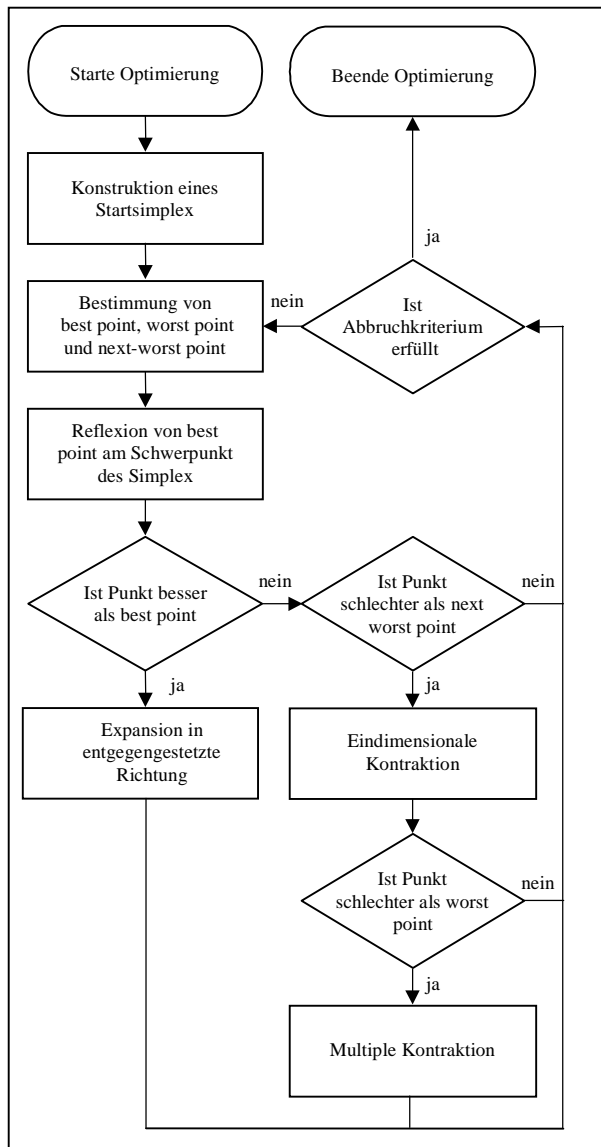
Wird das Abbruchkriterium erfüllt, terminiert das Verfahren, ansonsten wird eine Rotation des Koordinatensystems durchgeführt. Dazu wird die vielversprechendste Richtung der letzten Iteration als Hauptachse des neuen Koordinatensystems festgelegt und die restlichen  $n-1$  Achsen orthogonal zu dieser Achse aufgebaut. Mit diesem Koordinatensystem setzt sich die Suche des Optimums so lange fort, bis das Abbruchkriterium erfüllt wird.

### 2.1.3 Complex-Algorithmus

Das Complex-Verfahren wurde 1965 von Box [12] in Anlehnung an den Simplexalgorithmus für nichtlineare Funktionen von Nelder und Mead entwickelt. Der prinzipielle Ablauf ist in Abb. 3 dargestellt. Der Complex-Algorithmus wurde speziell für beschränkte Problemstellungen entwickelt und verwendet keinen der klassischen Ansätze: externe Straffunktion, interne Straf- bzw. Barrierefunktionen, Lagrange-Multiplier-Verfahren oder partielle Straffunktion.

Ausgehend von einem zulässigen Startpunkt, wird der Startcomplex gebildet. Es folgt die Ermittlung des schlechtesten Punktes sowie





**Abb.3: Prinzipieller Ablauf von Polyederverfahren**

des Scheitelpunktes der verbleibenden Punkte. Ist der Scheitelpunkt zulässig, wird der schlechteste Punkt an ihm reflektiert. Dabei handelt es sich um eine Normalreflexion, bei der gleichzeitig eine Expansion des Complex durchgeführt wird. Bei Nichtzulässigkeit des Scheitelpunktes wird ein neuer Startcomplex gebildet und das Verfahren neu gestartet. Auch der reflektierte Punkt muss daraufhin überprüft werden, ob er alle formulierten Bedingungen erfüllt. Ist dies nicht der Fall, wird der Reflexionsfaktor halbiert und eine erneute Reflexion durchgeführt, bis ein zulässiger Punkt gefunden wird oder der Reflexionsfaktor kleiner als eine bestimmte untere Grenze geworden ist. In diesem Fall wird der Reflexionsvorgang mit dem nächst schlechtesten

Punkt wiederholt. Kann für keinen Punkt des Complex ein zulässiger Reflexionspunkt gefunden werden, bricht das Verfahren vorzeitig ab. Ein zulässiger Reflexionspunkt wird daraufhin untersucht, ob mit ihm ein verbesserter Wert der Zielfunktion erzielt wird. Ist dies der Fall, wird das Abbruchkriterium überprüft; ist es erfüllt, terminiert das Verfahren regelmäßig, ansonsten wird der schlechteste Punkt ermittelt und die nächste Iteration des Optimierungsprozesses gestartet.

## 2.2 Hybridisierungsarten

Die Hybridisierung beruht auf folgenden vier Ansätzen:

- **Initialisierung der Startpopulation**  
Die Startpopulation wird ganz oder teilweise mit den Ergebnissen von Läufen des lokalen Verfahrens initialisiert. Ein eventueller Rest wird zufällig bestimmt. Der Grundgedanke besteht darin, die evolutionäre Suche bereits mit Lösungen einer gewissen Qualität beginnen zu lassen und so die Zeit zum Auffinden befriedigender Bereiche des Suchraums zu sparen.
- **Nachoptimierung der GLEAM-Ergebnisse**  
Die Nachoptimierung der GLEAM-Ergebnisse mit konventionellen Verfahren wird häufig auch als „*Local Hill Finding + Local Hill Climbing*“ bezeichnet. Man benutzt den EA zum Auffinden einiger vielversprechender Regionen des Suchraums und verwendet anschließend ein lokales Verfahren zur Bestimmung des exakten (lokalen) Optimums. Das beste dieser Optima ist dann die Lösung. Auf die Frage, wann der EA sinnvoll beendet und die lokale Suche begonnen werden kann, wird im nächsten Abschnitt eingegangen.
- **Direkte Integration**  
Hierbei werden alle oder nur der beste Nachkomme einer Paarungsoperation von GLEAM mit dem lokalen Verfahren optimiert. Damit operiert GLEAM quasi nur noch über den Bergspitzen der Fitnesslandschaft.
- **Verzögerte Direkte Integration**  
Bei vielen Aufgabenstellungen bringt eine lokale Optimierung ausgehend von

schlechten Startwerten meist nicht viel. Daher erscheint es sinnvoll, die direkte Integration erst nach einer anfänglichen Phase evolutionärer Suche dazuschalten.

Zu diesen vier Hybridisierungsarten kommen noch die Kombinationen von Voroptimierung und (verzögerter) direkter Integration. Bei der Nachoptimierung mit dem Complex wurden zwei Arten untersucht: Entweder wurden alle von GLEAM gelieferten Lösungen zur Bildung eines einzigen Start-Complexes genutzt oder sie dienten einzeln als Startpunkte für entsprechend viele Complex-Läufe. Daraus ergaben sich mit den beiden lokalen Verfahren 13 verschiedene Hybridisierungsarten.

### 2.3 Steuerung der Hybridisierung

Die auf dem OMID-Statusseminar [3] bereits ausführlich dargestellten Verfahren zur Steuerung der Hybridisierung beruhen auf der genotypischen Varianz in der Population. Diese wird nur bei auftretender Fitness-Stagnation bestimmt, um Rechenzeit zu sparen. Wenn es in der Population weniger als  $N_{max}$  Nischen von Individuen gibt, deren genotypische Differenz geringer als ein vorgegebenes  $\epsilon$  ist und alle Nischen eine geringere Differenz als  $\epsilon_{Pop}$  aufweisen, gilt die Population als „konvergiert“. Dann wird die Evolution bei der Nachoptimierung abgebrochen und alle Nischen

steuern einen Vertreter zum Gesamtergebnis bei, das noch um den Populationsbesten ergänzt wird, sofern er besser als alle Nischenvertreter ist. Diese Individuen werden mit dem jeweiligen lokalen Verfahren nachoptimiert und das beste ist schließlich das Endergebnis.

Im Falle der verzögerten direkten Integration wird das Konvergenzkriterium weniger streng als bei der Nachoptimierung parametrisiert. Es bestimmt den Zeitpunkt der Zuschaltung der lokalen Suche.

## 3 Testfälle

Untersucht wurden zusätzlich zu den in DEMIS [1] bereits benutzten Testfällen noch die Testfunktion nach Fletcher und Powel [9, Probl. 2.13] und die im Statusseminar [3] bereits behandelten Anwendungen aus dem laufenden Projekt. Tab. 1 zeigt wichtige Eigenschaften der Testfälle und gibt die Ergebnisse für die einzelnen Verfahren an. Alle Experimente wurden mit vergleichbaren Einstellungen der beteiligten Verfahren durchgeführt. Variiert wurden wie in [3] dargestellt lediglich beim Rosenbrock die Abbruchschranke und bei GLEAM die Populationsgröße. Abweichungen von den in [1] publizierten Ergebnissen erklären sich bei den beiden lokalen

Experiment	Parame- -teranz.	Modalität	Implizite Restrikt.	Genau- igkeit	Erfolgsrate [%] und Aufwand [Evaluationen]					
					GLEAM		Rosenbrock		Complex	
Bracken	2	multimod.	ja	1,0E-03	100	2509	4	1074	70	61
				1,0E-06	100	5000	-	-	98	196
Beale	3	unimod.	ja	1,0E-03	100	28936	37	4854	100	164
				1,0E-06	100	76229	2	1538	98	269
Bigg	5	multimod.	nein	1,0E-03	100	3370501	17	4001	20	164
				1,0E-06	-	-	18	18544	10	799
Helical Valley	3	unimod.	nein	1,0E-06	100	47288	71	15062	-	-
Rastrigin	20	multimod.	nein	1,0E-04	100	51721	-	-	-	-
Fletcher & Powel	5	multimod.	nein	1,0E-05	100	483566	10	1090	10	243
Fraktale Fkt.	20	multimod.	nein	Ziel: -0.05	100	195100	-	-	-	-
Heterodyn-Empf.	3	multimod.	nein		100	5773	8	94	12	102
Drucksensor	3	multimod.	nein		100	29791	-	-	-	-
Aktorplatte	dynam. n x 4	multimod.	ja							

Tabelle 1: Wichtige Eigenschaften der Testfälle und Ergebnisse der drei Verfahren

Verfahren durch den Verzicht, weitere Parameter wie z.B. die initiale Schrittweite anzupassen. Bei der Testfunktion nach Bracken wurde bei GLEAM (GADO) seinerzeit versehentlich der Aufwand zum Auffinden des exakten Optimums anstelle einer  $10^{-3}$ -Genauigkeit angegeben.

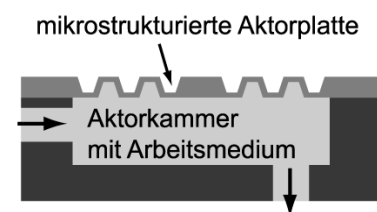
In den drei Ergebnisspalten von Tab. 1 wird die Erfolgsrate in Prozent und der durchschnittliche Aufwand in Evaluationen basierend auf jeweils 100 Läufen für die jeweils beste Parametrierung angegeben. Ein – bedeutet, dass das betreffende Verfahren nicht in der Lage war, das Problem unter den gegebenen Einstellungen zu lösen. Beim Rosenbrock wurden folgende normierte Abbruchschranken verwendet: 0.1 (s), 0.01 (n),  $10^{-4}$  (m),  $10^{-6}$  (h) und  $10^{-8}$  (x) und bei GLEAM wurde die Populationsgröße variiert. Ab der fraktalen Funktion sind in der Tabelle keine Genauigkeitsangaben vorhanden, da die exakte Lösung nicht bekannt ist. Bei der Aktorplatte handelt es sich auf Grund Ihrer aufwendigen Simulation um eine „echte Anwendung“ in dem Sinne, dass keine Läufe zum Verfahrensvergleich möglich sind, sodass hier nur die Eigenschaften gelistet sind.

Der Heterodynempfänger [2,13] ist eine Komponente eines mikrooptischen Kommunikationssystems bestehend aus zwei Kugellinsen und einer Photodiode. Der Abstand zwischen dem Lichtleiter und erster Kugellinse sowie die beiden Brechungsindizes sollen so bestimmt werden, dass nicht nur eine optimale Ausleuchtung der Photodiode im Kollimationspunkt erreicht wird, sondern die ganze Anordnung auch möglichst unempfindlich gegenüber Fertigungstoleranzen innerhalb vorgegebener Grenzen ist. Letzteres macht die ansonsten einfache Aufgabe zu einem trotz der nur drei Parameter schwierigen, stark multimodalen Optimierungsproblem.

Die Wandlung der mechanischen Größe Druck in ein elektrisch messbares Signal erfolgt bei dem betrachteten Drucksensorsystem [2] nach dem kapazitiven Wandlerprinzip. Der Drucksensor stellt eine Kondensatoranordnung dar, die aus der Sensormembrane (bewegliche Elektrode), der darunter liegenden Druckkammer und der implantierten n-Wanne (feste Elektrode) besteht. Bei der Designopti-

mierung des Drucksensors wurden zwei Kriterien berücksichtigt: die Sensitivität und die Linearität im Betriebsbereich. Optimierungsparameter sind die Kammerhöhe und –länge sowie die Membrandicke.

Der in Bild 4 dargestellte Mikroaktor [14] besteht aus zwei übereinander geschichteten dünnen Kunststofffolien, in denen mit einem Excimerlaser für die Aktorplatte, Aktorkammer und fluidischen Anschlüssen die entsprechenden Mikrostrukturen eingebracht wurden. Durch eine entsprechende Anzahl, Anordnung und Dimensionierung der ringförmigen und konzentrisch angeordneten Kerben kann das mechanische Verhalten der Aktorplatte gezielt beeinflusst werden. Ein wichtiges Optimierungsziel bei dem Kerbendesign besteht darin, bei einem gegebenen Arbeitsdruck eine möglichst große Auslenkung der Aktorplatte zu erreichen. Da die Aktorplatte gleichzeitig eine lange Lebensdauer haben soll, besteht das zweite Optimierungsziel darin, die Spannung in der Aktorplatte bei Auslenkung möglichst gering zu halten.



**Abb.4: Schematische Darstellung der Aktorplatte**

Die hier vorliegende Optimierungsaufgabe beschränkt sich somit nicht auf eine rein geometrische Optimierung, sondern sie umfasst zusätzlich die Optimierung der Topologie in Form von Anzahl und Anordnung der Kerben. Dies führt zu einem Parametervektor dynamischer Länge (pro Kerbe 4 Parameter), wodurch die meisten deterministischen und stochastischen Optimierungsverfahren nicht mehr anwendbar sind. Auch die Implementierungen vieler EAs erlauben im Gegensatz zu GLEAM keine Chromosomen dynamischer Länge, obwohl die Methode dies zulässt. Durch die diskrete Anzahl der Kerben ist der Lösungsraum zudem multimodal. Aufgrund der Genauigkeitsanforderungen und der Topologievariabilität ist der Einsatz von FEM-Modellen bei der Bewertung einzelner Lösungsvorschläge unumgänglich. Dies führt auf

einer Sun-Workstation (Ultra Sparc 10) zu Evaluationszeiten zwischen 3 und 10 Minuten. Die Optimierungen wurden daher mit einer Parallelversion von GLEAM durchgeführt [15]. Diese Ergebnisse dienen als Referenzwerte für die hier vorgestellten Resultate der hybriden Optimierung.

## 4 Ergebnisse

Die auf dem Statusseminar präsentierten Ergebnisse [3] und die Resultate weitergehender Untersuchungen [5] führen hinsichtlich der Hybridisierungsarten zu folgenden Schlussfolgerungen:

1. Die Vorooptimierung bringt bei einigen Testfällen eine gewisse Verbesserung, wobei einmal der Complex und ein anderes mal der Rosenbrock besser abschneidet.
2. Die Nachoptimierung verbessert zwar in der Regel die Ergebnisse der Evolution, erreicht aber lange nicht die Zuverlässigkeit bei der Lösungsfindung wie GLEAM.
3. Die (verzögerte) direkte Integration des Rosenbrock findet immer das Optimum (**GR** und **GvR**). Bei Verwendung des Complex (**GC** und **GvC**) wird meist weniger Aufwand benötigt als beim Rosenbrock, sofern die Complex-Integration als solche funktioniert.

Daher werden im folgenden nur die verzögerte und unverzögerte direkte Integration behandelt und ihre Resultate verglichen.

### 4.1 Ergebnisse der Testfunktionen

Bei der direkten Integration wurde Lamarck'sche Evolution verwendet, das heißt, dass der lokal verbesserte Nachkomme im Falle seiner Akzeptanz auch genotypisch an das gefundene (lokale) Optimum angepasst wurde. Bei der verzögerten direkten Integration werden die in Tab.2 angegebenen drei Parametrierungen für die Nischenbildung verglichen:

Parametrierung	$\epsilon$	$\epsilon_{Pop}$
P1	0.005	0.01
P2	0.002	0.005
P3	0.001	0.003

**Tabelle 2: Steuerparameter**

$N_{max}$  variiert in Abhängigkeit von der Populationsgröße P: 2 für  $P < 30$ , 3 für  $P = 30$ , 4 für  $P = 50$  und 5 für  $P > 50$ . Um die Ergebnisse der lokalen Verfahren mit GLEAM und den HyGLEAM vergleichen zu können, muss die Anzahl der Läufe bestimmt werden, bei der eine vergleichbare Sicherheit besteht, den Zielwert zu finden. Angesichts eines Stichprobenumfangs von 100 mag eine Sicherheit von 95,5% genügen. Damit ergeben sich die in Tab. 3 dargestellten Resultate.

Testfunktion	Genauigkeit	GLEAM	Aufwand bei 99,5% Sicherheit		HyGLEAM			
			Rosenbrock	Complex	GR	GvR	GC	GvC
Bracken	$10^{-3}$	2 509	139 620	305	31 264	21 834	439	545
	$10^{-6}$	5 000	-	392	314 834	546 499	1 433	1 785
Beale	$10^{-3}$	28 936	58 248	164	14 793	18 088	811	897
	$10^{-6}$	76 229	229 162	538	48 300	48 613	1 319	1 419
Bigg	$10^{-3}$	3 370 501	120 030	3 936	300 247	5 675 628	24 493	136 190
	$10^{-6}$	-	500 688	40 749	556 809	7 663 112	30 143	127 627
Helical Valley	$10^{-6}$	47 288	75 310	-	67 950	8 118	2 576	2 674
Rastrigin	$10^{-4}$	51 721	-	-	43 864	47 378	-	4 267 176
Fletcher&Powel	$10^{-5}$	483 566	55 590	12 393	13 535	10 206	4 684	4 635
Fraktale Fkt.		195 100	-	-	30 626	30 473	56 287	952 620

**Tabelle 3: Vergleich der benötigten Individuen**

(jeweils 100 von 100 Läufen erfolgreich oder hochgerechnet auf 95,5% Erfolgswahrscheinlichkeit)

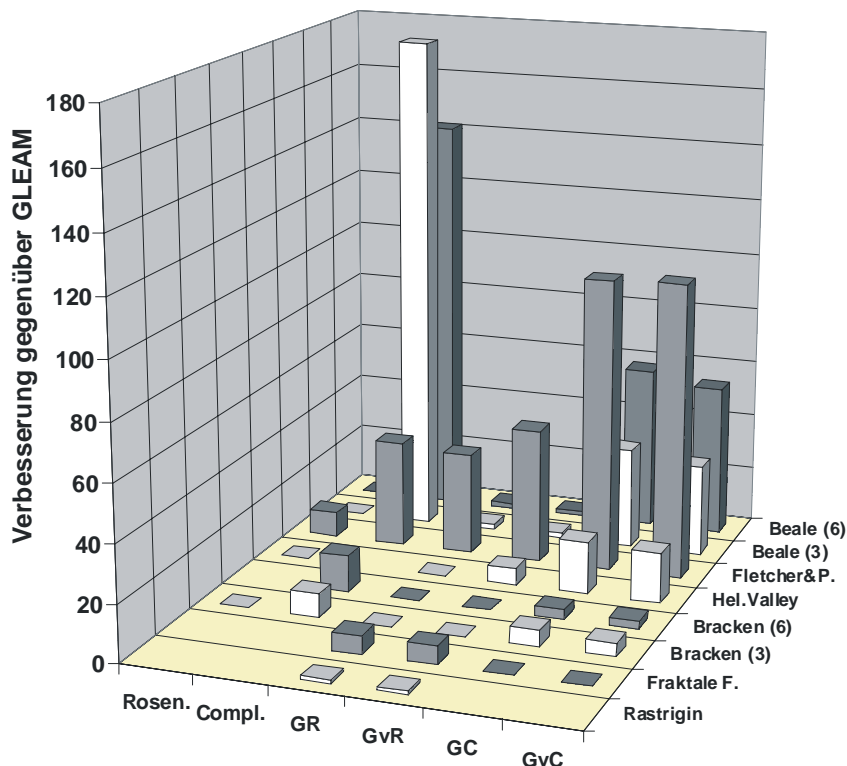
Es fällt auf, dass der Complex in den Fällen, in denen er konvergiert, das beste Ergebnis der drei Verfahren liefert. Tab. 4 und Abb. 5 vergleichen die erreichten Verbesserungen gegenüber GLEAM: kein Eintrag bedeutet, dass die Aufgabe nicht gelöst werden konnte; eine Null bzw. ein flaches Feld, dass zur erfolgreichen Lösung mehr Aufwand notwendig war. Die Testfunktionen sind entsprechend der besseren Übersichtlichkeit im Diagramm

sortiert. Die geforderte Genauigkeit des Zielwertes ist in Klammern angegeben. Bigg (6) wurde in der Tabelle weggelassen, da ein Vergleich mit GLEAM nicht möglich ist. Für die Hybridisierung ist der Fall aber insofern ein Erfolg, als dass die direkte Complex-Integration am besten abschneidet. In Abb. 5 musste Bigg (3) aus Gründen der Darstellbarkeit weggelassen werden.

Test-Funktionen	Verbesserung gegenüber GLEAM						Parametrierungen					
	Rosen.	Compl.	GR	GvR	GC	GvC	GLEAM	Rosen.	GR	GvR	GC	GvC
Rastrigin			1,2	1,1			5		5 m	5 m P3		
Fraktale F.			6,4	6,4	0	0	20		5 n	5 n P2	10	5 P2
Bracken (3)	0	8,2	0	0	5,7	4,6	20	m	5 n	5 n P2	5	5 P2
Bracken (6)		12,8	0	0	3,5	2,8	10		5 n	5 n P2	5	5 P2
Hel.Valley	0		0	5,8	18,4	17,7	5	m	5 m	10 m P2	5	5 P2
Fletcher&P.	8,7	37,4	35,7	47,4	103,3	104,2	600	m	10 m	5 m P2	5	5 P3
Beale (3)	0	176,4	2,0	1,6	35,7	32,3	10	x	5 n	5 n P2	5	5 P2
Beale (6)	0	141,7	1,6	1,6	57,8	53,7	10	x	5 h	5 x P2	5	5 P2
Bigg (3)	28,1	856,3	11,2	0	137,6	24,7	500	h	30 m	20 m P2	20	5 P2

**Tabelle 4: Verbesserung gegenüber GLEAM und beste Parametrierungen**

(Parametrierungen: Populationsgröße, Rosenbrock-Abbruchschranke, Nischenparametrierung (s. Tab.2))



**Abb.5: Vergleich der erreichten Verbesserungen des jeweils besten Laufs gegenüber GLEAM.**

(kein Eintrag: Zielwert nicht erreicht, flacher Eintrag: Zielwert bei höherem Aufwand als GLEAM erreicht)

Man kann feststellen, dass die hinsichtlich

Parameteranzahl und Suchraumeigenschaften

einfacheren Funktionen Bracken, Beale und Bigg am besten durch den Complex gelöst werden und die breiter angelegte globale Suche von GLEAM zu aufwendig für das Problem ist oder im Falle von Bigg GLEAM ohne Unterstützung des Complex-Verfahrens kaum eine Chance hat. Bei den schwierigeren Problemen versagt bis auf Fletcher & Powel der Complex und die Hybridisierung schneidet regelmäßig am besten ab. Allerdings arbeitet GLEAM dabei mit wechselnden Partnern. Betrachtet man nun die Parametrierungen der erfolgreichsten Läufe auf der rechten Seite von Tab. 4, so kann man folgendes feststellen: Alle hybriden Varianten kommen mit kleinen Populationsgrößen zwischen 5 und 30 aus, die Abbruchschranke des Rosenbrock-Verfahrens nimmt alle Werte zwischen  $n$  (niedrig) und  $x$  (sehr hoch) an und die Nischenparametrierung P2 dominiert. Es gibt also keine allgemein-günstige Parametrierung. Insbesondere ist unklar, welches lokale Verfahren besser geeignet ist und im Falle des Rosenbrock, mit welcher Abbruchschranke gearbeitet werden soll. Darauf wird im letzten Abschnitt dieses Kapitels noch weiter eingegangen werden.

## 4.2 Ergebnisse der Designoptimierungen

Der Vergleich der verschiedenen Verfahren und Parametrierungen erfolgt auf der Basis der zur Erreichung einer bestimmten Designqualität notwendigen Evaluationen. Da die Ergebnisse bereits auf dem Statusseminar [3] ausführlich dargestellt wurden, wird hier nur eine Zusammenfassung gegeben.

### 4.2.1 Heterodynempfänger

Wegen des relativ hohen Simulationsaufwands wurden jeweils nur 50 Läufe durchgeführt. Die lokalen Verfahren erreichen das Qualitätsziel mit unterschiedlicher Zuverlässigkeit, was in Tab. 5 mit den sich daraus ergebenden Aufwänden für eine Sicherheit von 99,5% dargestellt ist. GLEAM erreicht ab einer Populationsgröße von 150 zuverlässig die Zielqualität, wobei der niedrigste Aufwand bei einer Populationsgröße von 180 entsteht, siehe Abb.6.

Verfahren	Rosen. Abbr.	Erfolg [%]	Aufwand [Eval.]	95,5% Sicherheit:	
				Läufe	Eval.
Rosenbrock	n	8	94	64	6 016
	m	6	232	86	19 952
	h	30	764	15	11 460
Compl.		12	102	42	4 284

Tabelle 5: Ergebnisse der lokalen Verfahren

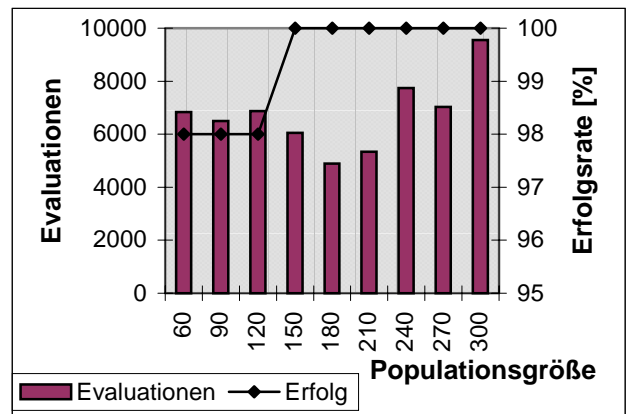


Abb.6: Aufwand und Erfolgsrate von GLEAM

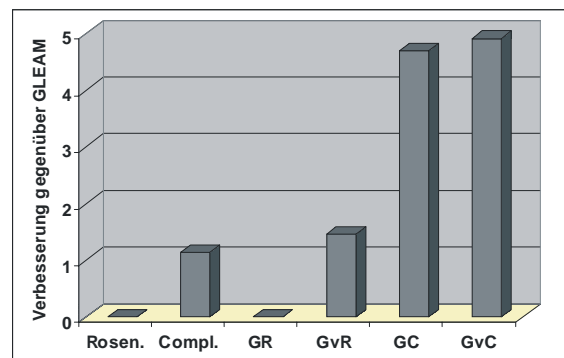


Abb.7: Vergleich der erreichten Verbesserungen des jeweils besten Laufs gegenüber GLEAM

Abb. 7 vergleicht die erreichten Verbesserungen gegenüber GLEAM basierend auf den jeweils besten Läufen der Hybriden und der berechneten Werte für die lokalen Verfahren. Alle Hybriden arbeiten mit einer Populationsgröße von 5 und die günstigste Hybridisierung GvC mit der Nischenparametrierung P3.

### 4.2.2 Drucksensor

Die Ergebnisse basieren auf 100 Läufen pro Parametrierung und Verfahren. Bei der hier benutzten strengeren Bewertung als in [2] erreichen die beiden lokalen Verfahren das Optimierungsziel in maximal 3% der Fälle (Rosenbrock) bzw. gar nicht (Complex). Bei GLEAM wird erst ab einer Populationsgröße

von 210 das Qualitätsziel zuverlässig erreicht und ab 360 auch bei niedrigen Evaluationsraten.

Die direkte Integration des Rosenbrock-Verfahrens mit Abbruchschranke  $s$  bringt hingegen stabile Verhältnisse, siehe Abb. 8. Der Aufwand ist mit dem der besten GLEAM-Läufe über einen weiten Bereich der Populationsgröße vergleichbar. Am besten schneidet eine Begrenzung der Iterationen des Rosenbrock auf 100 ab (Ros-100), während bei 250 (Ros-250) etwas mehr Evaluationen bei gleichem Erfolg benötigt werden. Andere Abbruchschranken führen zu deutlich schlechteren Ergebnissen. Die Integration mit dem Complex führt hier hingegen zu einem Mehraufwand.

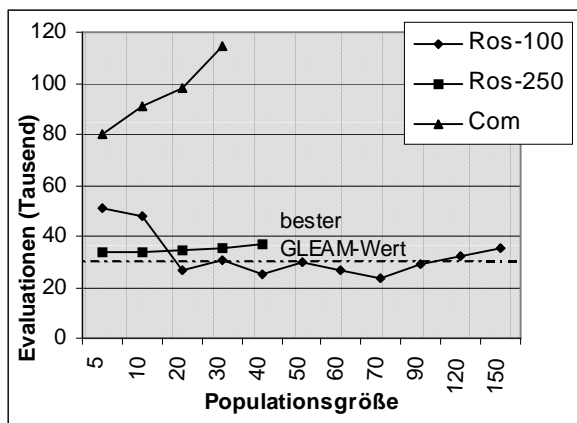


Abb.8: Direkte Integration

### 4.2.3 Aktorplatte

Bei der Aktorplatte verbieten sich auf Grund der langen Simulationszeiten vergleichende Läufe mit unterschiedlichen Parametrierungen. Stattdessen sind die bisher gemachten Erfahrungen praktisch anzuwenden. Verglichen werden Ergebnisse, die mit einer Parallelversion von GLEAM erzielt wurden [15], mit HyGLEAM unter Verwendung des Rosenbrock-Verfahrens bei direkter Integration und mittlerer Abbruchschranke, wobei die Hälfte der Startpopulation vorgeneriert und voroptimiert wurde [14]. Tab. 6 zeigt die Werte eines anfänglichen manuellen Designs und die der beiden Optimierungsergebnisse. Beim Vergleich des Aufwands ist folgendes zu beachten: Die Komponente zur Ankopplung des FEM-Simulators speichert die simulierten Parametervektoren zusammen mit den Ergebnissen ab und vergleicht neue Anforderungen mit den bereits errechneten, um so Simulationszeit zu sparen. Dies kommt besonders beim Rosenbrock-Verfahren zum Tragen, da in der Konvergenzphase relativ häufig sehr ähnliche Vektoren berechnet werden, die vom Filter als identisch bewertet werden. Dementsprechend wird in der Tabelle zwischen angeforderten und ausgeführten Simulationen unterschieden. HyGLEAM benötigt also nur etwa die Hälfte an real durchgeführten Simulationen, um ein besseres Ergebnis zu erzielen als die nichthybride Variante. Dies führte zu einer Laufzeit von 2 Tagen und 15 Stunden auf 8 Suns unterschiedlicher Leistung (504 CPU-Stunden) bei GLEAM, während HyGLEAM mit 12 Tage und 16 Stunden (304 CPU-Stunden) auf einer Sun Ultra Sparc 10 auskam. Neuere Ergebnisse zu variierenden Drücken und E-Modulen sind in [16] im gleichen Band zu finden.

rungen mit den bereits errechneten, um so Simulationszeit zu sparen. Dies kommt besonders beim Rosenbrock-Verfahren zum Tragen, da in der Konvergenzphase relativ häufig sehr ähnliche Vektoren berechnet werden, die vom Filter als identisch bewertet werden. Dementsprechend wird in der Tabelle zwischen angeforderten und ausgeführten Simulationen unterschieden. HyGLEAM benötigt also nur etwa die Hälfte an real durchgeführten Simulationen, um ein besseres Ergebnis zu erzielen als die nichthybride Variante. Dies führte zu einer Laufzeit von 2 Tagen und 15 Stunden auf 8 Suns unterschiedlicher Leistung (504 CPU-Stunden) bei GLEAM, während HyGLEAM mit 12 Tage und 16 Stunden (304 CPU-Stunden) auf einer Sun Ultra Sparc 10 auskam. Neuere Ergebnisse zu variierenden Drücken und E-Modulen sind in [16] im gleichen Band zu finden.

	Auslenkung		Spannung		Evaluierungen	
	$g$ [ $\mu\text{m}$ ]	[%]	[N/mm <sup>2</sup> ]	[%]	angef.	ausgef.
manuell	104	100	12	10		
GADO	238	229	16	13	2897	1130
HyGAD	251	241	17	14	1228	5935
O				2	47	

Tab.6: Optimierungsergebnisse

## 4.3 Schlussfolgerungen

Wie in den beiden vorigen Abschnitten dargestellt, gibt es keine problemunabhängige günstige Parametrierung für die Hybridisierung: Bei einer neuen Aufgabe ist weder klar, welches der beiden lokalen Verfahren günstiger ist, noch mit welcher Abbruchschranke der Rosenbrock betrieben werden soll oder ob und wenn ja mit welcher Verzögerung die lokale Suche zuzuschalten ist. Für die Populationsgröße kristallisiert sich immerhin heraus, dass kleine Werte genügen. Sicherheits halber sollte man einen Wert von zehn und bei wahrscheinlich schwierigen Problemen von 20 verwenden. Für die anderen Einstellungen sind selbstadaptive Lösungen anzustreben. So kann die Aufteilung zwischen Rosenbrock und Complex ausgehend von

einer Gleichgewichtung durch die Bewertung von jeweils erreichtem Erfolg und dazu notwendigem Aufwand angepasst werden.

Man kann davon ausgehen, dass bei der Nachkommensverbesserung anfänglich auf das genaue (lokale) Optimum verzichtet werden kann und die Präzision und damit der Aufwand erst im Laufe der Evolution ansteigen brauchen. Dies kann bei beiden Verfahren durch eine Begrenzung der Iterationen erreicht werden, die im Laufe der Evolution gelockert wird. Beim Rosenbrock kommt noch die Abbruchschranke als Einstellmöglichkeit hinzu. Beide Parameter können wieder durch das Verhältnis von erreichter Verbesserung bei investiertem Aufwand gesteuert werden. Als weitere Steuergröße kommt noch die Nischenbildung in Betracht, mit der der Fortschritt der Suche beurteilt werden kann.

Als Konsequenz aus den durchgeführten Untersuchungen kann gesagt werden, dass eine Hybridisierung auf der Grundlage der (verzögerten) direkten Integration bei hinreichend schwierigen Problemen eindeutig den geringsten Aufwand an Evaluationen verursacht. Dabei ist es problemabhängig, wie viel gegenüber GLEAM oder einem der lokalen Verfahren im Falle ihrer Konvergenz eingespart werden kann. Es wurden Einsparungen bis zum Faktor 104 (Fletcher & Powel) und 137 (Bigg (3)) beobachtet. Bei nicht zu schwierigen Problemen empfiehlt es sich, auch aus Gründen des geringen Aufwands erst den Complex auszuprobieren.

## 5 Webbasiertes HyGLEAM mit ORBIT

### 5.1 ORBIT

Abb. 9 gibt einen Überblick über Remote-Anwendungen mit dem in [8] ausführlicher beschriebenen ORBIT-System. Es besteht aus den in Abb. 10 dargestellten drei Komponenten Internet Web-Server, Data-Server und ORBIT-Engine, die im folgenden kurz beschrieben werden.

Der Internet Web-Server stellt die externe Schnittstelle zum ORBIT-System dar. Er verfügt über dynamische Webseiten, die je nach Anwendung automatisch generiert werden.

Hierzu zählen auch HTML-Formulare, die der Anwender zur Konfiguration der Modelle benutzt. Nach der Durchführung der jeweiligen Anwendung können auf den Seiten des Internet Web-Servers die Ergebnisseiten abgerufen werden.

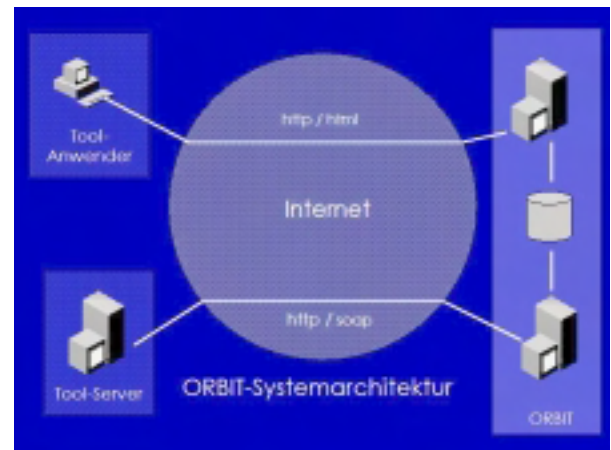


Abb.9: ORBIT-Systemarchitektur

Der Data-Server dient als Schnittstelle zwischen Internet Web-Server und ORBIT-Engine. In der Datenbank des Data-Servers werden Modell- und Ergebnisdaten sowie Laufzeitdaten von instanziierten Workflows vorgehalten. Der Data-Server verfügt über eine SOAP-Schnittstelle zur Generierung neuer Modellstrukturen und der Konfiguration von Tools bzw. Workflows.

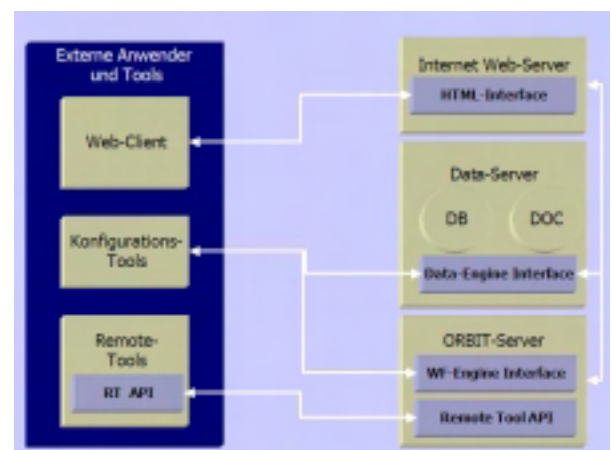


Abb.10: Komponenten des ORBIT-Systems

Die ORBIT-Engine stellt den eigentlichen Kern des Systems dar. Sie verwaltet Workflowdefinitionen und -instanzen und steuert den Fortschritt von laufenden Workflowinstanzen. Über die Remote-Tool-API werden externe Tools über SOAP in den Workflow-



prozess integriert und die Toolaufrufe gesteuert. Darüber hinaus verfügt die ORBIT-Engine über eine SOAP-basierte Service-Schnittstelle, mittels derer externe Tools Zugriff auf Konfigurationsdaten im ORBIT-System erhalten.

## 5.2 HyGLEAM in ORBIT

Zur Einbindung von HyGLEAM in das ORBIT-System wurde eine HyGLEAM-Engine ohne Benutzerinterface implementiert, deren Steuerung vollständig über die Aufrufparameter und Dateien erfolgt. Die HyGLEAM-Engine wird in einen eigenen in Java implementierten Tool-Agenten eingebettet, der die Kommunikation mit der ORBIT-Engine übernimmt. Abb. 11 zeigt das Einstiegsformular

des webbasierten Aufrufs eines Optimierungslaufs. Zuerst ist eine vorkonfigurierte Aufgabenstellung auszuwählen und dann das Optimierungsverfahren zu bestimmen. Je nach Wahl des Verfahrens gelangt man dann zu einem Unterformular, mit dessen Hilfe man die notwendige Parametrierung vornehmen kann, wobei für alle Eingabefelder Defaultwerte angeboten werden. In Abb. 12 ist das Formular für GLEAM selbst dargestellt während Abb. 13 die Parametrierung der Hybridisierungsart „Vorooptimierung“ zeigt. Während eines Laufs kann der Benutzer Zwischenergebnisse abfragen und gegebenenfalls den Lauf abbrechen. Am Ende eines Laufs erhält er die Ergebnisdaten.

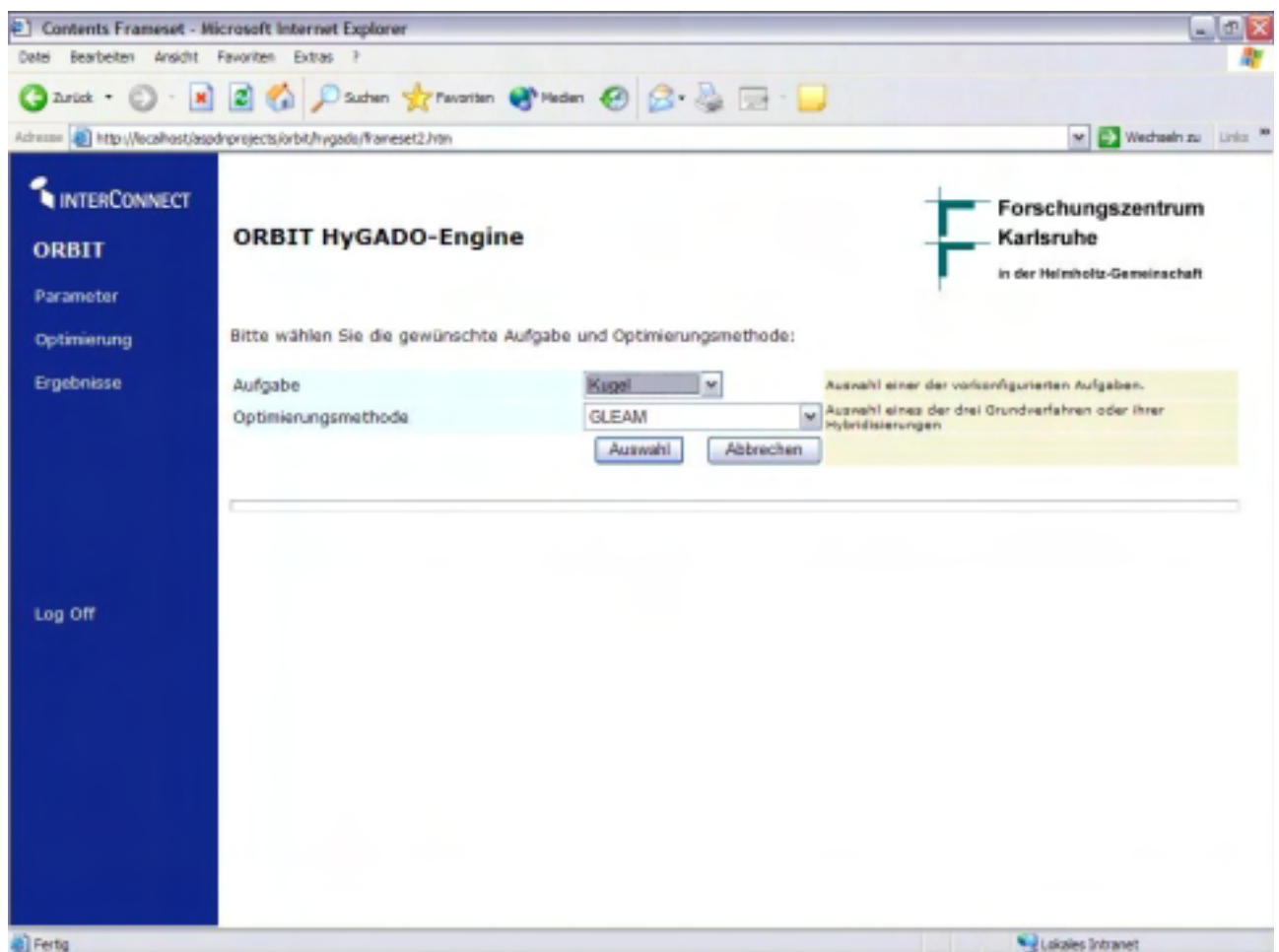


Abb.11: Einstiegsformular für einen Optimierungslauf

The screenshot shows a web browser window titled "Contents Frameset - Microsoft Internet Explorer" with the address bar displaying "http://localhost/ispd/projects/orbit/hygado/frameset1.htm". The page features a blue sidebar on the left with the "INTERCONNECT" logo and a menu containing "ORBIT", "Parameter", "Optimierung", "Ergebnisse", and "Log Off". The main content area is titled "ORBIT HyGADO-Engine" and includes the logo of the "Forschungszentrum Karlsruhe in der Helmholtz-Gemeinschaft". Below the title is the "Unterformular GLEAM" section, which contains a table of parameters for configuration:

Populationsgröße	128	5..1000
Dauer	1440	Dauer in Minuten, > 0
Zielnote	1000000	Zielqualität, 0..1000000
Generationslimit	5000	Max. Generationsanzahl, > 0
Generationen ohne lokale Verbesserung	500	Stagnationsindikator bei dem d. Lauf vorzeitig angebrochen wird, > 0
Generationen ohne lokale Akzeptanz	50	Stagnationsindikator bei dem d. Lauf vorzeitig angebrochen wird, > 0

At the bottom of the form are two buttons: "Auswahl" and "Abbrechen".

Abb.12: Unterformular zur Parametrierung eines GLEAM-Laufs

The screenshot shows the same web browser window as in Abb.12, but with the "Unterformular Voroptimierung" section active. This section contains a table of parameters for hybridization type configuration:

Populationsgröße	128	5..1000
Dauer	1440	Dauer in Minuten, > 0
Zielnote	1000000	Zielqualität, 0..1000000
Generationslimit	5000	Max. Generationsanzahl, > 0
Generationen ohne lokale Verbesserung	500	Stagnationsindikator bei dem d. Lauf vorzeitig angebrochen wird, > 0
Generationen ohne lokale Akzeptanz	50	Stagnationsindikator bei dem d. Lauf vorzeitig angebrochen wird, > 0
Lokales Suchverfahren	Rosenbrock	Verfahren
Rosenbrock-Präzision	niedrig	Rosenbrock-Abbruchkriterium
Voroptimierungsanteil	20	Prozentanteil voroptimisierender Individuen der Startpopulation

At the bottom of the form are two buttons: "Auswahl" and "Abbrechen".

Abb.13: Unterformular zur Parametrierung der Hybridisierungsart „Voroptimierung“

## 6 Zusammenfassung

Die wichtigsten Ergebnisse der umfangreichen Untersuchungen zum Verhalten globaler und lokaler Suchverfahren basierend auf dem Evolutionären Algorithmus GLEAM, dem Rosenbrock-Algorithmus und dem Complex-Verfahren unter Verwendung mathematischer Benchmarkfunktionen und realer Designprobleme sind, dass

1. es kein generell günstigstes Verfahren gibt,
2. bei einfacheren Aufgaben mit geringer Parameteranzahl die lokale Verfahren mit weniger Aufwand zum Ziel führen als das global arbeitende GLEAM,
3. ab einem gewissen Schwierigkeitsgrad die Hybridisierung in Form der direkten Integration zuverlässig und mit vergleichsweise geringem Aufwand zum Ziel führt,
4. hinsichtlich einer adaptiven Steuerung der Rechenzeitaufteilung zwischen globaler und lokaler Suche und der adaptiven Auswahl des geeigneten lokalen Verfahrens weiterer Forschungs- und Entwicklungsbedarf besteht, siehe Abschn. 4.3.

Mit der webbasierten Version von HyGLEAM wird eine Remote-Optimierung und damit eine wirtschaftlichere Nutzung der Simulationswerkzeuge durch einen Application Service Provider ermöglicht.

## Literatur

- [1] D. Peters, M. Gorges-Schleuter, W. Jakob, S. Parodat: *Modellbasierte Optimierung mit mathematischen Methoden – ein Vergleich von klassischen und evolutionären Verfahren*. 7. GMM-Workshop Methoden und Werkzeuge zum Entwurf von Mikrosystemen. Paderborn, 1999.
- [2] W. Jakob, B. Knorr, S. Parodat, D. Peters, A. Uhlig: *Optimierung von Mikrosystemen*. 8. GMM-Workshop Methoden und Werkzeuge zum Entwurf von Mikrosystemen. Berlin, 1999.
- [3] W. Jakob, D. Peters: *Designoptimierung mit HyGADO*. In: R. Laur, D. Peters (Hrsg.): OMID – Optimierung von Mikrosystemen für Diagnose- und Überwachungsanwendungen (Statusseminar), VDI/VDE, 2001.
- [4] W. Jakob: *HyGLEAM: Hybrid General purpose Evolutionary Algorithm and Method*. Proc. of 5<sup>th</sup> World Multiconference on Systemics, Cybernetics and Informatics (SCI'2001), 2001.
- [5] W. Jakob: *HyGLEAM - An Approach to Generally Applicable Hybridization of Evolutionary Algorithms*. Proc. of PPSN VII, LNCS 2439, Springer-Verlag, Berlin, 2002.
- [6] C. Blume: *GLEAM - A System for Simulated Intuitive Learning*. Proc. of PPSN I, LNCS 496, Springer-Verlag, 1991.
- [7] C. Blume, W. Jakob: *GLEAM - an Evolutionary Algorithm for Planning and Control Based on Evolution Strategy*. In: E. Cantú-Paz (Hrsg.): GECCO 2002, Vol. Late-Breaking Papers, L. Livermor National Laboratory, 2002.
- [8] A. Reiffer, P. Rohnacher: ORBIT – Internetbasierte Plattform für Remote-Anwendungen. In: H. Eggert, A. Reiffer (Hrsg.): OMID – Optimierung von Mikrosystemen für Diagnose- und Überwachungsanwendungen (Abschlußseminar des BMBF-Verbundprojekts OMID), FZK, 2002.
- [9] H.-P. Schwefel: *Evolution and Optimum Seeking*. John Wiley & Sons, Chichester, 1995.
- [10] C. Blume, W. Jakob: *Cutting Down Production Costs by a New Optimization Method*. Proc. of the Japan - U.S.A. Symposium on Flexible Automation, ASME, 1994
- [11] H. H. Rosenbrock: *An Automatic Method for Finding the Greatest or Least Value of a Function*, Comp. J. 3, 1960.
- [12] M. J. Box: *A New Method of Constrained Optimization and a Comparison with other Methods*, Comp. J. 8, 1965.
- [13] I. Sieber, H. Eggert, H. Guth, W. Jakob, K.-P. Scherer, P. Ziegler: *Design Optimization Considering Tolerance Effects of Microoptical Benches*. MicroSystem Technologies 98, VDE-Verlag, 1998.
- [14] A. Quinte: *Modellierung und Optimierung der Aktorplatte*. In: R. Laur, D. Peters (Hrsg.): OMID – Optimierung von Mikrosystemen für Diagnose- und Überwachungsanwendungen (Statusseminar), VDI/VDE, 2001.
- [15] W. Jakob, A. Quinte, K.-P. Scherer, H. Eggert: *Optimization of a Micro Fluidic Component using a Parallel Evolutionary Algorithm and Simulation Based on Discrete Element Methods*. S. Hernández, C.A. Brebbia (Eds.): Computer Aided Optimum Design of Structures VII, WIT Press, Southampton, 2001.
- [16] A. Quinte: *Modellierung und Optimierung der Aktorplatte*. In: H. Eggert, A. Reiffer (Hrsg.): OMID – Optimierung von Mikrosystemen für Diagnose- und Überwachungsanwendungen (Abschlußseminar des BMBF-Verbundprojekts OMID), FZK, 2002.

## ~~DESIGNOPTIMIERUNG VON MIKROSYSTEMEN MIT MODOS~~

D. Peters\*, H. Bolte\*, O. Nüssen\*, W. Jakob\*\*, A. Quinte\*\*, A. Schneider\*\*\*, R. Laur\*

\*Universität Bremen, ITEM, Postfach 33 04 40, 28334 Bremen

\*\*Forschungszentrum Karlsruhe, Institut für Angewandte Informatik

\*\*\* Fraunhofer Institut für Integrierte Schaltungen, Dresden

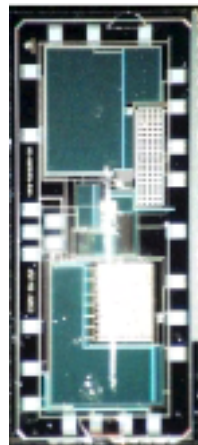
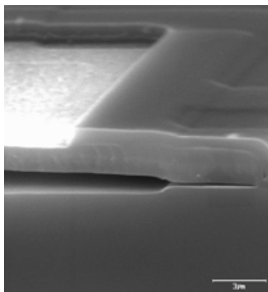
Email: peters@item.uni-bremen.de

### Kurzfassung

Die Designoptimierung von Mikrosystemen und deren Komponenten ist eine vielfältige Aufgabe, die nicht nur flexible Werkzeugunterstützung erfordert, sondern insbesondere auch die Entwicklung leistungsstarker Optimierungsstrategien erforderlich macht. In diesem Beitrag wird das Optimierungswerkzeug MODOS mit den im Rahmen dieses Forschungsvorhabens entwickelten angepassten Optimierungsstrategien anhand verschiedener Anwendungen vorgestellt.

### 1 Einleitung

Die Verfahren der Mikrostrukturierung werden in den verschiedensten technischen Bereichen zur Fertigung zuverlässiger und kostengünstiger Produkte eingesetzt. Aufgrund des Zusammenwirkens verschiedener physikalischer Domänen sowie aufgrund des wachsenden Integrationsgrades von Mikrosystemen nimmt die bei der Entwicklung zu beherrschende Systemkomplexität stark zu. Designer sind daher mehr und mehr auf die Unterstützung von CAD-Werkzeugen angewiesen.



**Abb. 1:**  
Mikrosystemtechnische  
Komponenten

Abbildung 1 zeigt zwei typische mikrosystem-

technische Komponenten: links, im Querschnitt, einen mikromechanischen Drucksensor und rechts daneben einen für medizintechnische Anwendungen entwickelten Telemetriechip.

Beide Komponenten sind Bestandteile medizintechnischer Produkte zur Messung des Drucks in verschiedenen Körperbereichen: z.B. Blutdruck, Druck im Inneren des Auges, Druck im Kopf. Bei jeder dieser Anwendungen ist der Sensor in den Körper des Menschen zu implantieren. Die Messsignale sind mit einer Signalvorverarbeitung so aufzubereiten, dass sie von der Telemetrieinheit aus dem Körperinneren in Form eines digitalen Datenwortes drahtlos nach außen gesendet werden können. Gleichzeitig übernimmt die Telemetrieinheit die Energieversorgung des intelligenten Sensorchips.

Neben einer stabilen Energieversorgung müssen die Messsignale des Sensors möglichst fehlerfrei übermittelt werden. Die geforderte Auflösung des Sensors ist zu realisieren und der maximal zulässige Leistungsverbrauch darf nicht überschritten werden. Allein diese Teilaspekte des Entwurfs des hier prototypisch vorgestellten Mikrosystems verdeutlichen die Entwurfskomplexität solcher Systeme. Üblicherweise wird ein modularer Ansatz gewählt, der das System in technisch weitgehend abgeschlossene Blöcke unterteilt. Diese werden unter Zuhilfenahme von Modellierung und Simulation zunächst „virtuell“ entwickelt. Soweit als möglich wird die Funktionalität des Gesamtsystems entsprechend der Spezifikation, basierend auf Simulationen, sichergestellt. Anschließend werden erste Prototypen gefertigt.

Doch auch dieser modulare Ansatz wird mit wachsender Integrationsdichte zunehmend komplizierter und schwieriger zu handhaben. Designer benötigen neben Modellierung und Simulation weitere Entwurfsunterstützungen,

mit denen aus der Vielzahl existierender Designvarianten diejenigen ausgewählt werden können, die die Spezifikation bestmöglich erfüllen. Derartige Entwurfsunterstützungen können durch Verfahren der modellbasierten Designoptimierung geliefert werden.

Ein Vorteil der modellbasierten gegenüber beispielsweise der auf symbolischen Berechnungen basierenden Designoptimierung besteht darin, dass sie sich nahtlos in den heute bereits etablierten Designprozess für Mikrosysteme einbinden lässt. Der Entwickler benötigt lediglich ein Werkzeug zur Optimierung, das mit beliebigen Simulatoren gekoppelt werden kann oder bereits gekoppelt ist. Problematisch dabei ist lediglich, dass die einzelnen Simulationen, die zur Bewertung eines Satzes von Designparametern benötigt werden, sehr zeitintensiv sein können. Des Weiteren liegt die meist nichtlineare Zielfunktion nicht in analytischer Form vor. Es sind also neben einem flexiblen Optimierungswerkzeug leistungsfähige Optimierungsalgorithmen zur Verfügung zu stellen, die möglichst sicher die beste Designalternative ermitteln können und gleichzeitig wenig Simulationen für die Suche benötigen.

In diesem Beitrag werden Ergebnisse der Arbeiten aus dem Forschungsvorhaben „Modellbasierte Designoptimierung von Mikrosystemen“ zur Weiterentwicklung eines Werkzeuges für die modellbasierte, rechnergestützte Designoptimierung (MODOS) präsentiert.

Zielsetzungen dieses Forschungsvorhabens hinsichtlich der Werkzeugentwicklung waren:

- die Entwicklung von an die Problemstellungen der Mikrosystemtechnik angepassten Optimierungsalgorithmen;
- die Entwicklung von hybriden Optimierungsalgorithmen;
- die Kopplung des am ITEM entwickelten Optimierungswerkzeugs MODOS an diverse Simulatoren;
- die Einbindung von MODOS in das verteilt arbeitende, web-basierte System MOSCITO [1].

In Abschnitt 2 wird exemplarisch die Anwendung des modellbasierten Designoptimierungssystems MODOS auf die Optimierung von Teilaspekten bei der Entwicklung des anfangs vorgestellten Mikrosystems dargestellt. Im Abschnitt 3 werden drei an die Erfordernisse der Mikrosystemtechnik angepasste Optimierungsverfahren vorgestellt. Abschnitt 4 und 5 stellen die Ergebnisse zur Optimierung zweier Anwendungen des Verbundprojektes OMID vor, bevor im Abschnitt 6 die Zusammenfassung folgt.

## 2 Exemplarische Anwendung von MODOS in der Mikrosystemtechnik

Das Modellbasierte Designoptimierungssystem MODOS (Abbildung 2) wurde speziell für die Belange des Designs von Mikrosystemen entwickelt. Im Folgenden wird das Konzept von MODOS wie auch eine seiner möglichen Anwendungen im Entwurfsprozess von Mikrosystemen erläutert.

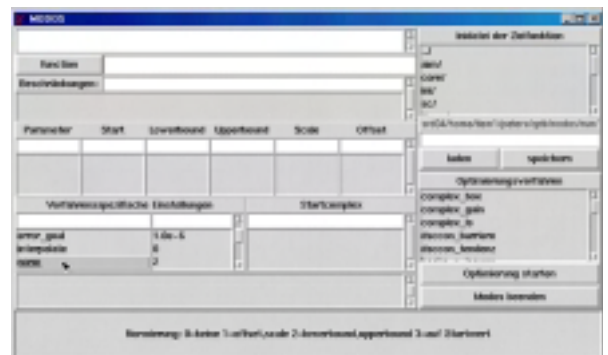


Abb. 2: MODOS - GUI

### 2.1 Modellbasierte Designoptimierung von Mikrosystemen

Die bei der Entwicklung von Mikrosystemen auftretenden Fragestellungen sind vielfältig. Dazu zählen unter anderem Modellparametrisierungen (insbesondere von Verhaltens- oder Makromodellen), Machbarkeitsstudien, Komponenten- und Systemoptimierungen hinsichtlich der vorgegebenen Spezifikation. Es ist also nicht sinnvoll, ein an feste Charakteristika einer Designaufgabe angepasstes Optimierungsverfahren zu entwickeln. MODOS stellt daher eine Auswahl verschiedener ein-

und mehrdimensionaler, beschränkt sowie unbeschränkt arbeitender Algorithmen zur Verfügung. Die Formulierung der Bewertungsfunktion (Zielfunktion) bleibt dem Anwender überlassen. Dies kann von Nachteil sein, wenn nur wenig Vorkenntnisse im Bereich der Optimierung vorhanden sind, birgt jedoch den Vorteil in sich, dass MODOS auf beliebige Problemstellungen angewendet werden kann.

Des Weiteren wurde MODOS mit diversen Simulatoren gekoppelt, da bei der Entwicklung von Mikrosystemen sowohl feld- als auch netzlistenbasierte und Verhaltenssimulatoren eingesetzt werden. Prinzipiell kann jeder beliebige Simulator mit MODOS betrieben werden. Realisiert wurden bislang folgende Kopplungen: PSPICE, HSPICE, TITAN, SABER, MATHEMATICA. Da MODOS als sogenannte TCL/Tk-Applikation realisiert wurde, kann es mit anderen TCL/Tk-Applikationen wie beispielsweise SABER direkt über Techniken der Interprozesskommunikation verbunden werden [2].

Um dem Anwender die Nutzung von MODOS zu erleichtern, wurde eine interaktive Hilfefunktion in die graphische Oberfläche eingebaut.

Die Anwendung von MODOS zur Optimierung von Mikrosystemen wird anhand des in der Einleitung bereits diskutierten Systems exemplarisch erläutert. Es handelt sich dabei um eine medizintechnische Anwendung von der im Folgenden der Drucksensor mit Signalverarbeitung näher betrachtet wird.

## 2.2 Modellevaluation

Bevor Verhaltensmodelle mikrosystemtechnischer Komponenten zur Designoptimierung eingesetzt werden können, ist sicherzustellen, dass alle wesentlichen Effekte in der abstrakteren Modellvariante enthalten sind. Zu diesem Zweck werden Verhaltensmodelle üblicherweise manuell mit Werten parametrisiert, die aus FEM-Simulationen oder Messungen bekannt sind. Ein visueller Vergleich zwischen gemessenen bzw. FEM-Kennlinien und den entsprechenden Kennlinien der Verhaltenssimulation liefert den Nachweis über die

Gültigkeit des Modells. Aufgrund der Vielzahl von Parametern, den zwischen ihnen auftretenden Wechselwirkungen und den über einen Wafer verteilten Parameterschwankungen ist diese Vorgehensweise sehr aufwendig.

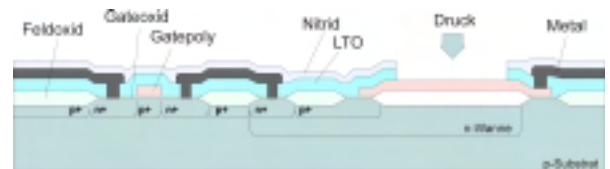


Abb. 3: Querschnitt des Drucksensors

Effektiver und auch zuverlässiger kann die Entwicklung eines Modells gestaltet werden, wenn Optimierungswerkzeuge wie MODOS unterstützend verwendet werden. Ausgehend von Messungen oder FEM-Simulationen wird die Kennlinie des Verhaltensmodells an die vorgegebene Kennlinie über Parametervariationen angepasst. Die Nominalwerte der Design- bzw. Prozessgrößen werden in physikalisch sinnvollen Grenzen vorgegeben, soweit sie bekannt sind. Kann das Modell die vorgegebene Kennlinie mit Abweichungen, die kleiner als die Messgenauigkeit sein sollten, nachbilden, und sind die Parameterwerte physikalisch sinnvoll gewählt worden, so bildet das Verhaltensmodell den Zusammenhang zwischen der gemessenen Größe und den Designparametern korrekt nach.

Anhand dieser Vorgehensweise wurde das Verhaltensmodell des Drucksensors mit MODOS evaluiert [3]. Verwendet wurden hierzu der Complex-Algorithmus [4], der ComplInt-Algorithmus [5], das Verfahren nach Rosenbrock [6] sowie der RCT-Algorithmus [7]. Um möglichst große Flexibilität zu bieten, wurde das Sensormodell in Form einer C-Funktion realisiert, die leicht in MODOS und in SABER eingebunden werden kann.

Ergebnis der Modellevaluation ist, dass der Zusammenhang zwischen der Sensorkapazität, den geometrischen Parametern sowie einigen Prozessparametern korrekt modelliert wurde und das Modell somit für Komponenten- wie auch Systemoptimierungen einge-

setzt werden kann.

### 2.3 Ableitung von Designvorgaben

Das Drucksensorsystem besteht aus einem kapazitiv arbeitenden Absolutdrucksensor (Abbildung 3) und einem SC-Oszillator zur Signalaufbereitung (Abbildung 4).

Die parasitären Kapazitäten  $C_{p1}$  am Eingang und  $C_{p2}$  am Ausgang des Sensors, die Verstärkung des Operationsverstärkers  $A_D$  sowie die Dimensionierung der Referenzkapazität  $C_R$  und der Feedbackkapazität  $C_F$  haben einen nicht zu vernachlässigenden Einfluss auf die Güte des Ausgangssignals. Sie führen zu einem Fehler  $\Delta u_a$ , der eine Veränderung des Pulsweitenverhältnisses und damit eine Verschlechterung der Auflösung nach sich zieht.

$$\Delta u_a = \Delta u_{a,ideal} - \frac{C_{p2}}{C_F} \cdot \frac{u_a}{A_D}$$

Um die vom Anwender geforderte minimale Auflösung realisieren zu können, ist der Einfluss der parasitären Kapazität  $C_{p2}$  zu minimieren. Dies kann erreicht werden, indem die Feedbackkapazität  $C_F$  bzw. die Verstärkung des Operationsverstärkers  $A_D$  möglichst groß gewählt werden. Zu beachten ist dabei, dass Kapazitäten sehr viel Chipfläche verbrauchen und deswegen nicht beliebig groß gewählt werden können. Da auch die Verstärkung des Operationsverstärkers nicht beliebig groß gewählt werden kann, ist genau das Design gesucht, das für bestimmte maximal zulässige parasitäre Kapazitäten die vom Anwender geforderte Auflösung erzielt.

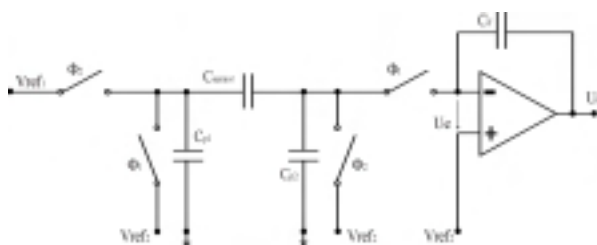


Abb. 4: SC-Oszillator

Beim manuell optimierten Nominaldesign konnte die geforderte Auflösung nicht erreicht werden. Die mit MODOS und SABER durchgeführten Designoptimierungen erzielen dage-

gen ein Design, dass bei einer Vergrößerung der Kantenlänge um ca. 15% und einer Verdoppelung der Verstärkung des Operationsverstärkers die Spezifikation erfüllt. Im nächsten Schritt wäre ein entsprechender Operationsverstärker zu entwickeln und das Sensordesign zu überarbeiten.

## 3 Entwicklung angepaßter Optimierungsverfahren

Die besondere Problematik der modellbasierten Designoptimierung besteht darin, daß Simulationsergebnisse zur Bewertung einer Designvariante benötigt werden, die Zielfunktionen also nicht in analytischer Form vorliegen. Da Simulationen in der Mikrosystemtechnik sehr zeitaufwendig sein können, sind für die modellbasierte Designoptimierung effiziente mathematische Suchverfahren zur Verfügung zu stellen. Sie sollten möglichst global konvergieren und möglichst wenige Iterationsschritte zur Ermittlung des Ergebnisses benötigen.

Neben der im Abschnitt 3.3 vorgestellten hybriden Strategie, die genetische Algorithmen mit direkten Suchverfahren kombiniert, wurden im Rahmen dieses Forschungsvorhabens zwei leistungsfähige stand-alone Verfahren implementiert: ein für beschränkte Problemstellungen entwickeltes Quasi-Newton-Verfahren sowie ein neuartiges Polyederverfahren. Quasi-Newton-Verfahren zeichnen sich in der Regel durch schnelle lokale Konvergenz aus, während das Polyederverfahren so konstruiert wurde, dass es lokal und global ausgerichtete Elemente bei der Suche miteinander kombiniert.

### 3.1 Quasi-Newton Verfahren

Quasi-Newton Verfahren zeichnen sich durch ihre schnelle Konvergenz aus. Sie weisen jedoch den Nachteil auf, dass sie in der Regel in das dem Startpunkt am nächsten gelegene lokale Optimum konvergieren. Dieses Verhalten resultiert daraus, dass zur Ermittlung der Suchrichtung die ersten sowie zweiten Ableitungen der Zielfunktion herangezogen werden.

Quasi-Newton Verfahren bilden im aktuellen Designpunkt ein quadratisches Modell der Zielfunktion mit Hilfe eines Taylorreihenansatzes zweiter Ordnung:

$$f(\mathbf{x}) \approx f(\mathbf{x}_k) + \mathbf{f}'(\mathbf{x}_k)^T (\mathbf{x} - \mathbf{x}_k) + \frac{1}{2} (\mathbf{x} - \mathbf{x}_k)^T \mathbf{A} (\mathbf{x} - \mathbf{x}_k)$$

Die Suchrichtung für den nächsten Schritt wird über das Optimum dieses lokal gültigen Modells ermittelt; die Schrittlänge in diese Richtung bestimmt man unter Verwendung eines beliebigen eindimensionalen Liniensuchverfahrens.

Problematisch wird die Verwendung dieser Verfahren dann, wenn Beschränkungen bei der Optimierung zu berücksichtigen sind. Dies ist bei fast jeder technischen Aufgabenstellung der Fall. Deshalb wurden im Rahmen dieses Forschungsvorhabens verschiedene Ansätze zur Berücksichtigung von Nebenbedingungen untersucht. Diese Untersuchungen basieren alle auf dem sogenannten Straffunktionsansatz. Kritisch gestaltet sich insbesondere die Implementation eines eindimensionalen beschränkten Suchverfahrens zur Ermittlung der optimalen Schrittlänge. Hier werden derzeit zwei Strategien untersucht: das Liniensuchverfahren nach Moré und Thuente [8] sowie ein Einschachtelungsansatz.

**Tab. 1: Testfunktion Beale**

	Boxing	Boxing	Thuente	Thuente
Startpkt	$\Delta$ in %	Sims	$\Delta$ in %	Sims
1	0,2	2793	3,6	67
2	1,2	8969	2,5	59
3	0,5	10113	1,4	59
4	0,5	9333	-	-
5	0,8	10022	-	-

**Tab. 2: Testfunktion Rosenbrock**

	Boxing	Boxing	Thuente	Thuente
Startpkt	$\Delta$ in %	Sims	$\Delta$ in %	Sims
1	0,2	2917	0,2	47
2	0,2	3104	0,8	81
3	0,2	3086	1,9	72
4	0,2	2948	4,4	32
5	0,2	2783	-	-

Bei den derzeit durchgeführten Evaluationen anhand von mathematischen Testproblemen

(Tabelle 1 bis 3) zeichnet sich der Trend ab, dass das Verfahren mit dem Thuente-Ansatz schneller, dafür aber ungenauere Ergebnisse liefert, während die das Einschachtelungsverfahren anwendende Strategie relativ lange für die Konvergenz benötigt, dafür aber genauere Ergebnisse liefert. Beide Ansätze weisen das Problem der Wahl des Strafparameters auf.

**Tab. 3: Testfunktion Schwefel 2.46**

	Boxing	Boxing	Thuente	Thuente
Startpkt	$\Delta$ in %	Sims	$\Delta$ in %	Sims
1	0,1	1717	0,1	532
2	1,0	1932	0,1	490
3	1,4	571	0,9	1093
4	0,7	808	0,3	1081
5	1,2	4180	0,1	216

### 3.2 Polyederstrategie

Polyederverfahren wie der Complex-Algorithmus nach Box [4] oder der Simplex-Algorithmus nach Nelder und Mead [9] zeichnen sich dadurch aus, dass sie unter Anwendung verschiedener Operationen (Spiegelung, Kontraktion, Reflektion) einen m-dimensionalen Polyeder durch den n-dimensionalen Raum in Richtung eines Optimums bewegen ( $m > n$ ). Die im Rahmen dieses Forschungsvorhabens entwickelte Polyederstrategie stellt ein Mischverfahren dar, das Elemente des Simplex und des Complex-Algorithmus miteinander kombiniert. Es ist daher sowohl zur Optimierung unbeschränkter als auch zur Optimierung beschränkter Problemstellungen geeignet.

#### 3.2.1 Theoretische Darstellung des Verfahrens

Neben den für Polyederverfahren üblichen Operationen Spiegelung, Reflektion und Kontraktion führt das hier vorgestellte Verfahren nach erfolgreicher Reflektion eine Überreflektion durch. Auch kann bei nicht erfolgreicher Kontraktion eine sogenannte multiple Kontraktion durchgeführt werden, bei der sich der gesamte Polyeder um den bislang besten Punkt herum zusammenzieht. Je nach Art der Optimierungsaufgabe führt das Verfahren mit multipler Kontraktion besser oder weniger gut zum gewünschten



Ergebnis. Daher hat der Anwender die Möglichkeit, dieses Feature gezielt auszuwählen. Außerdem wird der Reflexionsfaktor bei dem hier vorgestellten Verfahren automatisch an die Eigenschaften der zu bearbeitenden Optimierungsaufgabe angepaßt, bleibt während einer Optimierung jedoch konstant.

Neben diesen Charakteristika zeichnet sich das Verfahren im wesentlichen durch folgende Eigenschaften aus:

- üblicherweise arbeitet das Verfahren mit einem ausgehend vom durch den Anwender vorgegebenen Startpunkt zufällig aufgebauten Polyeder;
- kombiniertes Abbruchkriterium der beiden oben aufgeführten Verfahren;
- automatischer Teilrestart nach zu großer Anzahl von erfolglosen Kontraktionen;
- lokale oder globale Ausrichtung des Restarts, je nach Vorgabe des Anwenders.

Erste Evaluationen, basierend auf mathematischen Testfunktionen (Tabelle 4 und 5), führen zu vielversprechenden Ergebnissen, insbesondere bei den beschränkten Testfunktionen.

Bei den Benchmarkfunktionen ohne Beschränkungen ist jeweils die Anzahl der Simulationen des besten Laufs dargestellt<sup>1</sup>, da für statistische Auswertungen nicht genügend Referenzdaten zur Verfügung standen. In Tabelle 2 sind dagegen jeweils die Anzahl der Simulationen dargestellt, die die Algorithmen zur Erzielung einer 99%igen Konvergenzsicherheit benötigen würden.

Bei den unbeschränkten Testproblemen ist das Polyederverfahren von der Güte her vergleichbar mit dem speziell für diese Aufgabenstellungen entwickelten Verfahren nach Nelder und Mead; bei den beschränkten Testbeispielen schneidet das Polyederverfahren weitaus besser ab als die Verfahren nach Box und Guin.

**Tab. 4: Evaluierung Polyederverfahren (Benchmarks ohne Beschränkungen)<sup>2</sup>**

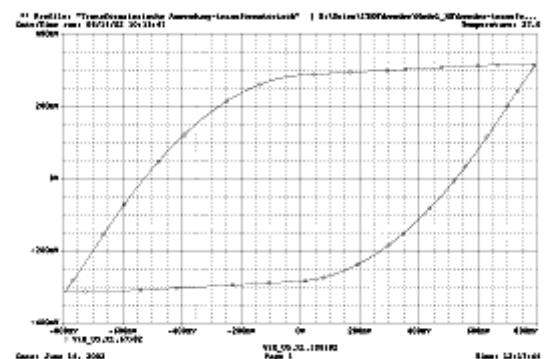
	T1	T2	T3	T4	T5	T6	T7
<b>Pol</b>	225	165	390	360	100	30	185
<b>Box</b>	300	400	710	n.e.	175	n.e.	295
<b>Sub</b>	290	250	n.e.	370	180	35	270
<b>Sim</b>	210	220	890	n.e.	80	n.e.	140
<b>Nel</b>	200	190	770	330	75	40	160

**Tab. 5: Evaluierung Polyederverfahren (Benchmarks mit Beschränkungen)**

	Smith	Wolfe	Bracken
<b>Pol</b>	270	545	59
<b>Box</b>	32676	7041	448
<b>Guin</b>	35896	9169	466

### 3.2.2 Modelloptimierung magnetischer Komponenten

Das Polyederverfahren wurde aufgrund seiner zuverlässigen Konvergenzeigenschaften zur Parametrisierung eines Kompaktmodells für hartmagnetische Materialien [14] eingesetzt.



**Abb. 5: minor loop**

Dabei waren die Modellparameter jeweils so zu wählen, dass das Modell die zu bestimmten Materialien vorgegebenen minor loops (Kleinsignalverhalten, Abb. 5) wie auch Hysteresekurven (Großsignalverhalten, Abb. 6) nachbilden kann. Die für Klein- und Großsignalverhalten ermittelten Parametersätze führten zu sehr guten Übereinstimmungen mit den Messwerten. Die Ergebnis-

<sup>1</sup> n.e. bedeutet "Optimum nicht erreicht"

<sup>2</sup> Pol: Polyederverfahren; Box: Complex Algorithmus nach Box; Sub: Simplexverfahren; Sim: Simplexverfahren; Nel: Simplex nach Nelder und Mead; Guin: Complex Algorithmus nach Box

kurven sind in Abbildung 5 und Abbildung 6 dargestellt.

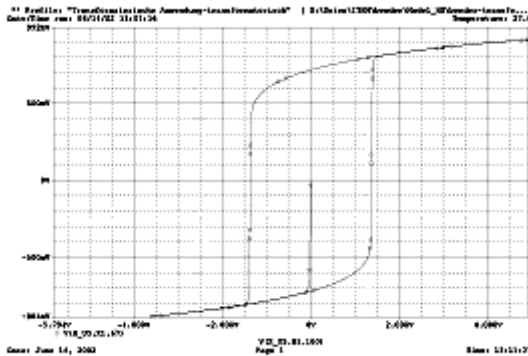


Abb. 6: Hysteresekurve

### 3.3 Verfahrensvergleiche mit Hilfe der MOSCITO-Schnittstelle

Im Laufe des Verbundprojektes OMID wurde MODOS in des System zur Verteilten Simulation und Optimierung, MOSCITO, eingebunden. Hierzu waren die entsprechenden Schnittstellen und Agentenbeschreibungen bereit zu stellen. Die Arbeiten wurden in enger Kooperation mit der Fraunhofer Gesellschaft Dresden durchgeführt.

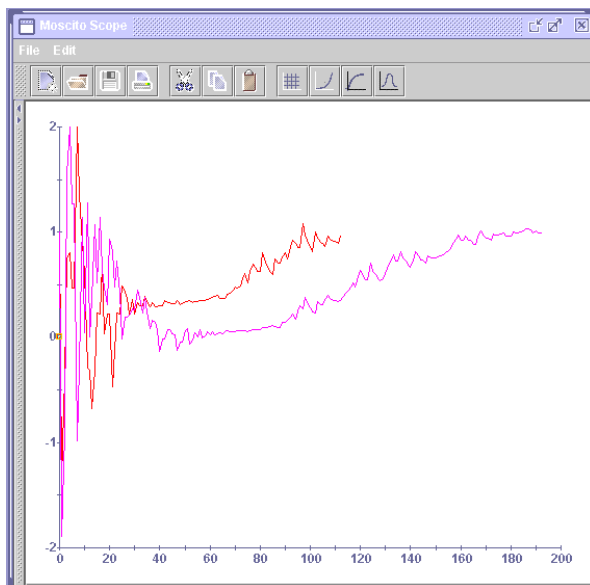


Abb. 7: MODOS-Optimierungen mit MOSCITO

In Abbildung 7 ist die Veränderung eines Designparameters während zwei verschiedener Optimierungsläufe dargestellt. Beide Läufe wurden mit dem in Abschnitt 3.2 vorgestellten Polyederverfahren durchgeführt

und unterscheiden sich lediglich in der Wahl des Abbruchkriteriums. Der erste Lauf (rot) arbeitete mit einer Fehlervorgabe von  $1.0e-3$  und erzielte im Ergebnis einen Zielfunktionswert von  $1.4823e-3$ <sup>3</sup>. Der zweite Optimierungslauf (rosa) arbeitete mit einer Fehlervorgabe von  $1.0e-4$  und erzielte im Ergebnis einen Zielfunktionswert von  $1.8054e-5$ .

An diesen beiden Ergebnissen sind die zentralen Eigenschaften des Polyederverfahrens deutlich zu erkennen:

- einstellbare Konvergenzgeschwindigkeit die mit der Genauigkeit des Ergebnisses korreliert ist;
- im hier gewählten lokalen Modus zieht sich der Polyeder relativ schnell zusammen, was insgesamt betrachtet zu einer verhältnismässig schnellen Konvergenz führt;
- durch die Überreflektionen und Neustart hat der Polyeder aber auch wieder die Möglichkeit zu expandieren (ist jeweils gegen Ende der Suche zu beobachten). Dieses Verhalten verhindert eine vorzeitige Konvergenz.

### 3.4 Hybride Strategien

In Kooperation mit dem Forschungszentrum Karlsruhe wurde während dieses Verbundvorhabens an der Entwicklung eines hybriden Optimierungsverfahrens gearbeitet. Grundlage dafür stellten zum einen das am Forschungszentrum entwickelte GADO-Verfahren [10] sowie die am ITEM implementierten direkten Suchverfahren nach Rosenbrock [6] und Box [4] dar. Das GADO-Verfahren verbindet die positiven Eigenschaften klassischer genetischer Algorithmen geschickt mit den Vorzügen der evolutionären Algorithmen.

#### 3.4.1 Eingesetzte Umschaltkriterien

Den kritischsten Punkt bei der Entwicklung hybrider Strategien stellt die Formulierung erfolgreicher Strategien zur Kombination der Optimierungsalgorithmen dar. Bei den hier

<sup>3</sup> Das tatsächliche Optimum liegt bei einem Zielfunktionswert von 0.0.

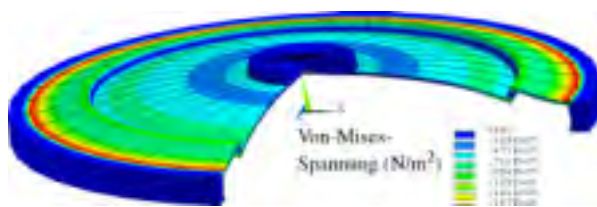
durchgeführten Arbeiten wurden folgende Kopplungsstrategien untersucht:

- Vorinitialisierte Startpopulation  
10% - 100% der Individuen der Startpopulation werden voroptimiert
- Nachkommensverbesserung  
lokales Suchverfahren wird zur Optimierung des besten Nachkommens eingesetzt
- Nachoptimierung  
Finetuning der Evolutionsergebnisse mit lokalem Suchverfahren

Die theoretischen Hintergründe dieser Strategie werden in [11] detailliert erläutert.

### 3.4.2 Ergebnisse zur Strukturoptimierung der Aktorplatte

Aufgrund der Möglichkeit, Problemstellungen mit variierender Anzahl an Designparametern behandeln zu können, eignet sich das hybride Verfahren insbesondere zur Strukturoptimierung der Aktorplatte (Abbildung 7).

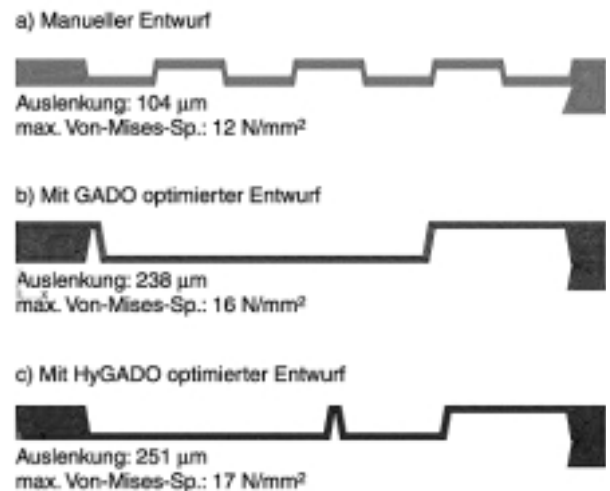


**Abb. 7: Aktorplatte**

Die Aktorplatte kann beispielsweise in fluidischen Systemen als Ventil eingesetzt oder zum Aufbau eines Braillearrays verwendet werden. Auf Realisierung und Funktion der Aktorplatte geht der Beitrag der Bartels mikrotechnik AG genauer ein [12]. Ziele der Optimierung dieses mikrotechnischen Elements war es, neben einer möglichst großen Auslenkung nur kleine Verspannungen in der Platte zu haben.

In Abbildung 8 sind die Ergebnisse zur Strukturoptimierung der Aktorplatte dargestellt. Mit dem manuell optimierten Entwurf wurden weitaus schlechtere Ergebnisse erzielt als mit den automatisierten Ansätzen zur Optimierung des Systems. Überraschend ist jedoch insbesondere, dass das beste Ergebnis

von der hybriden Strategie geliefert wurde. Auch liegt die Auslegung dieser Designvariante nicht unbedingt auf der Hand. Manuelle Designoptimierungen hätten wahrscheinlich kaum zu solch einem Ergebnis geführt.



**Abb. 8: Ergebnisse zur Strukturoptimierung der Aktorplatte**

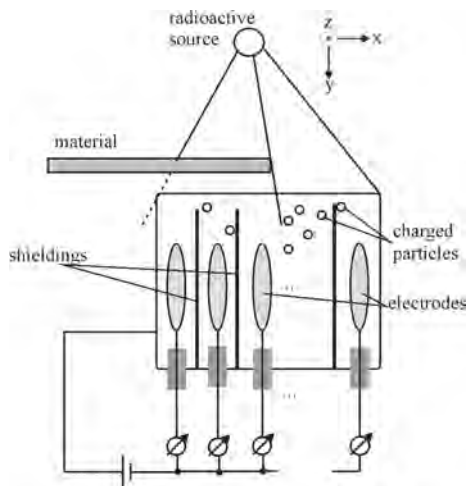
## 4 Multikriterielle Optimierung einer Mehrfachionisationskammer

Die im Verbundprojekt OMID im Rahmen einer vorwettbewerblichen Entwicklung untersuchte Mehrfachionisationskammer (Abbildung 9), kurz MFIK, der Firma VacuTec Meßtechnik GmbH, erweitert die Messwerte klassischer Ionisationskammern um eine eindimensionale Ortsauflösung. Mit der MFIK werden Materialbahnen während der Produktion bezüglich ihrer Kantenlage oder Flächenmasse überwacht.

Bei der Optimierung sind drei Kriterien zu berücksichtigen, die teils gegenläufig sind:

- der Strom in den Zellen soll absolut maximiert werden, um einen guten Rauschabstand zu erhalten
- die Steigung der Stromflanke soll maximiert werden, um eine exakte Kantendetektion zu ermöglichen
- der Durchstrahleffekt soll minimiert werden, um die Signalverfälschung zu verringern

Für die Optimierung sind der Kammerdruck, die Abstände von der Quelle zur Materialbahn und von der Materialbahn zur Kammer, der Elektrodenabstand und die Abschirmdicke variabel. Insgesamt sind also fünf Parameter für die Optimierung in den konstruktionsbedingten Grenzen frei, und es existieren drei Kenngrößen, anhand derer eine Aussage über die Optimalität getroffen werden soll.



**Abbildung 9: Skizze der Mehrfachionisationskammer**

Ziel der Designoptimierung für die MFIK war neben der eigentlichen Systemoptimierung die Untersuchung des Einflusses der Bewertungsfunktion.

Für die MFIK wurde am ITEM ein physikalisch parametrisierbares C-Modell entwickelt, das einfach mit MODOS gekoppelt und in Verhaltenssimulatoren wie SABER eingebunden werden kann. Die diversen Zielfunktionsformulierungen wurden ebenfalls als C-Funktionen zu MODOS hinzugefügt. Untersucht wurden verschiedene, statisch gewichtete Ansätze sowie ein Ansatz zur dynamischen Wichtung der Einzelkriterien. Die Durchführung der Optimierungen erfolgte mit dem Complex-Algorithmus nach Box [4] bzw. mit dem RCT-Algorithmus [7].

Im Vorfeld der eigentlichen Designoptimierungen wurden die Maxima sowie die Minima der Einzelkriterien ermittelt. Erst dadurch werden gezielte Skalierungen und Gewichtungen der Einzelkriterien möglich. Unabhängig

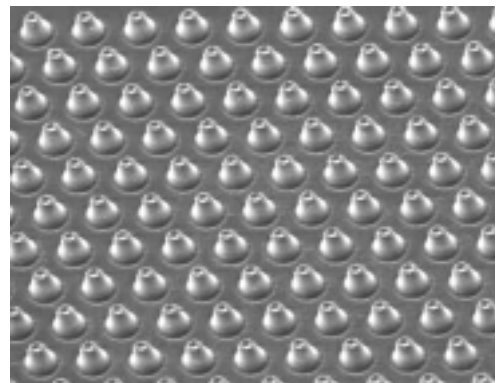
von der gewählten Zielfunktionsart ergaben die statischen Gewichtungen immer eines der folgenden qualitativen Ergebnisse:

- Verbesserung hinsichtlich des Absolutwertes des Stromes und hinsichtlich der Stromflankensteigung bei gleichzeitiger Verschlechterung des Durchstrahleffektes
- Verbesserung des Durchstrahleffektes bei gleichzeitiger Verschlechterung des Absolutwertes des Stromes sowie der Stromflankensteigung

Erst die in [13] vorgeschlagene dynamische Gewichtung der Einzelkriterien führt zu einer gleichzeitigen Verbesserung aller drei Ziele: die Durchstrahlung konnte um die Hälfte reduziert werden, wobei die Stromflankensteigung in etwa verdoppelt wurde und der Absolutwert des Stromes mit dem Faktor 2,5 zunahm.

## 5 Ergebnisse zur modellbasierten Designoptimierung medizinischer Teststreifen

Ebenfalls im Rahmen des Verbundprojektes OMID wurde eine Optimierung des Mikrosystems „Medizinische Teststreifen“ durchgeführt. Abbildung 10 zeigt den mit Säulen strukturierten Bereich eines Mikrokanals.



**Abbildung 9: Strukturierter Mikrokanal**

In Kooperation mit dem Industriepartner STEAG microParts GmbH und dem Forschungszentrum Karlsruhe wurde am ITEM ein Verhaltensmodell entwickelt, das die Grundlage der modellbasierten Optimierung bildet [15].

Ziel der Optimierung war es zunächst, einen optimalen Zusammenhang der Strukturhöhe

und des Säulenabstands zu finden. Für eine vorzugebende Befüllungszeit werden die entsprechenden Werte ermittelt, so dass der Industriepartner einen Wert vorgibt, beispielsweise nach prozesstechnischen Belangen und über den anderen die gewünschte Befüllungszeit einstellen kann. In der Anwendung entspricht diese Zeitvorgabe Ablösezeiten für Reagenzien oder chemischen Reaktionszeiten, die möglichst genau erreicht werden sollen.

Als Messtrecke dient der Kanalbereich zwischen der Position 1 mm hinter dem Kanaleinlass und 7 mm. Dieser Kanalbereich soll in einer bestimmten Zeit von der Fließfront überstrichen werden.

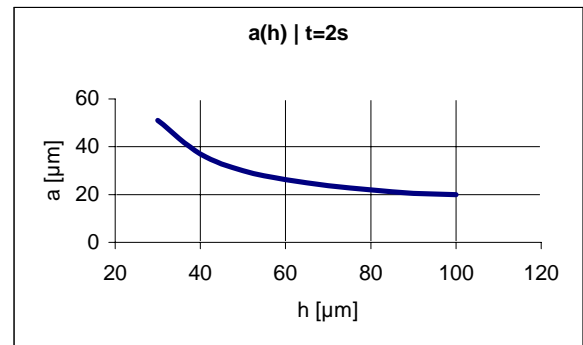
Es ist ein Zusammenhang zwischen Strukturhöhe, Säulenabstand, Druck und Volumenstrom gesucht, der den jeweils gültigen Arbeitspunkt bestimmt. Eine geschlossene Zielfunktion ist hierfür nicht definierbar. Die Grundlage zur Lösung dieses Problems bilden Variantensimulationen, in denen der Fluss in Abhängigkeit vom Druck einerseits, sowie von Strukturhöhe und Säulenabstand in verschiedenen Kombinationen andererseits bestimmt wird. Die Überführung der Simulationsergebnisse in eine analytische Beziehung erfolgte zunächst anhand der einzelnen Parameterkurven, die mit der gleichen Fitfunktion approximiert wurden, anschließend wurden die erhaltenen Fitparameter in Abhängigkeit der Optimierungsparameter erneut approximiert.

Für Aussagen über die Befüllungszeit der Teststrecke benötigt man eine genaue Erfassung von simulierter Zeit und erreichter Fließfrontposition. Daher wurde ein linearer Interpolationsansatz zwischen den räumlichen und zeitlichen Simulationswerten implementiert, welche die Start- und Endpositionen der Zeitmessung umgeben. Somit werden die genauen Zeit- und Positionsmarken für die Teststrecke ausreichend genau approximiert.

## 5.1 Optimierungsergebnis

Je nach Vorgabe des Optimierungszieles gibt es bei zwei gleichgerichtet wirkenden Parametern einen mehr oder weniger großen Lösungsraum an Arbeitspunkten, in dem sich für jeden Wert des einen Parameters ein Wert des anderen finden lässt, um die gleiche vor-

gegebene Befüllungszeit für eine vorgegebene Teststrecke zu erreichen. Das Optimierungsergebnis ist unendlich vieldeutig, d.h. es ergibt sich kein optimaler Arbeitspunkt, sondern ein als optimal ermittelter Zusammenhang zwischen den beiden Parametern, wie in Abbildung 10 dargestellt.



**Abbildung 10: Parameteroptimierung für Abstand und Höhe der Säulen**

Für eine frei gewählte Zeit von 2 Sekunden für die Teststrecke 1 mm bis 7 mm ergibt sich die obige Zuordnung von Wertepaaren. Qualitativ ist nachvollziehbar, dass sich bei einer großen Strukturhöhe der Kanalwiderstand verringert, was durch enger stehende Säulen, also einem kleineren Säulenabstand ausgeglichen wird, um die gleiche Befüllungszeit, d.h. den gleichen Gesamtfluss zu erhalten und umgekehrt.

Da es für eine zu formulierende Zielfunktion kein eindeutiges Minimum gibt, werden zunächst eindimensionale Optimierungsansätze verfolgt, bei denen für feste Werte der Strukturhöhe der Säulenabstand dem Zeitkriterium entsprechend eingestellt wird. Für die Aufnahme einer Kurve wie in Abbildung 10 wird somit je ein Optimierungslauf pro Stützstelle benötigt.

## 6 Zusammenfassung

In diesem Beitrag wurde der Hintergrund zur Entwicklung des modellbasierten Designoptimierungssystems MODOS kurz erläutert, seine Anwendung im Bereich der Mikrosystemtechnik an verschiedenen Anwendungen demonstriert sowie die Entwicklung verschiedener neuer mathematischer Optimierungsalgorithmen erläutert.

Die im Projekt betrachteten Anwendungen verdeutlichen verschiedene Aspekte, die bei

der automatisierten Auslegung von Mikrosystemen zu berücksichtigen sind. So sind in der Regel multikriterielle Optimierungsaufgaben zu bearbeiten, wobei die einzelnen Optimierungsziele auch gegenläufig sein können, wie dies bei der MFIK der Fall war. Die Strukturoptimierung der Aktorplatte hat auf beeindruckende Weise verdeutlicht, welchen Vorteil die Entwickler aus der Optimierung ziehen können.

## Literatur

- [1] Schneider, A.; Schneider, P.; Gramatová, E.; Ivask, E.: Internetbasierter Systementwurf mit MOSCITO. 10. E.I.S. Workshop, Dresden, 3.-5. April 2001
- [2] Gibron, M.: Universelle Schnittstelle zur Werkzeugintegration in SABER. 2. Statusseminar zum Verbundprojekt Modellbildung für die Mikrosystemtechnik, Paderborn, 1:75-82, Dezember 1997
- [3] Peters, D.; Laur, R.: Microsystem Model Evaluation Via Optimization Strategies. Proceedings of the 4<sup>th</sup> International Conference on Modeling and Simulation of Microsystems, South Carolina, März 2001
- [4] Box, M. J.: A new method of constrained optimization and a comparison with other methods. *Computer Journal*, 8:42-52, 1965
- [5] Peters, D.; Bolte, H.; Nüssen, O.; Marschner, C.; Laur, R.: Enhanced optimization algorithms for the design of microsystems. DTIP, April 2001
- [6] Rosenbrock, H. H.: An Automatic Method for finding the Greatest or Least Value of a Function. *Computer Journal*, 3:175-184, 1960
- [7] Peters, D.; Laur, R.: An Improvement of Rosenbrock's Algorithm for modelbased Design Optimization of Microsystems. Sixth IEEE International Conference on Electronics, Circuits and Systems, Zypern, 1999
- [8] Moré, J.J.; Thuente, D.J.: Line Search Algorithms with Guaranteed Sufficient Decrease. *ACM Transactions on Mathematical Software*, Vol. 20, 3:286-307, September 1994
- [9] Nelder, J.A.; Mead, R.A.: A simplex method for function minimization. *Computer Journal*, 7:308-313, 1965
- [10] Jakob, W.: HyGLEAM - An Approach to Generally Applicable Hybridization of Evolutionary Algorithms. in: J.J.Merelo, A. Panagiotis, H.-G. Beyer (Hrsg.): *Conf. Proc. PPSN VII, LNCS 2439*, Springer-Verlag, Berlin. 2002.
- [11] Jakob, W.; Peters, D.: Internetbasierte Designoptimierung mit HyGLEAM. Dieser Tagungsband.
- [12] Tahhan, I.: Die Aktorplatte – ein universeller fluidischer Mikroaktor. Dieser Tagungsband.
- [13] Bolte, H. et al.: Verhaltensmodelle und multikriterielle Optimierung im Entwurfsprozess von komplexen Systemen am Beispiel einer Mehrfachionisationskammer. ASIM 2001
- [14] Brachtendorf, H.G.; Laur, R.: A Hysteresis Model for Hard Magnetic Core Materials. *IEEE Transactions on Magnetics*, vol. 33, (1)723-727, 1997
- [15] Nüssen, O. et al.: Modellbasierte Optimierung Medizinischer Teststreifen aus Kunststoff, Beitrag im Tagungsband des OMID-Abschluss-Seminars, November 2002

# BETRACHTUNG DES FLÜSSIGKEITSVERHALTENS IN MEDIZINISCHEN TESTSTREIFEN AUS KUNSTSTOFF

C. Schön, I. Ballhorn, STEAG microParts GmbH  
Email: christian.schoen@microparts.de

## Kurzfassung

Um die Optimierungszyklen für den fluidischen Teil von medizinischen Reaktionsplattformen aus Kunststoff abkürzen zu können, wird angestrebt, Designvarianten mittels Modellbildung und Simulation auf ihre Funktion zu überprüfen und die Optimierung ebenfalls durch mathematische Verfahren zu beschleunigen.

Dazu werden typische mikrofluidische Strukturen (verschiedene Kanäle mit Mikrostrukturen) mittels Mikrospritzguss aus Kunststoff hergestellt und das fluidische Verhalten von Flüssigkeiten in diesen Kanälen mit den Simulationsergebnissen verglichen.

## 1 Einleitung

Mikrofluidische Einmalartikel aus Kunststoff bieten der medizinischen Diagnostik hohe Innovationspotenziale. So führen kleinere Reaktionsvolumina zu einem geringeren Chemikalienverbrauch und zu schnelleren Reaktionszeiten, die Integration von verschiedenen Funktionen und Analysen auf einer mikrostrukturierten Reaktionsplattform (Lab on a Chip-Technologie) ergeben dabei gleichzeitig neue Potenziale.

Zahlreiche, meist aufwändige Mikrostrukturierungsverfahren (z.B. optische Lithographie, Laserbearbeitung, Präzisionserodieren, Reaktives Ionenätzen von Silizium oder mikromechanische Bearbeitung) werden bei STEAG microParts GmbH zur Herstellung einer Master-Struktur eingesetzt. Hieraus wird anschließend ein Formeinsatz hergestellt, der die kostengünstige Massenproduktion von mikrostrukturierten Artikeln in Kunststoff ermöglicht.

Wegen der kostenintensiven Herstellung der Formeinsätze ist es anzustreben, die Optimierung nicht mittels Testreihen an verschiedenen Designvarianten von abgeformten Kunst-

stoffteilen durchzuführen, da dies die Herstellung eines Formeinsatzes für jede Variante bedeuten würde. Dies ist sehr zeit- und kostenintensiv und kann die Markteinführung eines Produktes um mehrere Monate verzögern.

Als Alternative zur iterativen Annäherung an die optimale Struktur mittels Testmustern werden häufig Simulation und computergestützte Optimierung genannt. Es existieren verschiedene Softwaretools, die für mikrofluidische Problemstellungen geeignet erscheinen, wie z.B. CFX der Firma AEA Technology oder FiDAP der Firma Fluent.

Diese Softwarepakete bieten Unterstützung für die Untersuchung von mikrofluidischen Vorgängen, bei denen die Kapillarkraft die treibende Kraft ist bzw. lassen das Erweitern des Paketes durch eigenen Code zu. Diese Pakete sind jedoch nicht ursprünglich für die Simulation von mikrofluidischen Aufgabenstellungen entwickelt worden. Sie kommen aus dem Makrobereich (z.B. Pipelines, Rauchausbreitung in Gebäuden usw.) und wurden dann durch zusätzliche Module um die erforderlichen mikrofluidischen Bestandteile erweitert.

In der Mikrofluidik spielen jedoch Effekte wie Benetzung einer Oberfläche, Adhäsion der Flüssigkeit an der Kanalwand oder Oberflächenspannung eine deutlich größere Rolle [1] als dies im Makrobereich der Fall ist, wo bei Simulationen die genannten Phänomene vernachlässigt werden dürfen.

Es ist daher zu prüfen, ob die durch Simulation ermittelten Parameter tatsächlich mit den realen Größen übereinstimmen. Dazu werden typische mikrofluidische Strukturen durch Abformung mittels Mikrospritzguss in Kunststoff hergestellt. Das sind hier verschiedene Kanäle mit Tiefen zwischen 25µm und 50 µm, die mit unterschiedlichen Mikrostrukturen (Säulen verschiedener Formen und Querschnitte) besetzt sind. An diesen werden Mes-

sungen durchgeführt, insbesondere wird die Fließgeschwindigkeit unter verschiedenen Bedingungen gemessen und mit den durch Simulation ermittelten Werten verglichen. Wenn die Kennwerte von Messung und Simulation bei den vorliegenden Teststrukturen übereinstimmen, kann davon ausgegangen werden, dass die modellbasierte Simulation und Optimierung von mikrofluidischen Strukturen eine Alternative zur Optimierung mittels Testmustern und Prototypen ist. Bei Einsatz der computergestützten Verfahren wäre eine deutlich schnellere Entwicklung von medizinischen Teststreifen aus Kunststoff u.ä. möglich.

## 2 Herstellung von medizinischen Teststreifen

### 2.1 Formeinsatzherstellung

Die Formeinsatzherstellung ist die Basis für die Abformung von medizinischen Einmalartikeln aus Kunststoff. Durch die tausendfache Abformung eines Formeinsatzes können solche Kunststoffteile auch unter Einhaltung der erforderlichen hohen Genauigkeit kostengünstig in großen Stückzahlen hergestellt werden.

Bei der Formeinsatzherstellung für medizinische Teststreifen und Reaktionsplatten (Lab on a Chip-Technologie) werden spezialisierte Verfahren eingesetzt. Dazu zählen optische Lithographie, Excimer-Laserbearbeitung, Präzisionsrodieren, Reaktives Ionenätzen von Silizium oder mikromechanische Bearbeitung mit Formdiamanten. Durch diese Bearbeitungsverfahren wird eine Masterstruktur hergestellt, die in nachfolgenden Schritten durch Galvanoformung in einen Formeinsatz kopiert wird.

Diese Metallkopie der Masterstruktur (also das Negativ) wird nach einigen zusätzlichen Bearbeitungsschritten als Formeinsatz eingesetzt.

### 2.2 Abformung

Der metallische Formeinsatz wird zur Replikation des Kunststoffteiles eingesetzt. Dazu stehen STEAG microParts GmbH neben Kunststoff-Spritzguss auch Heißprägeverfah-

ren und andere zur Verfügung. Die im Projekt genutzten Teststrukturen wurden durch Spritzguss hergestellt.

Neben einfachen Mikrokanälen ohne Mikrostrukturen (siehe Abbildung 1) wurden in einer 2. Projektstufe mittels lithographisch hergestellter Formeinsätze auch kompliziertere Mikrostrukturen abgeformt. Diese bestehen aus einem Kanal mit den gleichen Abmessungen wie die einfachen Kanäle (Breite 3 mm, Tiefe 25  $\mu\text{m}$  oder 50  $\mu\text{m}$ ), haben jedoch im Kanal Mikrostrukturen (Säulen).



**Abbildung 1: Teststruktur: einfache Mikrokanäle**

Um die Messungen auch untereinander vergleichen zu können, wurden mehrere Varianten der mikrostrukturierten Kanäle hergestellt. Es wurden sowohl die Art der Mikrostrukturen als auch die Geometrie variiert. Außerdem wurden verschiedene Deckel hergestellt, um auch den Einfluss des Deckels auf das fluidische Verhalten beurteilen zu können.

Eine Auflistung der verwendeten Teststrukturen ist in Tabelle 1 dargestellt.

**Tabelle 1: Verwendete Teststrukturen**

Name	Mikrostruktur	Kanaltiefe
Bodenplatte 1	Runde Säulen, $\varnothing 10\mu\text{m}$	25 $\mu\text{m}$
Bodenplatte 2	Runde Säulen, $\varnothing 10 \text{ mm}$	50 $\mu\text{m}$
Bodenplatte 3	Lamellen, $b = 10 \mu\text{m}$ , $l = 50\mu\text{m}$	25 $\mu\text{m}$

Eine Aufnahme mit dem Rasterelektronenmikroskop (REM) der Mikrostruktur von Bodenplatte 1 ist in Abbildung 2 zu sehen. Das Bild zeigt einen Ausschnitt des Abform-



teils vom Kanalanschluss. Das gezeigte Kunststoffteil wurde mittels Mikrospritzguss hergestellt.

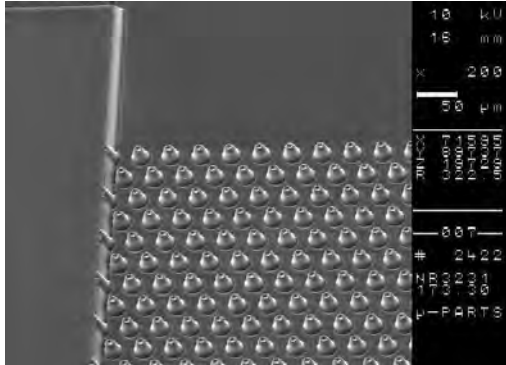


Abbildung 2: Kanal mit Mikrostrukturen

### 2.3 Oberflächenmodifikation

Bei medizinischen Teststreifen und anderen Reaktionsplattformen aus Kunststoff folgt nach der Herstellung meist noch eine Modifikation der Oberflächeneigenschaften, da die Oberflächeneigenschaften des reinen Kunststoffes erfahrungsgemäß das gewünschte fluidische Verhalten nicht zufriedenstellend unterstützen. Dabei sind je nach Einsatzzweck sowohl hydrophobe als auch hydrophile Oberflächeneigenschaften einstellbar. Weiterhin ist es möglich, durch Beschichtungen bestimmte chemische Eigenschaften zu erzielen. So kann z.B. durch Beschichtungen ein besonderes gutes Bindungsverhalten bestimmter Molekülketten an der Oberfläche erreicht werden. Für eine Befüllung mittels Kapillarkraft ist natürlich eine hydrophile Behandlung der Oberfläche nützlich. Durch die Behandlung sinkt der Kontaktwinkel von Wasser mit der Kunststoffoberfläche von typischerweise  $80^\circ$  auf ca.  $40^\circ$ , was eine gute Benetzbarkeit und damit auch Kapillarfüllung der Kanäle sicherstellt.

### 2.4 Assemblierung

Nach der Oberflächenmodifikation werden die Kanalstrukturen noch gedeckelt, d.h. die bis dahin noch offene Oberseite der Kanäle wird verschlossen. So wird der Kanal durch Boden, Seitenwände und Deckel an allen 4 Seiten begrenzt und eine gleichmäßige Befüllung des Kanals durch Kapillarkräfte erreicht.

Dazu stehen mehrere Verfahren zur Verfügung, wie z.B. Ultraschallschweißen, Thermi-

ches Bonden, Heißsiegeln, Schweißen, Kleben usw. Bei den vorliegenden Teststrukturen wurden die Kanäle mit einer speziellen selbstklebenden Folie gedeckelt. Der mit der Flüssigkeit in Kontakt stehende Kleber der Folie hat ähnlich hydrophile Benetzungseigenschaften wie der Kunststoff nach der Modifikation. Dies ist wichtig für die Simulation, da die verwendete Simulationssoftware nicht die Möglichkeit bietet, im Modell verschiedene Oberflächeneigenschaften für die einzelnen Kanalseiten zu definieren.

## 3 Messungen an medizinischen Teststreifen

### 3.1 Allgemeines

In den ersten Projektstufen wurden sehr einfache Mikrokanäle vermessen und fluidischen Tests unterzogen. Diese Kanäle hatten eine Tiefe zwischen  $8\ \mu\text{m}$  und  $100\ \mu\text{m}$ , die Länge und Breite waren konstant  $20\ \text{mm}$  (Länge) bzw.  $3\ \text{mm}$  (Breite).

Diese einfachen Kanäle wurden zum einen deshalb verwendet, weil so einige Effekte in den Hintergrund treten und die Messungen leichter zu interpretieren sind. Zum anderen sind einfache Kanäle leichter herstellbar, da bei der Formeinsetzung auf mikromechanische Bearbeitung (evtl. mit direkter Herstellung des Formeinsetzes) anstelle von aufwändigeren Verfahren zurückgegriffen werden kann. Dadurch kann die Kanalgeometrie ohne größeren Aufwand hochgenau vermessen und in das Modell übertragen werden, ohne dass das Modell sehr groß wird. Zusätzlich sinkt aufgrund des kleineren Modells der Zeitbedarf für die Simulation.

Nach Abschluss dieser Messungen wurden in einer weiteren Projektstufe Mikrokanäle mit Mikrostrukturen vermessen, die realen Mikrofluidikkomponenten näher kommen.

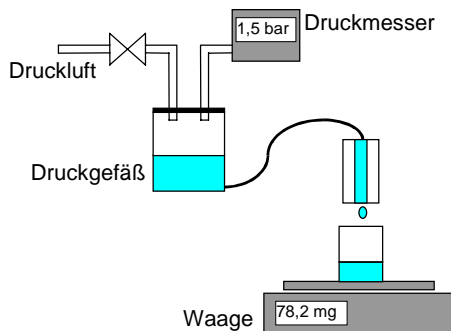
An beiden Strukturen wurden mehrere Messungen durchgeführt. Zum einen wurden zur Validierung des Modells Messungen benötigt, die den Volumenstrom bei einer bestimmten Druckdifferenz zum Ergebnis haben. Dieser Versuch kann relativ einfach simuliert werden, da hier die Kapillarkraft nicht als treibende Kraft in der Simulation berücksichtigt

werden muss. Solche Problemstellung sind einfacher zu handhaben als Kapillarströmungen, da einige Effekte wie Phasenübergänge und Kapillareffekte nicht vorkommen.

Zum anderen wurden Messungen der Strömungsgeschwindigkeit bei Kapillarbefüllung durchgeführt. Auch dies ist eine wichtige Messung, da ein optimales Befüllverhalten des Teststreifens eines der Hauptziele der Entwicklung ist.

### 3.2 Messaufbau

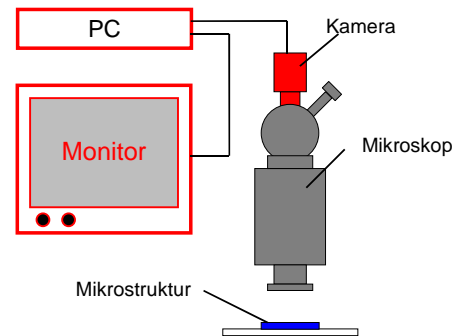
Zur Messung des Durchflusses bei Beaufschlagung der Mikrokanäle mit einer bestimmten Druckdifferenz wurde ein Messaufbau verwendet, wie er in Abbildung 3 dargestellt ist.



**Abbildung 3: Messaufbau Druckbefüllung**

In einem Druckgefäß mit konstantem Druck befindet sich die Flüssigkeit, die durch den Kanal fließt und nach dem Austreten aufgefangen wird. Das Volumen wird über das Gewicht der Flüssigkeit bestimmt. Effekte wie Temperaturabhängigkeit der Viskosität, Einfluß der Kapillarkraft, Verluste durch Arbeit beim Abreißen des Tropfens, Verdunstung im Auffanggefäß oder Ausbildung einer Helmholtz-Schicht wurden entweder durch apparative Maßnahmen eingeschränkt oder nach eingehender Untersuchung vernachlässigt.

Die Befüllung der Kanäle mittels Kapillarkraft wurden unter einem Mikroskop durchgeführt und mittels Framegrabberkarte digital aufgezeichnet. Der prinzipielle Messaufbau ist in Abbildung 4 dargestellt.



**Abbildung 4: Messaufbau Kapillarbefüllung**

Die Auswertung der Aufnahmen erfolgt dann am PC. Ausgewertet werden Werte wie Fließgeschwindigkeit, Winkel der Fließfront zur Seitenwand, Geschwindigkeitsprofil usw.

## 4 Vergleich von Messung und Simulation

### 4.1 Messungen an einfachen Mikrokanälen

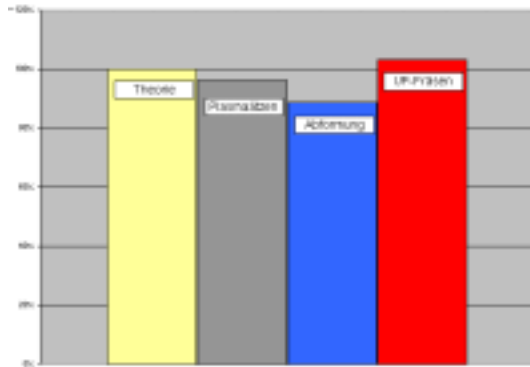
Die Messungen an einfachen Mikrokanälen ergaben eine gute Übereinstimmung von Simulation und Messung. Die Übereinstimmung lag im Bereich von 90% (siehe Diagramm 1). Diese Werte wurden zusätzlich an Testteilen überprüft, die mit anderen Verfahren (reaktives Ionenätzen/ Plasmaätzen von Silizium sowie Ultrapräzisionsfräsen) hergestellt wurden.

Die Simulationen und Berechnungen basieren auf dem Modell der planparallelen Platten, bei dem davon ausgegangen wird, dass die Kanaltiefe klein gegenüber der Kanalbreite ist.

$$\dot{V} = \frac{d^3 b}{12\eta \cdot l} \cdot dp$$

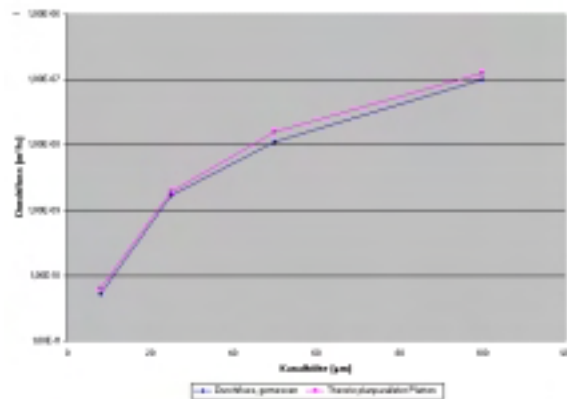
Dass diese Annahme gültig ist, wurde durch Versuche mit Kanälen verschiedener Breite belegt.

Die Übereinstimmung von 90% wurde als ausreichend angesehen. Mit dieser Grundlage konnte begonnen werden, auch Messungen und Simulation von komplizierteren Strukturen durchzuführen, wo der Aufwand besonders bei der Modellbildung und Simulation deutlich höher liegt.



**Diagramm 1: Einfache Kanäle: Übereinstimmung Theorie/ Messung**

Die Abhängigkeit des Durchflusses von der Kanaltiefe ist in Diagramm 2 dargestellt. Auch hier wird die Übereinstimmung der gemessenen Werte mit den theoretisch zu erreichenden deutlich.

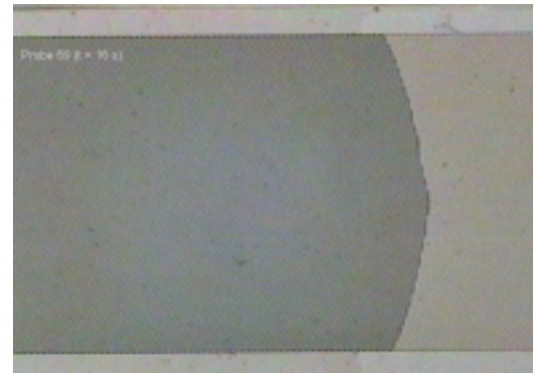


**Diagramm 2: Abhängigkeit des Durchflusses von der Kanaltiefe**

## 4.2 Messungen an Kanälen mit Mikrostrukturen

Als nächste Stufen wurden die Messungen an mikrostrukturierten Kanälen in Angriff genommen. Auch hier wurden sowohl Durchflussmessungen mit Druckbeaufschlagung als auch die Befüllung mittels Kapillarkraft ausgewertet.

Abbildung 5 zeigt beispielhaft eine Momentaufnahme der Befüllung eines mikrostrukturierten Kanals mit einer Flüssigkeit durch Kapillarkraft. Der helle Bereich auf der rechten Seite des Kanals ist die Flüssigkeit, der dunkle Bereich auf der linken Seite ist noch mit Luft gefüllt. Die typische Ausbildung der Fließfront („Smiling“-Effekt) ist gut zu sehen. Die diskreten Stufen in der Fließfront lassen die Säulenstruktur im Kanal erkennen.



**Abbildung 5: Kapillarbefüllung einer mikrofluidischen Teststruktur**

Die so ermittelten Werte von Druck- und Kapillarbefüllung wurden im nächsten Schritt mit den Simulationsergebnissen der Projektpartner Forschungszentrum Karlsruhe/ IAI und Universität Bremen/ ITEM verglichen [2].

Als letzter (und wichtigster Schritt) erfolgte dann eine modellbasierte Optimierung der Strukturen [2]. Ziel der Optimierung war die Befüllung des mikrostrukturierten Kanals durch Kapillarkraft in einer vorgegebenen Zeit.

## 5 Zusammenfassung

Es wurden mikrofluidische Strukturen mittels Mikrospritzguss hergestellt, wie sie in abgewandelter Form auch in medizinischen Teststreifen oder Reaktionsplattformen (Lab on a Chip-Technologie) zum Einsatz kommen könnten. Das Verhalten von Flüssigkeiten in diesen Strukturen wurde gemessen und beobachtet und mit Simulationsergebnissen verglichen. Eine gute Übereinstimmung zwischen Simulation und Messung wurde bei einfachen Mikrokanälen gefunden.

Der Vergleich von Simulationsergebnissen und Messungen bei komplizierteren Strukturen, wie z.B. 50 µm tiefen Kanälen mit Mikrostrukturen in Form von Säulen mit einem Durchmesser von 10 µm wurden durchgeführt. Der Vergleich ergab, dass die Werte von Messung und Simulation in der gleichen Größenordnung liegen.

Es kann also davon ausgegangen werden, dass die rechnergestützte Simulation und Optimierung in der Lage ist, den Entwickler bei der Auslegung mikrofluidischer Netzwerke

zu unterstützen. Die Ergebnis stimmten bei den hier untersuchten mikrofluidischen Strukturen zwar nicht 100%ig mit den Messwerten überein, geben dem Entwickler aber deutliche Hinweise, in welche Richtung weiter entwickelt werden muss, um ein zufriedenstellend funktionierendes Design zu finden.

Dies hat zur Folge, dass ein (vor-)optimiertes Design für eine mikrofluidische Struktur nicht mehr mittels Versuchen an realen Funktionsmustern gefunden werden muss, sondern auch (deutlich schneller) über mathematische Optimierungsalgorithmen ermittelt werden kann.

## **6 Literatur**

[1] M.J. Harper, C.M. Turner, J.E.A. Shaw: Fluid modelling in micro flow channels, *Microsystems and Microstructures*, S. 155-163

[2] O. Nüssen, S. Halstenberg, D. Peters: Modellbasierte Optimierung medizinischer Teststreifen aus Kunststoffen, *Abschlussseminar OMID*, Nov. 2002, Proceedings

# MODELLBASIERTE OPTIMIERUNG MEDIZINISCHER TESTSTREIFEN AUS KUNSTSTOFF

Dipl.-Ing. O. Nüssen, ITEM, Universität Bremen  
Dr.-Ing S. Halstenberg, IAI, Forschungszentrum Karlsruhe  
Dr.-Ing. D. Peters, ITEM, Universität Bremen  
nuessen|peters@item.uni-bremen.de, halsten|eggert@iai.fzk.de

## Kurzfassung

In Kooperation zwischen dem Industriepartner STEAG microParts GmbH, dem Forschungszentrum Karlsruhe und der Universität Bremen wurden Simulationsmodelle für die Optimierung strukturierter Mikrokanäle erstellt. Die generierten Verhaltensmodelle basieren auf Berechnungen mit dem Fluidiksimulator CFX unter Einsatz der Finite-Volumen-Methode (FVM), welche Strömungswiderstände und Kapillardrücke von Teilstrukturen als Kennwerte in Abhängigkeit der Strukturgeometrien liefern. Die Ergebnisse erlauben Rückschlüsse auf das Gesamtsystem unter Ausnutzung der gegebenen Symmetrien.

## Einleitung

Anders als beim hochstandardisierten Entwurf mikroelektronischer Schaltungen, bringt das Design komplexer Mikrosysteme hochspezifische Herausforderungen mit sich. Die durch einzelne Problemlösungen gewonnenen Erkenntnisse lassen sich oft nur unzureichend verallgemeinern, um methodische Ansätze für nachfolgende Aufgaben zu erzielen. Schwierigkeiten liegen einerseits in der Komplexität der meist hybriden Systeme selbst, andererseits stammen die für die Systemeigenschaften relevanten Größen oft aus ganz unterschiedlichen Domänen, was beispielsweise in sehr divergenten numerischen Größenordnungen resultieren kann, denkt man an Verarbeitungszeiten einer Auswerte-Elektronik im Vergleich zu thermischen Ausgleichsvorgängen, welche gleichberechtigt in einem Modell erfasst sein müssen. Darüberhinaus sind nur wenige Werkzeuge überhaupt in der Lage, eine Gesamtsystembetrachtung durchzuführen.

Zwar lassen sich prinzipielle Aussagen für viele Mikrosysteme treffen, die Umsetzung in

quantitative Beschreibungen ist jedoch stärker als bei rein elektrischen Schaltungselementen von der Beschaffenheit des Systems abhängig. Somit existieren noch vergleichsweise wenige allgemeingültige MST-Modellbibliotheken, derer sich ein Entwickler heute bedienen kann, weshalb aktuelle Bestrebungen dazu dienen, den Entwurf von Mikrosystemen stärker zu modularisieren und über das Baukastenprinzip zu einer einheitlicheren Entwurfsmethodik zu gelangen.

Über die oftmals neuartigen Eigenschaften hybrider Mikrosysteme lassen sich exakte Aussagen nicht selten nur über die Herstellung von Prototypen gewinnen, was erheblichen Einsatz an Produktionsmitteln und -zeit durch iterative Annäherung an ein optimales Design mit sich bringt. Es ist Ziel des Projektes OMID, durch den Einsatz von Modellbildung, Simulation und Optimierung Wege aufzuzeigen, um bereits in einer frühen Entwurfsphase zu einem optimalen Satz an Entwurfsparametern zu gelangen und somit zu deutlichen wirtschaftlichen Vorteilen bezüglich der Innovationszyklen und der Zeit zur Markteinführung [1].

Gerade im Bereich der Mikrosystemtechnik finden Modellierung, Simulation und modellbasierte Optimierung noch keinen flächendeckenden Einsatz in der Entwurfsmethodik. Die vorliegende Problemstellung bei der Modellierung des Fließverhaltens in medizinischen Teststreifen kann einschließlich der bisher erzielten Ergebnisse aus dieser Sicht heraus als exemplarisch für die grundsätzliche Vorgehensweise bei der modellbasierten Optimierung von Mikrosystemen angesehen werden.

## Medizinische Teststreifen

Der Einsatzbereich medizinischer Kunststoff-Teststreifen liegt hauptsächlich in der Analyse von Körperflüssigkeiten hinsichtlich spezieller Eigenschaften oder des Nachweises von Antikörpern auf bestimmte Krankheitserreger. Die Herstellung miniaturisierter Einweg-Analysesysteme verspricht eine kostengünstige Anwendung und einen Schritt auf dem Weg zu einem integrierten Mikrolabor ('Lab on a Chip').

### 1.1 Systembeschreibung

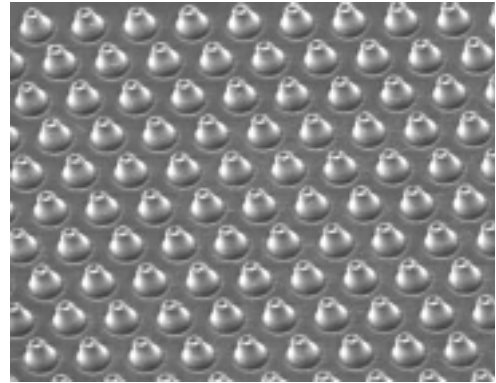
Die aus Kunststoff gefertigten medizinischen Teststreifen unterteilen sich in verschiedene Funktionsbereiche, die geometrisch unterschiedlich gestaltet sind. Die Reaktionszonen, in welche die Probenflüssigkeit aufgrund der Kapillarkwirkung hineintransportiert wird, sind für das Systemverhalten von hoher Relevanz. Die Reaktion der Probenflüssigkeit mit den dort befindlichen Reagenzien bewirkt beispielsweise eine Farbveränderung, eine Änderung der Fluoreszenz oder eine Änderung der elektrischen Eigenschaften, die mit entsprechenden Messverfahren ausgewertet werden kann. Die strukturierten Bereiche der Mikrokanäle besitzen jeweils typische Verweildauern bzw. Befüllungsgeschwindigkeiten, die sich aus dem strömenden Medium, aus dem Strukturmaterial und aus den geometrischen Abmessungen ergeben [2].

Die zugrundeliegenden physikalischen Effekte sind zwar teilweise bekannt, jedoch auf mikroskopischer Ebene zum Teil noch nicht analytisch beschreibbar. Beispielsweise ändert sich das Verhältnis von Oberflächeneffekten zu Volumeneffekten signifikant mit der Reduzierung der Strukturgröße, was eine Umformulierung der makroskopischen Gesetzmäßigkeiten erforderlich machen kann. Eine wichtige Rolle spielen hierbei die Kapillarkwirkung und die weiteren Interaktionen an den Grenzschichten zwischen unterschiedlichen Medien. Derartige Einflussfaktoren sind noch nicht vollständig quantifizierbar.

### 1.2 Aufbau der Teststrukturen

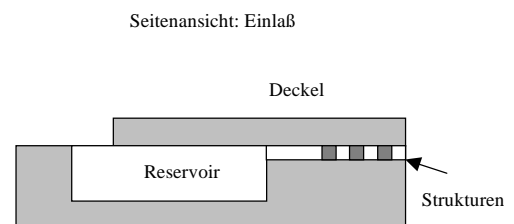
Um die Eignung von Modellbildung und Simulation im Entwurfsprozess der Teststreifen

zu verifizieren, wurden Teststrukturen vereinbart. Es handelt sich dabei um rechteckige Kanäle von 20 mm Länge und einer Breite von 3 mm, die im Durchflussbereich zum einen mit Säulen wie in Abb.1 und zum anderen mit Lamellen bestückt sind, um einen definierten Durchflusswiderstand zu erhalten.



**Abb.1 Teststrukturen**

Um zusätzliche Parametervariationen in die Betrachtung aufzunehmen, beträgt die Tiefe der Kanalstrukturen einmal 25  $\mu\text{m}$  und einmal 50  $\mu\text{m}$ . Die Säulen haben einen festen oberen Durchmesser und verbreitern sich mit dem anisotropen Ätzwinkel des Herstellungsprozesses zum Kanalboden hin. Die Anzahl der Strukturelemente im Kanal ist zum einen abhängig von den Kanalabmessungen, zum anderen wird der Abstand zwischen zwei Säulenmittelpunkten als variabel betrachtet.



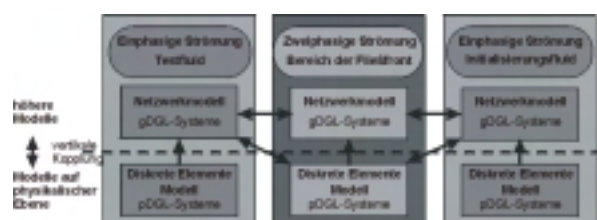
**Abb.2: Kanal mit Deckel und Einlass**

Der Kanal wird mit einem Deckel sowie mit einem Reservoir zur Befüllung gemäß Abb.2 versehen. Als Deckel fungiert eine hydrophile Klebefolie, die Testflüssigkeit ist ein Gemisch aus Polyethylenglykol (PEG) und Wasser, dessen fluidische Kenngrößen wie Viskosität, Dichte und Kontaktwinkel sich definiert einstellen lassen und als Richtparameter für die anschließenden FVM-Simulationen dienen.

## Ansatz zur Erstellung eines Gesamtmodells

Das Voranschreiten der Flüssigkeit und ihrer Fließfront in einem mikrostrukturierten Kanal geht mit einer Änderung der Grenzflächenform einher. Um diese Änderung zu erfassen, können Methoden eingesetzt werden, die mit einer hohen Diskretisierung des Strömungsraumes arbeiten, wie z.B. die Finite-Volumen-Methode (FVM). Aufgrund des verwendeten numerischen Verfahrens liegen jedoch die erforderlichen Simulationszeiten selbst bei Modellen einfacher Mikrokanäle unter Verwendung der vorliegenden Symmetrieeigenschaften bereits im Stundenbereich [3]. Für komplexere Geometrien sind erheblich größere Modelle notwendig und damit verbunden ein erhöhter Simulationsaufwand, der eine Systemoptimierung auf diesem Wege nicht mehr zulässt. Deshalb wurde folgendes Konzept zur Erstellung eines geeigneten Gesamtmodells erarbeitet, welches das Strömungsgebiet in drei Bereiche zerlegt (siehe Abb.3):

- Einphasige Strömung hinter der Fließfront, die das Verhalten des einströmenden Testfluids beschreibt
- Zweiphasige Strömung im Bereich der Fließfront
- Einphasige Strömung vor der Fließfront, d.h. Verhaltensbeschreibung des Initiationsfluids, das aus dem Kanal verdrängt wird



**Abb.3: Konzept zur Erstellung eines Gesamtmodells**

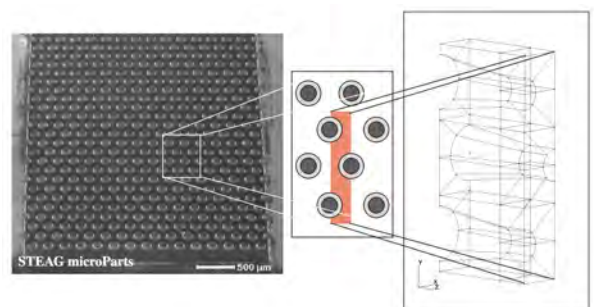
Für jeden Bereich werden Modelle auf der physikalischen Modellebene erstellt und deren Simulationsergebnisse bei der Erstellung von Modellen der höheren Modellierungsebene (z.B. Netzwerkmodelle, Verhaltensmodelle) verwendet. Diese werden zu einem Gesamtmodell gekoppelt, welches einen deutlich geringeren Simulationsaufwand aufweist.

## FVM-Modellbildung und Simulation

Die Reaktionszonen der medizinischen Teststreifen weisen eine hohe Anzahl an regelmäßigen internen Strukturen auf. Diese Zonen sollen dahingehend optimiert werden, dass eine vollständige Reaktion innerhalb einer möglichst geringen Strecke erfolgt, sodass das benötigte Volumen des Testfluids gering bleibt. So wurde das Strömungsverhalten in Kanälen untersucht, die entweder eine Säulenstruktur oder eine Lamellenstruktur aufweisen. Druckverteilung und Geschwindigkeiten sowie die Position und die Form der Fließfront sind die interessierenden Simulationsergebnisse.

Die Geometrie des zu untersuchenden Kanals ist durch die Parameter Länge, Breite und Höhe des Kanals bestimmt. Die Geometrie der Säulenstrukturen wiederum ist durch die Höhe der Säulen und den Durchmesser ihrer Deckfläche sowie den Abstand der in einem gleichseitigen Dreieck angeordneten Säulen gegeben. Die Neigung der Säulenwände resultiert aus dem gegebenen Ätzwinkel.

Aufgrund der großen Kanalbreiten (einige Millimeter) im Vergleich zu den Säulendurchmessern und -abständen (einige Mikrometer) ist die Anzahl der integrierten Säulen bzw. Lamellen sehr hoch. Das resultierende FVM-Modell besäße eine immense Komplexität, weshalb zur Reduzierung der Modellgröße und somit der Elementzahlen die Symmetrie ausgenutzt wird (s. Abb.4).



**Abb.4: Kanal mit integrierten Säulen (Struktur und 3D-Modell)**

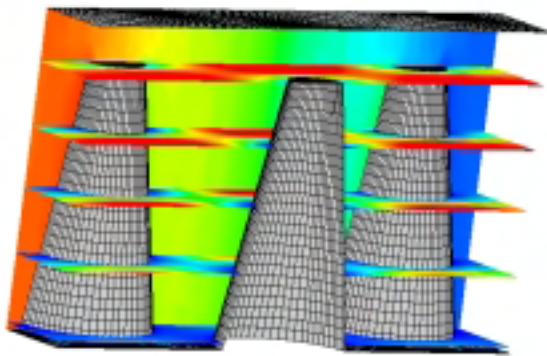
Die seitlichen Ränder des Modells werden bei inneren Zellen als Symmetrieebenen definiert, bei äußeren Zellen wird eine Seite als Wand betrachtet.

Das Modell ist parametrisierbar und wird aus

den oben genannten Geometrieparametern mit dem Präprozessor von CFX (Firma AEATechnology) erzeugt. Punkte, Linien, Flächen, Volumen und auch das FVM-Netz werden automatisch generiert. Zur Durchführung der Simulationen mit CFX werden sowohl die Symmetrie-, als auch die Ein- und Auslassbedingungen festgelegt. Wie schon im Konzept dargestellt, wurden sowohl Modelle zur Betrachtung der einphasigen Strömung als auch entsprechende zweiphasige Modelle erstellt.

### 1.3 Simulation der einphasigen Strömung

Bei der einphasigen Strömung wird das Verhalten eines Fluids untersucht, das den betrachteten Kanalauschnitt vollständig ausfüllt. Zur Simulation des Strömungsverhaltens ist die anliegende Druckdifferenz (Druck am Kanaleinlass bzw. -auslass) als Eingabeparameter vorzugeben. Der resultierende Volumenstrom durch die Kanalstruktur dient als Referenzgröße zur Erstellung höherer Modelle (siehe Abb.5).



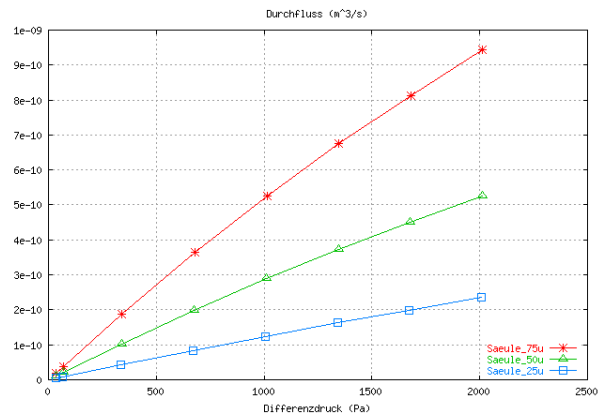
**Abb.5: Volumenstrom durch einen mikrostrukturierten Kanal**

Die horizontalen Flächen zeigen die Strömungsgeschwindigkeit in Kanallängsrichtung an, die vertikalen den Druckverlauf über der Zelle.

Mit dem vorliegenden parametrisierbaren Modell können für unterschiedliche Säulenhöhen und unterschiedliche Säulenabstände die Volumenströme in Abhängigkeit von der anliegenden Druckdifferenz berechnet werden. Diese Ergebnisse werden dann zur Generierung höherer Verhaltensmodelle benutzt.

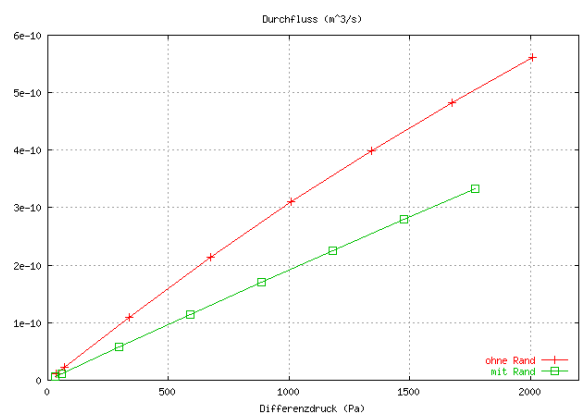
Durch Kopplung der höheren Modelle zu einem Gesamtmodell kann der Volumenstrom bei einer gegebenen Druckdifferenz für den Gesamtkanal bestimmt werden.

In Abb.6 sind die Simulationsergebnisse für drei unterschiedliche Kanalausstrukturen dargestellt.



**Abb.6: Simulierter Volumenstrom mit Säulen der Höhe 25 µm, 50 µm und 75 µm**

Die Simulationen benötigten auf einer Sparc Ultra II Maschine mit 440 MHz Rechenzeiten zwischen 3 und 20 Minuten. Diese Zeiten sind von der Modellgröße und von der Elementgröße des Netzes abhängig. Zur Simulation des Gesamtkanals wären Zehntausende der gewählten Kanalauschnitte zu berücksichtigen, was zu immensen Rechenzeiten führen würde. Ein FVM-Modell des Gesamtkanals ist nicht mehr berechenbar und kann somit nicht für eine Designoptimierung eingesetzt werden.

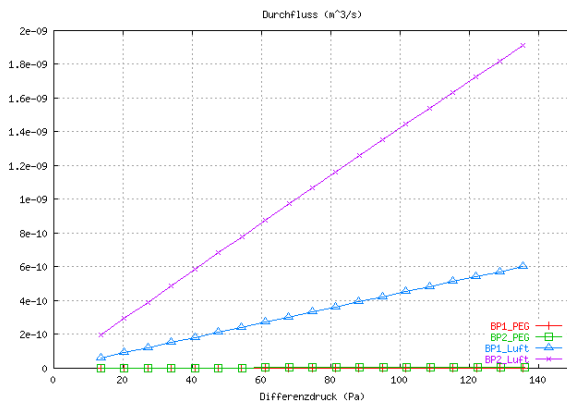


**Abb.7: Durchfluss einer Rand- und einer inneren Zelle**

Um die Effekte zu untersuchen, die von der Seitenwand des Kanals herrühren, werden nicht nur innere Zellen, sondern auch eine



Randzelle betrachtet, deren eine Seite aus einer Wand besteht (s. Abb. 7). Ist jedoch die Anzahl der inneren Zellen hoch, d.h. die Kanalbreite groß gegenüber dem Säulenabstand, so kann dieser Effekt vernachlässigt werden. In Abb. 8 sind Simulationsergebnisse zweier Bodenplatten mit den Säulenhöhen 24 bzw. 49  $\mu\text{m}$  dargestellt.



**Abb.8: Vergleich der Volumenströme von Bodenplatte 1 und 2 für Luft bzw. PEG**

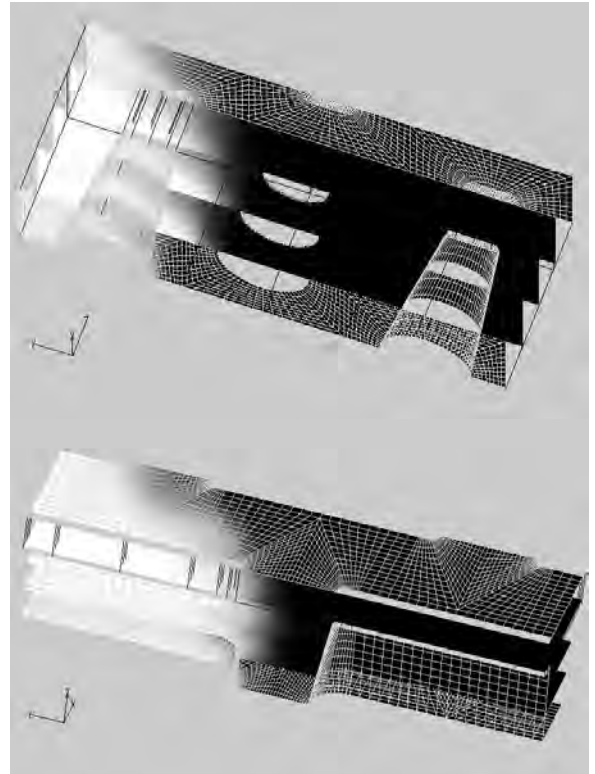
Die Bodenplatten wurden vom Projektpartner hergestellt, sodass die Simulationsergebnisse mit den entsprechenden Messdaten verglichen werden können. Im Gegensatz zu den bisher gezeigten Strukturen weisen sie keinen Abstand zwischen Säule und Deckel auf, da zur Deckelung eine Klebefolie verwendet wurde.

#### 1.4 Simulation der Fließfront

Auch für die zweiphasige Strömung wurden entsprechende Modelle erstellt. Neben den Eingabeparametern, die auch bei der einphasigen Strömung benötigt werden, müssen zusätzlich die Materialparameter Oberflächenspannung und Kontaktwinkel angegeben werden. Ebenso ist anzugeben, wie weit der Kanal zu Beginn der Simulation befüllt ist. Die Ausgabeparameter dieser Simulationen sind der Volumenstrom bzw. der Strömungswiderstand und die Form und Position der Fließfront in Abhängigkeit von der Zeit.

Die Simulation des transienten Verhaltens erfolgt bis zu einem vorgegebenen Zeitpunkt. Aufgrund des Voranschreitens der Fließfront werden die Simulationen mit einer dynamischen Kontrolle der Zeitschrittweite durchgeführt, die von der Fließfrontgeschwindigkeit und von der Elementgröße abhängig ist.

In Abb. 9 sind Simulationsergebnisse zum Zeitpunkt  $t = 0,14 \text{ ms}$  dargestellt. Es wurden die Oberflächenspannung von Wasser mit  $0,0725 \text{ N/m}^2$  und ein stationärer Kontaktwinkel von  $36^\circ$  zugrundegelegt.



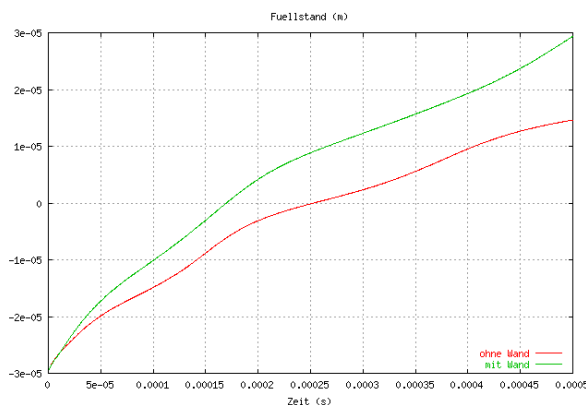
**Abb.9: Voranschreiten der Fließfront aufgrund der Kapillarität**

Zu Beginn der Simulation ist der Kanalauschnitt zur Hälfte mit Flüssigkeit befüllt. In der anderen Hälfte befindet sich Luft. Da zu Beginn der Simulation jedoch aufgrund der vorgegebenen ebenen Fließfront eine unphysikalische Situation vorliegt, bildet sich zuerst die Meniskusform der Fließfront zwischen zwei benachbarten Strukturen (Säulen bzw. Lamellen) aus. Danach beginnt die Flüssigkeit, aufgrund der Kapillarkraft vorwärts zu strömen (in der Abbildung verläuft die Fließrichtung von rechts nach links). Die Flüssigkeit ist schwarz dargestellt, die Luft dagegen weiß.

Die Flüssigkeit wird an den Wänden der Struktur, den Säulen- bzw. Lamellenwänden, dem Kanalboden und dem Kanaldeckel in den Kanal hineingezogen. Bei der Säulenstruktur konnte in der Nähe der Fließfront ein Kapillardruck von etwa  $76 \text{ hPa}$  beobachtet werden. Im Vergleich hierzu liegt in einem Kanal entsprechender Höhe, der keine

Strukturen enthält, ein Kapillardruck von etwa 47 hPa vor. Wie zu erwarten, führen somit die Strukturen zu einer Verstärkung des Kapillareffektes.

Auch diese Simulationsergebnisse werden entsprechend aufbereitet, um sie als Referenzdaten zur Erstellung höherer Modelle bereitzustellen. Hierzu wird beispielsweise der Füllstand entlang des Kanalausschnittes in Abhängigkeit von der Zeit aufgetragen (siehe Abb.10).



**Abb.10: Füllstand über der Zeit für eine Wand- und eine innere Zelle**

Um auch bei der Simulation der Fließfront die Kanalseitenwände zu berücksichtigen, wurde die Strömung sowohl einer inneren (untere Kurve) als auch einer Randzelle des Kanals (obere Kurve) untersucht. Neben dem Füllstand wird auch der Volumenstrom am Modelleinlass und der Kapillardruck über der Zeit bestimmt.

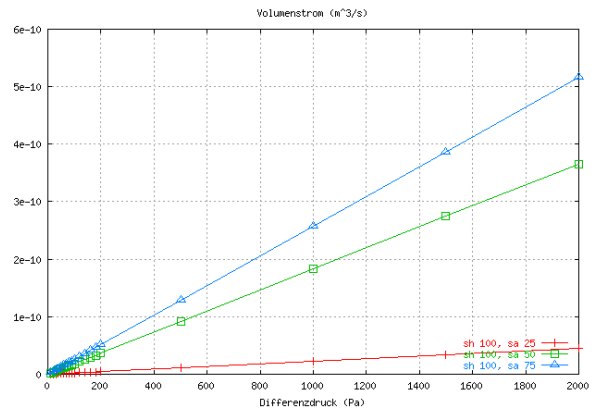
Mittels der gekoppelten höheren Modelle (Kopplung von mehreren einphasigen und zweiphasigen Modellen) wird dann das Strömungsverhalten für den gesamten Kanal berechnet und kann mit Messergebnissen verglichen werden.

## 1.5 Variantensimulationen

Zur Durchführung einer ersten Optimierung sollen der Säulenabstand und die Säulenhöhe variiert werden. Die Säulendurchmesser sind sowohl an der Deckfläche als auch am Boden fest vorgegeben, weshalb der Wandwinkel der Säulen von ihrer Höhe abhängig ist.

Folgende Geometrieparameter wurden für die Durchführung der Variantensimulationen festgelegt: Betrachtet werden sollen Säulenhöhen

von 20  $\mu\text{m}$ , 60  $\mu\text{m}$  und 100  $\mu\text{m}$  und Säulenabstände von 25  $\mu\text{m}$ , 50  $\mu\text{m}$  und 75  $\mu\text{m}$ . Für diese Geometrien wurden Simulationen der einphasigen Strömung in Abhängigkeit vom anliegenden Differenzdruck einerseits für Luft und andererseits für die PEG-Testflüssigkeit durchgeführt (s. Abb.11).



**Abb.11: Variation des Säulenabstandes bei einer Säulenhöhe von 100  $\mu\text{m}$**

Die Simulationen der einphasigen Strömung für neun Varianten (je Variante müssen mehrere Differenzdrücke simuliert werden) dauerten ca. 2-2½ Tage. Bei den entsprechenden zweiphasigen Simulationen muss pro Variante zwar nur eine Simulation durchgeführt werden, jedoch handelt es sich dabei um eine transiente Simulation, sodass viele Zeitschritte berechnet werden müssen, womit eine Simulation mehrere Stunden dauert.

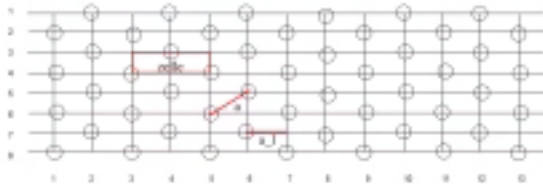
Aus den Variantensimulationen der zweiphasigen Strömung, die ebenfalls für innere wie auch für Randzellen durchgeführt werden, wird der in den entsprechenden Kanalstrukturen vorliegende Kapillardruck bestimmt, der die Triebkraft für das Hineinströmen der Testflüssigkeit in den Kanal darstellt.

## Verhaltensmodell

Aus den im vorigen Abschnitt beschriebenen FVM-Simulationen wurde ein Durchflussmodell für die gesamte Teststruktur erstellt. Nach guter Übereinstimmung von Theorie und Messung für den Durchfluss durch unstrukturierte Kanäle erfolgte die Übertragung der Simulationsergebnisse für Teilstrukturen in ein Gesamtmodell.

## 1.6 Modellierungsansatz

Zunächst wurde die einphasige Strömung betrachtet, die durch eine Druckdifferenz über dem Kanal hervorgerufen wird und bei welcher der Kanal vollständig mit dem strömenden Medium ausgefüllt ist. Die Unterteilung des Kanals gemäß den simulierten Zellen resultierte in einer Struktur entsprechend Abb. 12.



**Abb.12: Kanalstruktur aus Fließzellen**

Die Anzahl der Zellen variiert mit den Parameterwerten der jeweiligen Kanalgeometrie und dem Abstand der Säulenmittelpunkte. Die Länge einer Zelle ergibt sich beispielsweise aus der Projektion des Säulenabstands in Fließrichtung, d.h.:

$$a_l = \sin(\pi/3) \cdot a.$$

Entsprechend gilt für die Anzahl der Fließzellen im Kanal (s. Abb.4):

$$n_l = \lfloor l_{kanal} / 2a_l \rfloor \quad \text{in Längsrichtung und}$$

$$n_q = \lfloor b_{kanal} / (a/2) \rfloor \quad \text{in Querrichtung.}$$

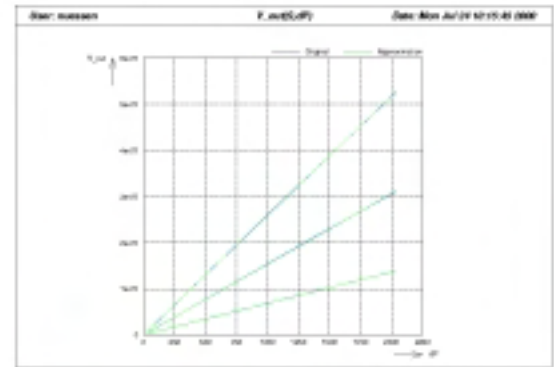
## 1.7 Strömungswiderstand

Die Ergebnisse der FVM-Variantensimulationen liefern den Zusammenhang zwischen Druck und Durchflussmenge für jede Zelle. Diese Kennlinie lässt sich mit dem Approximations- und Synthese-Werkzeug *MagStatic* in eine funktionale Beziehung überführen (Abb.13) [4].

Hierbei erfolgt die Nachbildung des Strömungsverhaltens durch eine frei wählbare Zielfunktion mit den  $n$  anzupassenden Parametern  $x_0 \dots x_{n-1}$ , beispielsweise:

$$\text{Vol} = (x_0 + x_1 \cdot h + x_2 \cdot h^2) + (x_3 + x_4 \cdot h + x_5 \cdot h^2) \cdot p + (x_6 + x_7 \cdot h + x_8 \cdot h^2) \cdot p^2.$$

Die Funktion liefert eine Beziehung zweiter Ordnung hinsichtlich der Kanalhöhe  $h$  und des Außendruckes  $p$ .



**Abb.13: Approximation der Durchflusskennlinien**

Der auf diese Weise ermittelte Strömungswiderstand wird jeder einzelnen Zelle zugewiesen. Dies erlaubt eine Berechnung des Gesamtwiderstandes für den spezifizierten Kanal mit

$$R_{Zelle} = \frac{R_{ges} \cdot n_q}{n_l / 2}, \quad R_{ges} = \frac{R_{Zelle} \cdot n_l / 2}{n_q}.$$

Im Verhaltensmodell wurde die Berechnung der Kanalstruktur aus den Geometrieparametern automatisiert. Widerstandskennlinien, Struktur- und FVM-Simulationsdaten werden in Dateien abgelegt werden. Diese modulare Handhabung der dem Verhaltensmodell zugrundeliegenden Informationen ermöglicht hohe Flexibilität und große Übersichtlichkeit bei der Simulation spezifischer Kanäle, besonders im Hinblick auf die Optimierung.

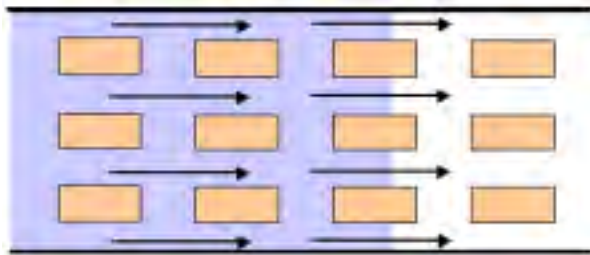
## 1.8 Dynamisches Verhalten

Für den Einsatzbereich der Mikrokanäle ist das zeitliche Verhalten von besonderem Interesse, da die Zeiträume für chemische Ablösevorgänge und Reaktionen eingehalten werden müssen. Um die Befüllungszeit berechnen zu können, beinhaltet das Modell eine transiente Ablaufsteuerung.

Im Hinblick auf den zweiphasigen Strömungsfall, den die Befüllung eines luftgefüllten Kanals mit einem anderen Fließmedium darstellt, werden zunächst die Strömungswiderstände der einzelnen Zellen bestimmt. Aufgrund der Abhängigkeit vom umgebenden Fluid sind zwei Fälle zu unterscheiden: zunächst gilt für alle Zellen der Luftwiderstand, nach der Flutung einer Zelle

ändert sich der Zellwiderstand zum Durchflusswiderstand für den einphasigen Strömungsfall. Beide Widerstände sind aus FVM-Simulationen bekannt. Vom Standpunkt der Simulation aus gesehen, bedeutet das Vorschreiten der Fließfront somit eine dynamische Änderung der Modellstruktur bzw. -instantiierung während der Analyse.

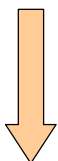
Die maximale räumliche Auflösung des Modells beträgt eine Widerstandszelle, d.h. der Befüllungsvorgang der einzelnen simulierten Zellen wird nicht betrachtet. Zunächst wird von einer geraden Fließfront ausgegangen, d.h. der Fluss ist konstant über der Breite des Kanals. Zu jedem Zeitschritt gilt gemäß Abb. 14 für die Position bis zur Mediengrenze der Durchflusswiderstand, darüberhinaus der Widerstand für unbefüllte Zellen.



**Abb.14: Befüllungstand der Zellen**

Aus der Position der Fließfront ergibt sich über den aktuell gültigen Gesamtwiderstand der Fluss in  $m^3/s$ . Mit dem Strömungsraum im Kanal resultiert aus der Projektion des in Flussrichtung strömenden Volumens zu jedem Zeitschritt in  $m/s$  die neue Position der Fließfront, welche ihrerseits den Ausgangspunkt einer iterativen Durchflussbestimmung darstellt. Ist bei einer neu berechneten Position eine weitere Widerstandszeile überschritten, ändert sich über die Zellenwiderstände der Gesamtwiderstand und damit der Fluss. Das nachfolgende Ablaufschema erläutert den Simulationsablauf mit dem das C-Modell beinhaltenden Programm *flowsim*. Die transienten Simulationsparameter für die Zeitschrittsteuerung sind wie die Geometriedaten Bestandteil der durch das Modell lesbaren Input-Dateien.

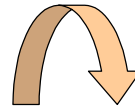
**Datenfile:**



- [ Lese Strukturdaten ];
- [ Lese Druck ];
- [ Lese Simulationszeit & Schritt-

weite ];

**Modellauswertung:**

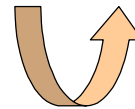


[ Berechne Zellenwiderstand und Geometrie];

[ Start Loop ];

[ Berechne Gesamtwiderstand ];

[ Berechne Fluss ];



[ Berechne Ausbreitung ];

[ End Loop ];

**[ Berechne Gesamtwiderstand ]**

→ Aus allen Zellenwiderständen Gesamtwiderstand (pro Iteration)

**[ Berechne Fluss ]**

→ In Abhängigkeit des Gesamtwiderstandes Fluss ermitteln (pro Iteration)

**[ Berechne Ausbreitung ]**

→ Aus dem Fluss in  $m^3/s$  und Geometrie auf  $m/s$  schließen

→ Berechnung für vorgegebenen Zeitschritt

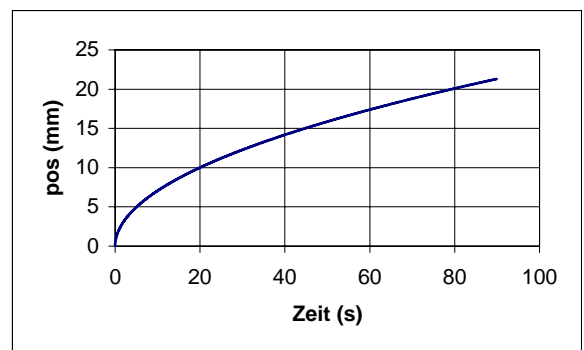
→ Ausgabe Position der Grenze bei Zeitschritt

→ Ermittlung der neuen Grenze Luft/Fluid

→ Merken der Grenze

→ Neuer Gesamtwiderstand, wenn Elementzeile überschritten

Abb.15 zeigt das Simulationsergebnis einer transienten Simulation für einen 21,3 mm langen Kanal. Man erhält die Position der Fließfront und den Fluss als Funktion der Zeit.



**Abb.15: Zeitliches Befüllungsprofil**

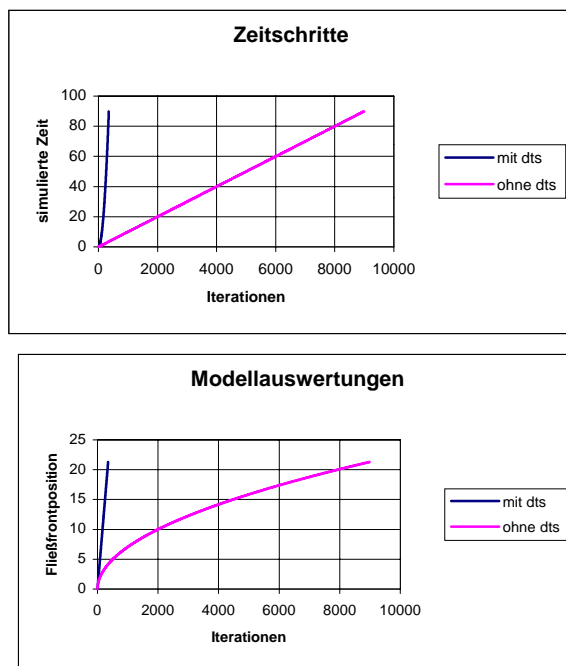
**1.9 Zeitschrittsteuerung**

Simulationen mit dem Verhaltensmodell haben gezeigt, dass die vom Anwender vorzugebende Größe der Zeitschritte erheblichen

Einfluss auf die Rechenleistung und -genauigkeit besitzt. Eine konstante Schrittweite ist zur Vermeidung falscher Werte derart klein zu wählen, dass im Endbereich der Simulationen die Rechenzeit in keinem Verhältnis mehr zur simulierten Zeit steht. Daher wurden in das Modell 3 Varianten aufgenommen:

- Konstante Zeitschrittweite
- Konstante Schrittweite mit Neuberechnung des Kanalwiderstandes nur bei Änderung der Zelleninstantiierung
- Dynamische Zeitschrittweite

Bei der dynamischen Zeitschrittsteuerung (dts) wird das Berechnungsprinzip für das Voranschreiten der Fließfront umgekehrt. Es wird die Fließfrontposition ermittelt, bei der sich das nächste Mal die Zellenzusammensetzung nach luft- oder fluidgefüllten Zellen ändert. Daraus lässt sich über den Volumenstrom die zu simulierende Zeit für das Erreichen der nächsten Zellenreihe zurückrechnen. Den Erfolg der Methode zeigt Abb.16.



**Abb.16: Dynamische Zeitschrittsteuerung**

Im Vergleich mit einer Schrittweite, die zugunsten der Rechenzeit noch kleinere Ungenauigkeiten tolerierte, konnte die Effizienz um drei Größenordnungen gesteigert werden, noch dazu mit vollständig korrekter Berechnung. Dies ist für die Modelloptimierung von wesentlicher Bedeutung.

## 1.10 Kapillareffekt

Die Kapillarwirkung an den Grenzschichten zwischen den beteiligten Medien besitzt einen doppelten Einfluss auf das Fließverhalten.

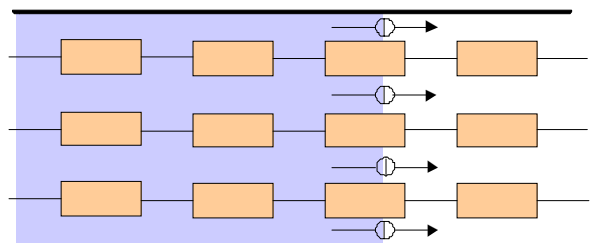
Einerseits ist bei ihrer Einbeziehung die getroffene Vereinfachung einer geraden Fließfront nicht mehr haltbar, da sich unter Einwirkung der Kapillarkräfte ein Meniskus wie in Abb.17 ausbildet.



**Abb.17: Aufnahme der Fließfront**

Andererseits ist die Kapillarwirkung mit einer Antriebskraft gleichzusetzen, die einen eigenen Beitrag zur Befüllung liefert. Dieses ist nicht zu vernachlässigen, sondern bildet einen zentralen Punkt der Betrachtungen, da die Teststreifen in der Praxis ohne äußere Druckdifferenz betrieben werden.

Für die Modellierung mussten die Kapillarwirkung für die Zellenstrukturen quantifiziert werden, wie in Kapitel 4 dargestellt. Um die simulierten Kapillardrücke mit einer Antriebskraft gleichzusetzen, wurden mit der Fließfront mitwandernde Spannungsquellen wie in Abb.18 eingesetzt, welche den Kapillardruck wiedergeben.

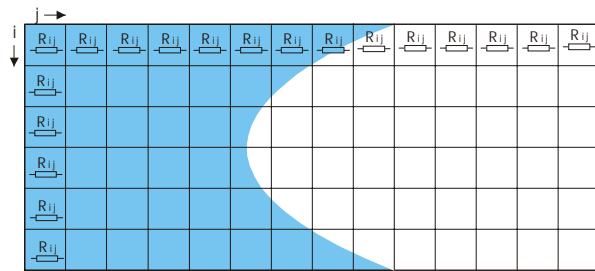


**Abb.18: Kapillarwirkung**

Falls sich bei weiteren Untersuchungen der Kapillardruck noch als ortsabhängig erweist, kann auf diese Weise den Quellen ein positionsabhängiger Wert zugewiesen werden.

## 1.11 Meniskus der Fließfront

Um nicht geradlinig verlaufende Fließfronten modellieren zu können, wurde die Partitionierung der Fließzellen in eine zweidimensionale Matrix überführt. Jede einzelne Zelle kann somit auf ihren Befüllungszustand geprüft werden. Anhand einer vorzugebenden Krümmungsfunktion sowie ihrer zeitlichen Charakteristik wird die Fließfrontposition für jede Matrixzeile separat ausgewertet, bei symmetrischem Verlauf genügt eine Spiegelung der für die halbe Kanalbreite ermittelten Werte.



**Abb.19: Meniskus der Fließfront**

Die Berechnung des Gesamtwiderstandes muss ebenfalls angepasst werden, der Strömungswiderstand des Kanals ergibt sich nun allgemeingültig zu

$$R_{ges}^{-1} = \sum_{i=1}^{n-q} \sum_{j=1}^{n-l} (R_{ij})^{-1} .$$

Derzeitige Untersuchungen sollen den Einfluss der Fließfrontform auf die Befüllungsdynamik quantifizieren.

## Optimierung

Die vorhergehenden Erläuterungen, fluidischen und Verhaltenssimulationen führten mittels rechnergestützter, entwurfsbegleitender Modellbildung, Simulation und modellbasierter Optimierung zu einem ersten optimalen Satz an Designparametern.

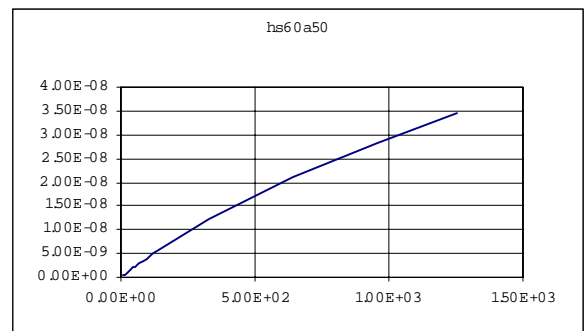
### 1.12 Optimierungsziel

Eine optimale Beschreibung des Strömungsverhaltens innerhalb der beschriebenen Teststrukturen ist erreicht, wenn die Befüllungszeit für eine zuvor definierte Strecke des Kanals über die Änderung von Designparametern eingestellt werden kann. Es wird eine Strecke von 6 mm zwischen der Position 1 mm hinter dem Kanaleinlass und 7 mm festgelegt, die in einer bestimmten Zeit von der Fließfront über-

strichen werden soll. Dies entspricht dem Anspruch an ein definiertes Verhalten im realen Teststreifen, welches für eine gleichmäßige Ablösung von Reagenzien oder eine präzise Einstellung der Zeitspanne für chemische Reaktionen verantwortlich ist.

### 1.13 Lösungsraum der Parameter

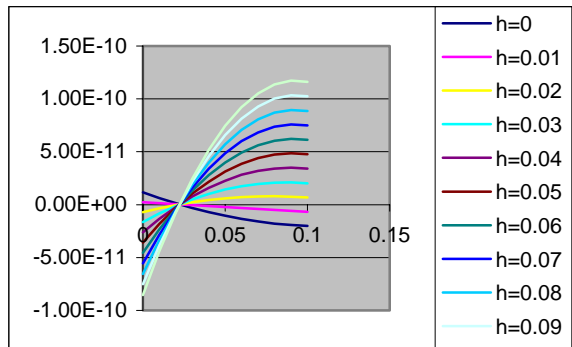
Als Optimierungsparameter fungieren die Strukturhöhe und der Säulenabstand. Da kein prozessbedingter Zusammenhang zwischen den beiden Größen besteht, ergibt sich ein vierdimensionaler Zusammenhang für Strukturhöhe, Säulenabstand, Druck und Volumenstrom, der den jeweils gültigen Arbeitspunkt bestimmen soll. Eine geschlossene Zielfunktion ist hierfür aus der Anschauung nicht mehr zu definieren. Grundlage der Betrachtung sind die bereits in Kapitel 4 dargestellten Variantensimulationen für unterschiedliche Geometrien, d.h. für unterschiedliche Werte der genannten Optimierungsparameter. Die Überführung der Simulationsergebnisse in eine analytische Beziehung erfolgt zunächst anhand der einzelnen Parameterkurven, wie in Abb.20 für den Fluss bei einer Strukturhöhe von 60  $\mu\text{m}$  und einen Säulenabstand von 50  $\mu\text{m}$  dargestellt.



**Abb.20: Variantensimulation-Einzelkurve**

Für die Einzelkurven wurde ein quadratischer Zusammenhang angenommen, was sich als ausreichend genau erwiesen hat. Die bei gleicher Zielfunktion gefitteten freien Parameter für alle Einzelkurven wurden in Abhängigkeit der bei den FVM-Simulationen variierten Parameter gesetzt und ihrerseits mit einer komplexeren Zielfunktion approximiert. Den so gefundenen Zusammenhang für jeden Fitparameter der ersten Fitfunktion, wie er in Abb.21 exemplarisch für einen beliebigen Parametersatz dargestellt ist, kann man wieder-

um in selbige Funktion einsetzen und erhält auf diese Weise einen geschlossenen vierdimensionalen Ausdruck über den gesamten Lösungsraum der Optimierungsparameter.



**Abb.21: Fitparameter als  $f(hs,a)$**

Durch Reproduktion der simulierten Einzelkurven sowie durch Nachweis der Stetigkeit im übrigen Raumgebiet wurde die Gültigkeit des gefundenen analytischen Ausdrucks verifiziert.

### 1.14 Zeitmessung

Jede Aussage über die Befüllungszeit der im Modell frei wählbaren Teststrecke stützt sich auf eine genaue Erfassung von simulierter Zeit und erreichter Fließfrontposition. Dies beinhaltet gleich zwei Probleme: einerseits ist die räumliche Auflösung bei fester Zeitschrittweite abhängig vom jeweiligen Fluss, erhöht sich also mit zunehmender Befüllung des Kanals aufgrund des sich erhöhenden Kanalwiderstandes und der einhergehenden Reduzierung der Fließfrontgeschwindigkeit. Andererseits korreliert die zeitliche Auflösung im Falle der im Kapitel 5.4. dargestellten dynamischen Zeitschrittsteuerung mit der räumlichen Diskretisierung, beträgt also einen ‚zeitlichen Zellenabstand‘, da hier mit einer konstanten räumlichen Schrittweite gerechnet wird. Bei letzterer Variante ist die Genauigkeit mit Sicherheit, bei der ersten unter Umständen nicht mehr ausreichend für eine Bewertung im Sinne einer Optimierungsrechnung. Daher wurde ein linearer Interpolationsansatz zwischen den räumlichen und zeitlichen Simulationsergebnissen implementiert, welche die Start- und Endpositionen der Zeitmessung umgeben. Somit werden die genauen Zeit- und Positionsmarken für die Teststrecke ausreichend genau approximiert:

$$x_t = \frac{x_2 - x_1}{y_2 - y_1} (pos - y_1) + x_1$$

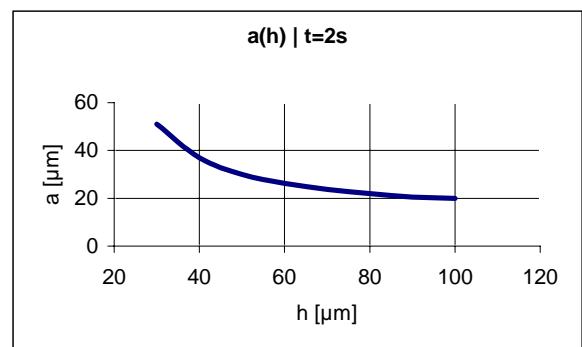
mit:

- $x_1, x_2, y_1, y_2$ : Zeit- und Positionswerte der um die Messposition liegenden Simulationsergebnisse
- $pos$ : Start- oder Endposition für die Zeitmessung, bspw. 1 mm oder 7 mm
- $x_t$ : für Position  $pos$  gültiger Zeitwert

Der verbleibende Fehler der linearen Interpolation liegt bei einem näherungsweise exponentiellen Befüllungsverlauf wie in Abb.15 im Sub-Promill-Bereich und erlaubt eine eindeutige Bewertung der Optimierungsparameter.

### 1.15 Optimierungsergebnis

Eine Besonderheit des vorliegenden Optimierungsproblems besteht in der Wirkungsweise der Optimierungsparameter Strukturhöhe und Säulenabstand. Je nach geforderter Befüllungszeit als Vorgabe des Optimierungszieles gibt es einen mehr oder weniger großen Lösungsraum an Arbeitspunkten, in dem sich für jeden Wert des einen Parameters ein Wert des anderen finden lässt, um die gleiche vorgegebene Befüllungszeit für eine vorgegebene Teststrecke zu erreichen. Da sich die Wirkungen der Parameteränderungen gegenseitig beeinflussen, ist das Optimierungsergebnis unendlich vieldeutig, d.h. es ergibt sich kein optimaler Arbeitspunkt, sondern ein als optimal ermittelter Zusammenhang zwischen den beiden Parametern, wie in Abb.22 gezeigt.



**Abb.22: Optimierte Parameter a und h**

Für eine frei gewählte Zeit von 2 Sekunden für die Teststrecke 1 mm bis 7 mm ergibt sich die obige Zuordnung von Wertepaaren. Qua-

litativ ist nachvollziehbar, dass sich bei einer großen Strukturhöhe der Kanalwiderstand verringert, was durch enger stehende Säulen, also einem kleineren Säulenabstand ausgeglichen wird, um die gleiche Befüllungszeit, d.h. den gleichen Gesamtfluss zu erhalten und umgekehrt.

Eine wesentliche Rolle kommt dem ebenfalls geometrieabhängigen Kapillardruck zu, der sich noch stärker auf das Endergebnis auswirkt als die Optimierungsparameter. Aus diesem Grund wurden die in Kapitel 4.2 erläuterten zweiphasigen Simulationen ebenfalls als Variantensimulation in Abhängigkeit der Optimierungsparameter durchgeführt und lieferten Werte für die Kapillarwirkung als essentielle Eingangsgröße des Modells. Da der Befüllungsverlauf einer einzelnen Zelle im Gegensatz zu den wirkenden Kapillarkräften keine für das Modell verwertbaren Informationen beinhaltet, können die FVM-Ergebnisse für die zweiphasigen Variantensimulationen in einen dreidimensionalen analytischen Zusammenhang für den gesamten Lösungsraum überführt werden und liefern somit für jede Kombination der Optimierungsparameter einen gültigen Arbeitspunkt des Modells. Die übrigen geometrischen Abmessungen des Kanals sind ebenfalls variabel, bleiben aber für die Optimierung weitgehend ohne Einfluss. Da der fluidische Widerstand einer luftgefüllten Zelle um etwa zwei Größenordnungen unter dem einer mit der Testflüssigkeit gefüllten Zelle liegt, dominieren die Fluidzellen die Befüllungszeit, sobald die Fließfrontposition die erste Zellenspalte erreicht hat, nahezu unabhängig davon, wieviele luftgefüllte Zellen im verbleibenden Kanalraum noch folgen. Eine Änderung der Kanalbreite wird durch den sich ergebenden Fluss kompensiert, so dass sich die Befüllungszeit ebenfalls nicht merklich ändert.

Da es für eine zu formulierende Zielfunktion kein eindeutiges Minimum gibt, werden zunächst eindimensionale Optimierungsansätze verfolgt, bei denen für feste Werte der Strukturhöhe der Säulenabstand dem Zeitkriterium entsprechend eingestellt wird. Für die Aufnahme einer Kurve wie in Abb.22 wird somit je ein Optimierungslauf pro Stützstelle benötigt.

## Zusammenfassung

Auf der Basis fluidischer FVM-Simulationen, d.h. der Lösung partieller Differentialgleichungen, ist ein Verhaltensmodell für die Befüllung strukturierter Mikrokanäle als wesentlicher Bestandteil medizinischer Teststreifen entwickelt worden. Die Teilergebnisse für partielle Kanalbereiche wurden in Abhängigkeit der gewählten Optimierungsparameter Strukturhöhe und Säulenabstand aufgenommen und der Systemsymmetrie entsprechend zur Berechnung eines Gesamtkanals herangezogen. Die diskreten einzelnen Simulationsergebnisse konnten mit Hilfe numerischer Verfahren durch geschlossene analytische Ausdrücke wiedergegeben werden, welche einen kontinuierlichen Lösungsraum für die Optimierung eröffneten und somit für jede Kombination der Eingangsgrößen einen gültigen Arbeitspunkt der Verhaltenssimulationen definierten. Davon ausgehend wurde als Optimierungsergebnis für eine exemplarisch definierte Befüllungszeit ein eindeutiger Zusammenhang zwischen den Optimierungsparametern gefunden. Der Industriepartner kann somit einen Optimierungsparameter innerhalb des gültigen Arbeitsbereiches vorgeben und weiß, wie der andere für eine definierte Befüllungszeit zu wählen ist.

## Literatur

- [1] O. Nüssen, H. Bolte, D. Peters, S. Bechtold, R. Laur: An application specific design methodology for microsystems, DTIP2001, Cannes, Proceedings pp. 175-186
- [2] C. Schön, I. Ballhorn: Betrachtung des Flüssigkeitsverhaltens in medizinischen Teststreifen aus Kunststoff. Abschlußseminar OMID, Nov. 2002, Karlsruhe
- [3] S. Halstenberg, A. Quinte, H. Eggert, C. Schön, R. Peters: Einsatz numerischer Verfahren zur Modellbildung und Simulation von kapillarisch getriebenen Fluiden in Mikrostrukturen. 4. Statuskolloquium des Programms Mikrosystemtechnik, 30./31. März 2000, Forschungszentrum Karlsruhe, FZKA 6423, S. 143-148
- [4] M. Anton, St. Bechtold, O. Nüssen, R. Laur: Computer aided modeling of static and dynamic transfer characteristics, Technical Proceedings of MSM98, April 6-8, 1998, Santa Clara, California
- [5] O. Nüssen, H. Bolte, D. Peters, R. Laur, S. Halstenberg, H. Eggert: Behavioural Modelling of Medical Test Strips for Optimization, DTIP 2002, May 2002, Cannes, Proceedings pp. 385-393



# DIE AKTORPLATTE – EIN UNIVERSELLER FLUIDISCHER MIKROAKTOR

Dipl.-Ing. Isam Tahhan, F&E, Bartels Mikrotechnik GmbH

Email: tahhan@bartels-mikrotechnik.de

## Kurzfassung

Das Konzept eines Mikroaktors, welcher fluidisch betrieben und elektrisch gesteuert werden kann, wird vorgestellt. Der prinzipielle Herstellungsweg des Aktors und des zugehörigen Ventils sowie unterschiedliche Betriebsarten werden dargelegt, und potenzielle Anwendungsmöglichkeiten aufgezeigt.

## 1 Einleitung

Die Mikrotechnik beschäftigt sich seit geraumer Zeit und mit wachsendem Einsatz an der Realisierung einzelner Funktionen für fluidische Bauelemente sowie fluidische Gesamtsysteme. Dabei besteht insbesondere an Pumpen, Ventilen und Aktoren reges Interesse, da diesen aktiven Elementen eine Schlüsselrolle bei dem Aufbau von Systemen zukommt.

Dringend benötigt werden hier Aktorelemente, die sich kostengünstig und in großen Flächen innerhalb eines 2-D-Arrays realisieren lassen.

Solche Aktorarrays könnten für viele Applikationen eingesetzt werden, so z.B.:

- taktiles Array für die Realisierung von Blindenschrift, der Bewegung von kleinen Objekten etc.
- hochpoliges flächenhaftes flexibles elektrisches 2-D-Relais
- Aktoren zum Antrieb von Ventilarrays
- Verkippung der Aktoren, so dass damit eine Spiegelverstellung möglich wird.

Ein effizientes Konzept zur Verwirklichung solcher Arrays stellt die sog. „Aktorplatte“ dar. Diese besteht im wesentlichen aus einer mikrostrukturierten Membran, die sich über einem Hohlraum befindet. Bei Druckbeaufschlagung wölbt sich diese Membran nach außen (bzw. bei Unterdruck

nach innen), so dass die daraus resultierende Bewegung nutzbar gemacht werden kann.

## 2 Prinzipien

Die folgenden Skizzen verdeutlichen den Aufbau und die Wirkungsweise der Aktorplatte und des dazugehörigen Ventils.

### 2.1 Die Aktorplatte

Die Aktorplatte selber besteht im wesentlichen aus einer Kunststoffmembran, die durch Excimerlaserstrukturierung an bestimmten, genau definierten Stellen ausgedünnt ist. Als Ergebnis dieser Ausdünnung erreicht man eine maximale Auslenkung bei minimalem Druckabfall unter der Platte und minimalen Spannungen, d.h., die Platte ist soweit als möglich strukturoptimiert.

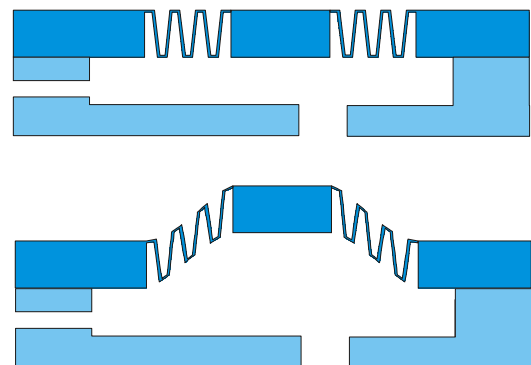


Abbildung 1: Funktionsweise der Aktorplatte

Während die Aktorplatte im drucklosen Zustand eben mit der Oberfläche abschließt (s. oben), hebt sie sich bei Überdruck ein gewisses Maß aus der Oberfläche heraus (darunter). Bei Unterdruck würde das entstehende Gegenteil gelten.

### 2.2 Das Ventil

Zum Betrieb der Aktorplatte ist ein Ventil nötig, welches dafür sorgt, daß im Falle einer gewünschten Auslenkung den Druck in der

Hohlkammer unter der Platte ansteigen läßt. Dieses Ventil wird mittels elektro-rheologischen Flüssigkeiten (kurz: ERF) realisiert werden.

### 2.2.1 ERF

ERF sind Flüssigkeiten, die ihre Viskosität unter Einfluß eines elektrischen Feldes ändern. Im Extremfall kann es so zu einer Quasi-Verfestigung kommen, aus der ein Erliegen des Flüssigkeitsstromes resultiert.

### 2.2.2 Betrieb

Die folgende Abbildung zeigt schematisch das Ventil in Verbindung mit einer ruhenden bzw. ausgelenkten Aktorplatte.

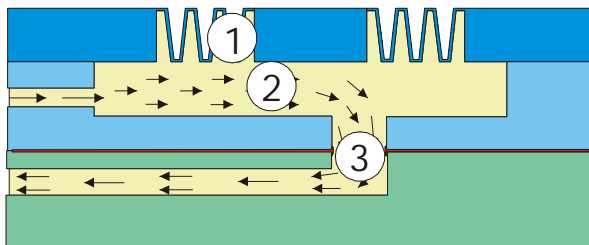


Abbildung 2: Ruhezustand, Ventil offen

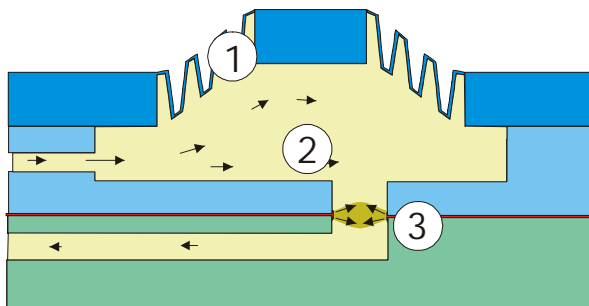


Abbildung 3: Ausgelenkter Zustand

(1) ist die Aktorplatte, (2) der ERF-Flüssigkeitsstrom, und (3) das ERF-Ventil. Im Normalfall strömt die ERF ungehindert unter der Aktorplatte hindurch und verläßt die Hohlkammer durch das offene Ventil. Wird das Ventil jedoch durch Anlegen einer Spannung und somit der Erzeugung eines elektrischen Feldes aktiv, so entsteht dort ein Propf, der zu einer Druckerhöhung in der Hohlkammer führt: die Platte wird ausgelenkt.

### 2.2.3 Druckquellen

Als Drucklieferanten können unterschiedliche Quellen dienen. Für nicht raumkritische

Anwendungen können normale Pumpen eingesetzt werden, die allerdings (bei Einsatz partikelhaltiger ERF) nicht verstopfen dürfen (Membranpumpen). Hier sind Drücke bis mehrere Bar problemlos erzeugbar. Im Falle eines beschränkten Raumangebots können Gaskartuschen, die über eine Mechanik ein ERF-Reservoir mit Druck beaufschlagen, genutzt werden, allerdings mit dem Nachteil, daß nach Versiegen des Kartuschendruckes auch der ERF-Strom zum Erliegen kommt. Schließlich können für besondere Anwendungen auch Mikropumpen eingesetzt werden, die allerdings nur Drücke bis ca. 0,1 bar Überdruck erzeugen, was die Anwendungsbandbreite etwas einschränkt.

## 2.3 Aufbauvarianten

Je nach Anwendungsfall kann zwischen den Aufbauvarianten „Ventil hinter der Platte“ und „Ventil hinter und vor der Platte“ ausgewählt werden.

## 3 Variationsmöglichkeiten

Ein kurzer Einblick ins Innere von Platte und Ventil soll die Einfachheit der dahinter stehenden Idee, sowie die Gründe für die Robustheit der Lösung illustrieren.

### 3.1 Aktorplatte

#### 3.1.1 Material

Die Herstellung der Aktorplatte erfolgt vornehmlich in Folientechnik. Das bedeutet, daß man auf Halbzeug hoher Güte, geringen Preises und sehr guter Verfügbarkeit zurückgreifen kann, welches in unterschiedlichsten Materialien und Dicken vorliegt. Je nach Anwendungsfall kann so eine optimale Material- und Geometriewahl getroffen werden.

#### 3.1.2 Strukturierung

Für kleine Stückzahlen erfolgt die Strukturierung der Folie (z.B. Polycarbonat) mittels Excimerlaser-Strahlung. Die dabei erzielbaren Genauigkeiten von  $\pm 5 \mu\text{m}$  sind ausreichend für alle hier diskutierten Anwendungen. Für große Stückzahlen ist die Prägetechnik nutzbar.

### 3.1.3 unterschiedliche Plattenformen

Im allgemeinen ist die Platte als kreisrunder Aktor ausgeführt, da sich hier die besten Ergebnisse bezüglich Auslenkung und Spannungsverteilung erzielen lassen. Es sind aber auch Sonderformen denkbar, in denen z.B. mittels einer Keilform eine eindeutige Vorzugsrichtung beim Auslenken erzeugt wird.

### 3.1.4 Funktionsplattform

Die Funktionsplattform kann einerseits ohne weitere Strukturierung oder Behandlung direkt zur Übertragung beispielsweise haptischer Information genutzt werden. Sie kann jedoch auch durch entsprechende Metallisierung elektrischen Strom leiten, so daß sie Teil eines Relais wird, oder durch Verspiegeln zum Ablenken eines Strahles genutzt werden.

## 3.2 Ventil

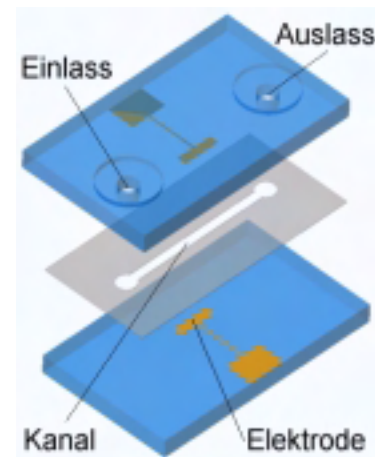
Das ERF-Ventil besteht aus Gründen der Wirtschaftlichkeit und Integrationsfähigkeit aus den gleichen Materialien wie die Aktorplatte. Daher kommen auch hier Folien bzw. Platten aus Kunststoff zum Einsatz. Üblicherweise befinden sich im Unterbau der Platte ohnehin die Kanäle für die ERF, welche mit unterschiedlichen Verfahren an entsprechenden Stellen zu metallisieren sind, um das nötige elektrische Feld zu erzeugen. Mögliche Verfahren hierzu sind:

- Einlegeteile aus dünnem Metall, die mittels feiner Drähte nach außen kontaktiert werden.
- Direktmetallisierung auf die Kanalflächen.

Folgendes Bild zeigt den Aufbau eines getesteten ERF-Ventils.

Das Ventil besteht aus zwei Polycarbonat-Platten, auf welche die Elektroden, deren Zuleitungen und außenliegende Kontaktierungsflächen aufgebracht werden. Zwischen die Platten wird eine dünne geschlitzte Membran gelegt.

Der Schlitz dient als Flüssigkeitskanal; die Höhe des Kanals entspricht der Dicke der Membran und ist somit genau einstellbar.



**Abbildung 4: Schema des ERF-Testventils**

Mittels der Ein- und Auslaßöffnungen wird elektrorheologische Flüssigkeit durch den Kanal gepumpt. Über die Außenkontakte kann der Flüssigkeitsstrom je nach Druck und Spannung vermindert oder auch völlig gestoppt werden.

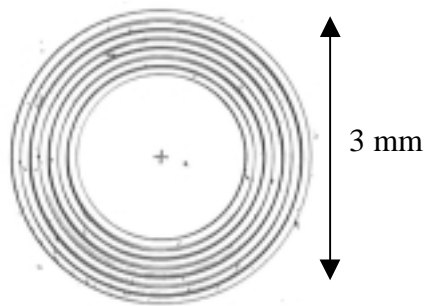
## 4 Realisierte Prototypen der Aktorplatte

Im Rahmen des Förderprojektes OMID konnten bereits mehrere Prototypen der Aktorplatte sowie das dazugehörige Ventil hergestellt werden. Eine Integration von Aktorplatte und Ventil in ein Bauteil wurde aus Zeitgründen noch nicht realisiert; das Zusammenspiel beider Komponenten konnte jedoch mittels Verbindung durch Schlauchverbindungen erfolgreich demonstriert werden.

### 4.1 Vorversionen

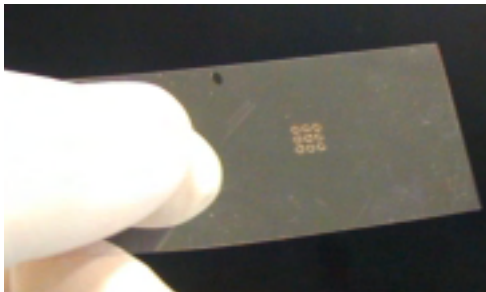
Im Laufe des Projektes wurden verschiedene Versionen von Aktorplatten aufgebaut und charakterisiert.

Die nächste Abbildung zeigt eine Aktorplatte aus Polycarbonat mit dem Durchmesser 3,2 mm. Im Zentrum ist die große Funktionsfläche zu sehen, die kreisrunden Kerben sind noch äquidistant.



**Abbildung 5: Einfache Aktorplattenmembran (Lichtmikroskopaufnahme)**

Auch ein 3x3-Array konnte hergestellt werden. Die einzelnen Platten haben einen Durchmesser von ca. 1 mm.



**Abbildung 6: 3x3-Membran-Array**

## 4.2 Prüfaufbau

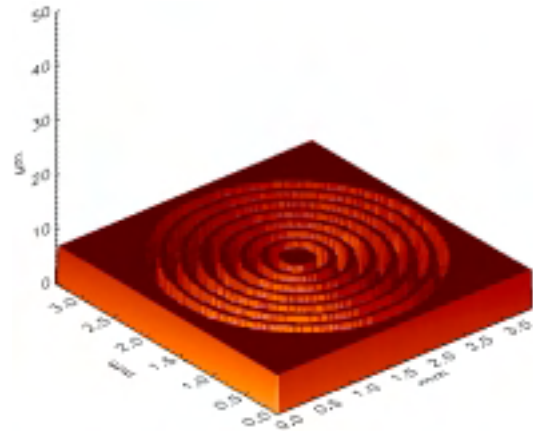
Zur Prüfung der mechanischen Eigenschaften der Platten wurde ein einfacher Testaufbau realisiert. Die mit einer externen, regelbaren Druckquelle verbundenen Platten wurden schrittweise mit (Luft-)Druck beaufschlagt und der sich daraus ergebende Hub gemessen. Die Abbildung zeigt den Halter, in dem sich zwei Platten auf einem Polycarbonat-Träger befinden. Die Auslenkungen wurden mittels eines Oberflächenmeßgerätes bestimmt.



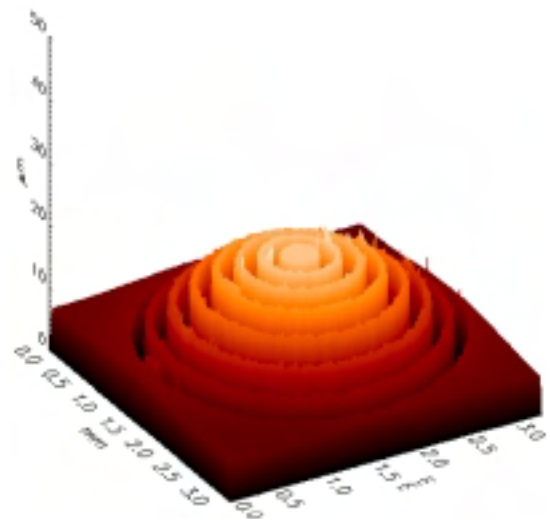
**Abbildung 7: Versuchsaufbau**

Die Ergebnisse des 3-D-Scans über die

ruhende und ausgelenkte Aktorplatte zeigen die folgenden beiden Abbildungen. Die Auslenkung dieses Prototypen beträgt noch ca. 20  $\mu\text{m}$ , je nach Bauart und Betriebsdruck sind aber deutlich größere Hübe realisierbar.



**Abbildung 8: Aktorplatte im Ruhezustand**



**Abbildung 9: Aktorplatte ausgelenkt**

## 4.3 Optimierte Version

Am Forschungszentrum Karlsruhe wurden extensive Untersuchungen zur Optimierung des Plattenprofils betrieben. Die festen Randbedingungen betrafen die Gesamt-Plattendicke, das Materialparameter, den Gesamtdurchmesser, die Größe der Funktionsplattform und die minimale und maximale Kerbenanzahl sowie die maximale Kerbtiefe. Variiert werden konnten die Lage und Tiefe der Kerben. Als Optimierungsziel wurde eine möglichst große Auslenkung bei gleichem Druck und unter

Einhaltung der maximalen Vergleichsspannung angegeben.

#### 4.3.1 optimiertes Profil

Aufgrund der vom FZK gelieferten Optimierungsergebnisse wurde eine entsprechende Platte gefertigt. Als hervorstechendes Merkmal besitzt diese Platte lediglich je eine relativ breite Kerbe auf der Ober- und der Unterseite. Das Profil welches je nach gewünschtem Maximaldruck und E-Modul leicht variiert, zeigt nachstehende Abbildung.

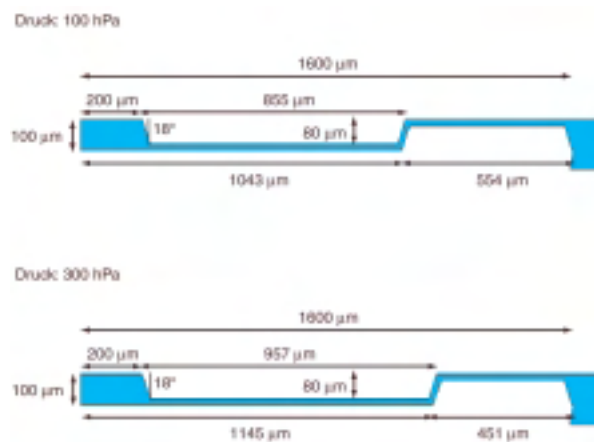


Abbildung 10: optimierte Profile

#### 4.3.2 Druckverlauf

Das optimierte Profil wurde mit Drücken bis ca. 1 bar belastet. Die sich daraus ergebenden Verformungen sind im nachstehenden Graphen aufgetragen. Zusätzlich sind die Daten der entsprechenden Simulation angegeben, welche mit den Original-Profildaten durchgeführt wurde. Die Abweichung der beiden Kurven liegt bei unter 10%. Die eigentlichen Hübe stimmen jeweils gut überein, und auch deren schwächeres Anwachsen bei höheren Drücken läßt sich mittels einer logarithmischen Kurve gut annähern.

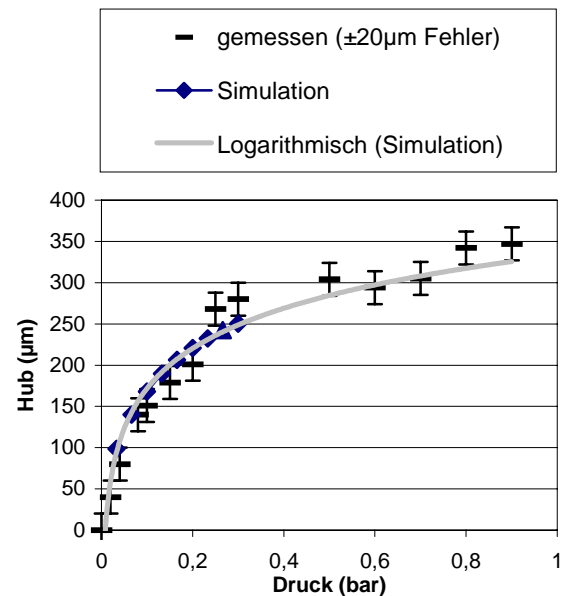


Diagramm 1: Hub der optimierten Aktorplatte (real und simuliert)

#### 4.3.3 Hysterese

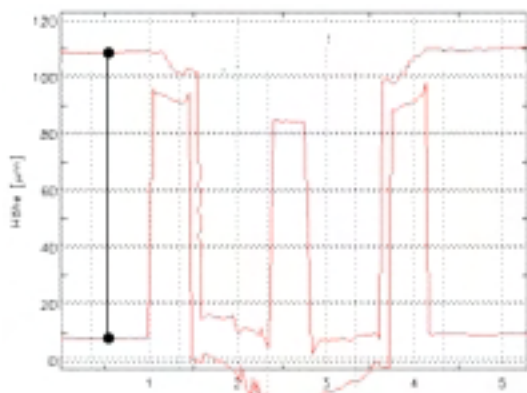
Bei den Messungen wurde eine gewisse Hysterese festgestellt. Diese bewegte sich im Bereich um 5% des Maximalhubes. Auch eine Nullpunktverschiebung in dieser Größenordnung wurde bemerkt. Der Grund hierfür liegt in der teilplastischen Verformbarkeit des Werkstoffes sowie einer teilweise unvollständigen Verklebung von Platte und Körper. Während der ersten Druckbeaufschlagung lösen sich diese im Randbereich befindlichen Teilverklebungen voneinander, so daß sich der Wirkdurchmesser der Platte geringfügig vergrößert. Mittels geeigneter Verbindungstechnik kann letzteres Problem teilweise umgangen werden, wobei eine Resthystere aufgrund des zuvor beschriebenen Grundes nicht vermeidbar sein wird. Dies darf jedoch für die meisten Anwendungen als unproblematisch angesehen werden.

#### 4.3.4 Fertigungsaspekt / reales Profil

Im Rahmen des Projektes wurde festgestellt, daß sich doppelseitige Strukturen, die nahe beieinander liegen (wie es bei den gegenüberliegenden Kerben er Fall ist), nicht ohne weiteres herstellen lassen, wenn der sich bei der Fertigung ergebende Wandwinkel nicht ausreichend berücksichtigt wird. Als

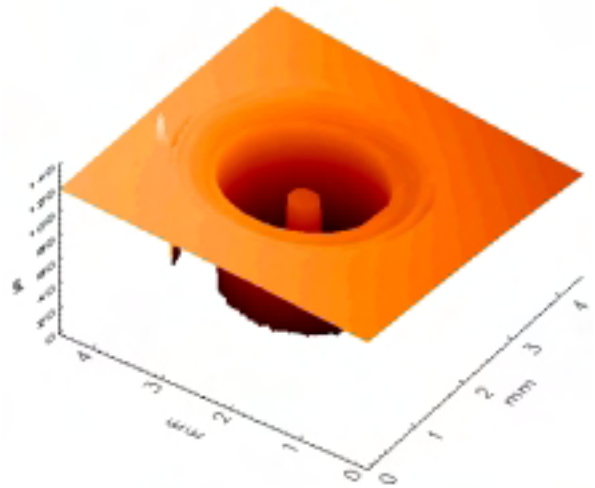
Folge dessen konnten zunächst nur Platten gefertigt werden, bei denen die Kerbtiefe nicht ganz bis zur Materialmitte reichte bzw. der Abstand gegenüberliegender, aneinander angrenzender Kerben (d.h. die Dicke der in axialer Richtung verlaufenden Platten-segmente) sehr groß war.

Umfangreiche Versuche ergaben schließlich entsprechende Richtwerte, die dazu führten, daß die zu Beginn des Projektes anvisierten Maße, nämlich eine Restdicke der Platte von ca. 20  $\mu\text{m}$ , erreicht werden konnten. Das Schnittprofil einer solchen Platte zeigt das nächste Bild.

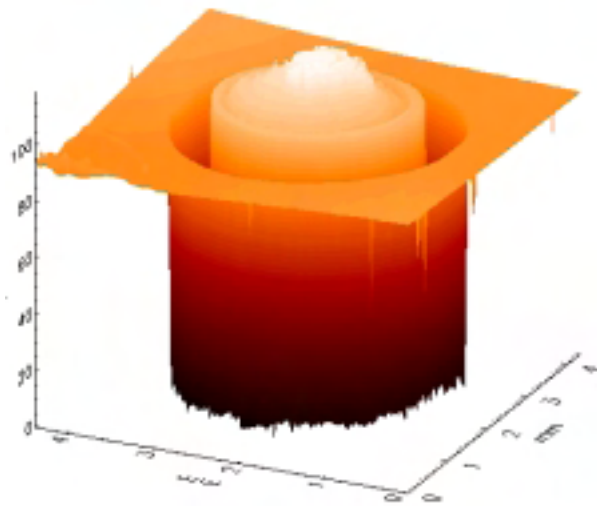


**Abbildung 11: reales Profil der opt. Platte**

Die an den Kantenübergängen sichtbaren Zacken sind weitgehend auf Meßfehler zurückzuführen. Die am linken Bildrand befindliche Maßstab deutet die Materialstärke von 100  $\mu\text{m}$  an. Die Platte scheint nach unten durchgebogen zu sein; dies wird auf materialimmanente Verspannungen zurückgeführt und muss noch Gegenstand weiterer Untersuchungen sein. Bei anderen Materialien traten diese Verformung nicht auf. Die nächsten beiden Abbildungen zeigen die 3D-Ansichten von Ober- und Unterseite der optimierten Platte.



**Abbildung 12: 3D-Ansicht der optimierten Aktorplatte (Oberseite)**



**Abbildung 13: 3D-Ansicht der optimierten Aktorplatte (Unterseite)**

## 5 Aktorplattensystem: Aktor und Ventil

In einem abschließenden Funktionstest wurde die optimierte Aktorplatte mit dem ERF-Ventil verbunden und aktuiert. Abbildung 14 zeigt schematisch den Aufbau der Teststruktur.



**Abbildung 14: Schematischer Systemaufbau**

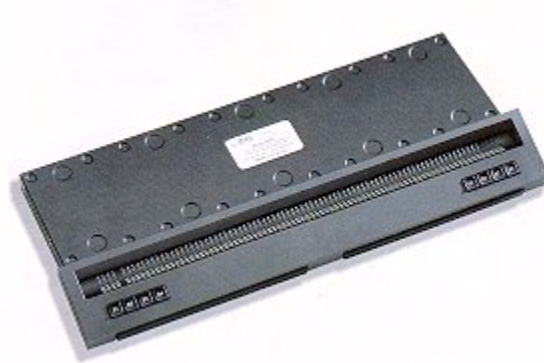
Der Aufbau ähnelt dem des Testventils (s.o.), ist jedoch mit zwei anstatt einer Elektrode bestückt und um eine Aktorplatte, welche sich zwischen den beiden Elektroden befindet, erweitert. Mittels diesem Aufbau wurde erfolgreich nachgewiesen, daß der durch das Ventil schaltbare Druck ausreicht, um die (ebenfalls mit ERF unterspülte) Aktorplatte zu betätigen.

## 6 Anwendungsbeispiele

Aufgrund des einfachen Aufbaues und universell einsetzbaren Prinzips lassen sich eine Vielzahl von Anwendungen finden, in denen die Aktorplatte sinnvoll eingesetzt werden kann. Die folgenden Beispiele geben eine Übersicht.

### 6.1 Braille-Display

Bestes Beispiel ist die fluidisch angesteuerte Braille-Zeile (Blindenschrift-Anzeige). Eine Braille-Zeile ist ein Hilfsmittel für blinde Personen, die mittels dem Ertasten der Blindenschrift lesen können. Die Blindenschrift besteht aus Buchstaben, die sich aus punktförmigen Erhebungen zusammensetzen. Von erhabenen gedruckten Braille-Schriftstücken abgesehen, benötigt ein Blinder eine Anzeige, die elektronisch angesteuert wird und die Darstellung einer oder mehrerer Zeilen in Blindenschrift ermöglicht.



**Abbildung 15: herkömmliche Braille-Zeile mit Tastaturauflage**

Durch die kostengünstige Fertigung eines großen Aktorplatten-Displays könnte einen Quantensprung von den maximal dreizeiligen Liniendisplays hin zu echten „Braille-Bildschirmen“ mit VGA-Auflösung bedeuten.

Wenngleich herkömmliche Braille-Anzeigen einen Hub von 1-2 mm bieten, ist auch bereits deutlich unter 1 mm eine Erkennung der Schriftsymbole möglich, wenn eine entsprechende Scharfkantigkeit der Pixel gewährleistet ist. Dies kann mittels einer entsprechenden nadelförmigen Erhebung in der Mitte der Platte erreicht werden.

### 6.2 hochohmiges Relais

Eine weitere Anwendung ist die Herstellung großflächiger Relais, die zur wahlweisen Kontaktierung bzw. Trennung von Flip-Chips eingesetzt werden könnten. Als besonders vorteilhaft ist hier die Integrationsfähigkeit der Aktoren in die Leiterplatte zu sehen.

Bestimmte Kunststoffe besitzen eine (begrenzte) immanente Leitfähigkeit, beispielsweise aufgrund von Kohlenstoff-Einlagerungen (z.B. Pokalon® C), so dass je nach Anwendungsfall eine Nachmetallisierung teilweise oder ganz entfallen kann.

### 6.3 Chip-Kontaktierung

Eine dem Aktorplatten-Relais ähnliche Anwendung ist die schaltbare Flip-Chip-Kontaktierung. Dabei befindet sich unter jedem der Chipkontakte, welche bis zu mehreren hundert betragen können, eine Aktorplatte, die bei Betätigung die Verbindung zur Leiterplatte herstellt.

## 6.4 Kippspiegel

Eine leichte Abwandlung der Konstruktion der Aktorplatte erlaubt eine seitliche Verkippung anstatt dem gleichmäßigen Ansteigen und Absenken der Funktionsfläche. Befindet sich dort eine spiegelnde Schicht, so erhält man einen fluidisch angetriebenen Ablenkspiegel, der bei geeigneter Wahl der Basismaterialien ebenfalls in eine Leiterplatte integriert werden könnte. Daraus resultiert eine platzsparende und kostengünstige Fertigung der Ablenkssysteme, da die Aktorik mit der Leiterplatte in einem Arbeitsgang gefertigt wird, und dabei auch die Einbringung der notwendigen Elektroden zur Steuerung problemlos möglich ist.

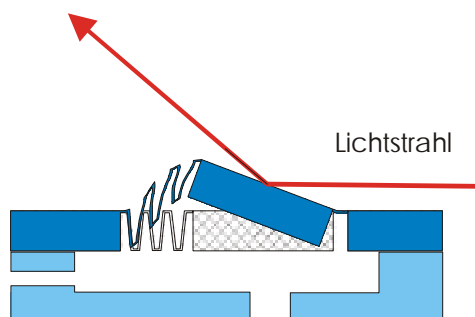


Abbildung 16: Aktorplatte als Ablenkspiegel

## 6.5 Manipulationsfläche

Bei einer genügend dichten flächigen Packung vieler Aktorplatten nebeneinander, ist eine Funktionsfläche denkbar, die durch gezieltes Ansteuern der einzelnen Elemente darauf abgelegte Objekte bewegt. Limitierender Faktor sind hier allerdings die zur Bewegung nötigen recht hohen Kräfte, die entsprechend hohe Anforderungen an die Aktorplatte stellen. Mittels entsprechend dicker Plattenmaterialien und hoher Drücke kann jedoch auch für solche Anwendungen eine geeignete Aktorplatte entworfen werden.

## 7 Zusammenfassung

Die Aktorplatte ist als ein „universelles“ Instrument anzusehen, welches überall dort eingesetzt werden kann, wo Aktorik aufgrund von

- zu hohen Robustheitsanforderungen
- zu hohen Kosten
- zu eingeschränkter Flexibilität

bisher keine Verwendung fand. Im Detail betrachtet, ergeben sich nocheinmal folgende Vorteile:

### 7.1 Vorteile des Konzeptes

- **Serienprodukt:** Da die Aktoren in großer Anzahl parallel und in einem preisgünstigen Verfahren fertigbar sind, liegen die Kosten deutlich geringer als bei vergleichbaren Mikroaktoren anderer Herstellungsweise.
- **Vielfältig:** Das neue Konzept bietet eine große und relativ leicht realisierbare Kombinationsvielfalt bezüglich der am Aktor angekoppelten Zusatzfunktionen, so dass hier der Gedanke eines echten modularen Mikrosystems zum Tragen kommt.
- **Preisgünstig:** Das Konzept ist kostengünstig, weil zum einen die Materialien (Kunststoffe), aber auch die Fertigungsverfahren (Lasertechnik in der Prototypenphase, Prägetechnik in der Serienproduktion) vergleichsweise geringe Kosten verursachen.

### 7.2 Vorteile des Aktors

- **Robustheit:** Der Aktor ist deutlich robuster, weil die Materialien nicht brüchig (Silizium o.ä.) und trotzdem chemisch ausreichend resistent sind (je nach Materialwahl).
- **Hohe Kraft:** Er ist stärker, weil die krafterzeugende Quelle sich nicht mehr direkt am Ort des Geschehens lokalisiert sein muss, sondern entfernt vom Wirkort befinden kann. So kann auf „makroskopische“ Druckquellen (Standard-Pumpen) ausgewichen werden, wenn dies nötig ist. Zudem hält der Baukasten der Mikrotechnik (MATCH-X) mittelfristig auch eine Mikropumpe bereit, welche direkt in modulare Mikrosysteme integriert werden kann. Auch der Aktor selber ist



leicht in einen MATCH-X-Baustein integrierbar und erweitert so den bestehenden Baukasten um ein wichtiges Element.

- **Modulares Konzept:** Der Aktor ist so flexibel, weil die Aktorplatte alleine, in einer Linie oder auf einer Fläche angeordnet sein kann, und weil mit denselben Fertigungsschritten unterschiedliche Mikrosysteme realisiert werden können (Braillezeile, Relais, ...). Insbesondere die Kombinierbarkeit des Aktors mit anderen, fluidischen Bauelementen, die praktisch im gleichen Arbeitsgang hergestellt werden, sei hier hervorgehoben.

Prototypisch sind verschiedene Aktorplattentypen hergestellt und getestet worden. Ein im auf die Aktorplatte abgestimmtes ERF-Ventil wurde ebenfalls aufgebaut und erfolgreich als Ansteuerventil für den Aktor getestet.

## 8 Firma / Autor

**Bartels Mikrotechnik** bietet Produkte aus dem Anwendungsbereich der Mikrostrukturtechnik und der Mikrofluidik an. Das Ziel von Bartels Mikrotechnik ist es, Unternehmen aus den Anwendungsgebieten der Biotechnologie, Pharmaforschung und medizinischen Diagnostik sowohl bei der Entwicklung innovativer Mikroprodukte zu unterstützen als auch Produkte aus dem Bereich der Mikrofluidik anzubieten.

Dipl.-Ing **Isam Tahhan**, Jahrgang 1968, studierte Maschinenbau Richtung Feinwerk- und Mikrotechnik an der Technischen Universität Berlin und der University of California in Berkeley. Seit 1998 ist er bei der Firma Bartels Mikrotechnik GmbH, Dortmund, als Konstrukteur und Entwickler tätig.

## MODELLIERUNG UND OPTIMIERUNG DER AKTORPLATTE

A. Quinte, Institut für Angewandte Informatik, Forschungszentrum Karlsruhe GmbH  
quinte@iai.fzk.de

### Kurzfassung

In Zusammenarbeit mit den Projektpartnern Bartels Mikrotechnik GmbH und SIMEC GmbH wird an der Entwicklung einer universellen fluidisch betriebenen Aktorplatte gearbeitet, die als integrierte Komponente eines Mikrosystems für die Umwandlung fluidischer Energie in mechanische Energie eingesetzt werden soll.

Dieser Beitrag beschäftigt sich mit der Topologieoptimierung mit Evolutionären Algorithmen unter Verwendung eines parametrisierbaren FEM-Simulationsmodelles. Der Einsatz von Evolutionären Algorithmen erfordert für die Evaluierung kurze Simulationszeiten, so dass bislang physiknahe Simulationsmodelle auf Basis diskreter Elemente Methoden aufgrund ihrer üblicherweise hohen Rechenzeiten (im Vergleich zu höheren Modellen) hierfür nicht eingesetzt werden konnten. Die an unserem Institut entwickelte Parallelversion des Optimierungswerkzeuges GADO (Genetischer Algorithmus zur Design Optimierung) ermöglicht die Verwendung von physiknahen Simulationsmodellen bei der Optimierung mit einer globalen Suchstrategie.

## 1 Einleitung

### 1.1 Die Aktorplatte

Der fluidisch betriebene Mikroaktor (Abbildung 1) besteht aus zwei übereinander geschichteten dünnen Kunststofffolien, in die mit einem Excimerlaser für die Aktorplatte, Aktorkammer und fluidischen Anschlüssen die entsprechenden Mikrostrukturen eingebracht wurden [1]. Die Strukturierung der ca. 100 µm dünnen Kunststoffolie für die Aktorplatte erfolgt beidseitig zunächst in Form von ringförmigen und konzentrisch angeordneten Kerben. Durch die Aktorkammer zirkuliert ein flüssigförmiges Arbeitsmedium über das ein Arbeitsdruck aufgebaut werden kann.

Diese bewirkt eine vom Druck abhängige mechanische Auslenkung der Aktorplatte. Der Druckaufbau in der Aktorkammer wird durch ein aktives Ventil am Ausgang der Aktorkammer gesteuert.

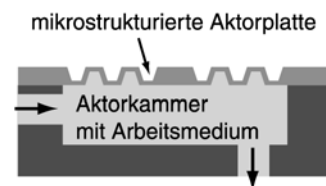


Abbildung 1: Mikroaktor mit Aktorplatte

### 1.2 Das Optimierungsproblem

Durch eine entsprechende Anzahl, Anordnung und Dimensionierung der Kerben kann das mechanische Verhalten der Aktorplatte gezielt beeinflusst werden. Ein wichtiges Optimierungsziel besteht darin, bei einem gegebenen Arbeitsdruck eine möglichst große Auslenkung der Aktorplatte zu erreichen. Gleichzeitig soll die Aktorplatte eine lange Lebensdauer besitzen, d.h. sie muss widerstandsfähig gegenüber häufigen Lastwechseln und äußeren mechanischen Einflüssen sein. Dazu trägt sowohl die Verwendung eines geeigneten Materials als auch eine entsprechend gewählte Geometrie bei.

Die vorliegende Optimierungsaufgabe beschränkt sich somit nicht auf eine reine Dimensionierungsoptimierung, bei der eine fest vorgegebene Anzahl von Geometrieparametern optimiert werden muss, sondern sie umfasst zusätzlich auch die Optimierung der Topologie in Form der Anzahl und Anordnung der Kerben. Die Optimierungsaufgabe führt damit zu einem Parametervektor mit dynamischer Länge in Verbindung mit einer hohen Parameterzahl. Durch die diskrete Anzahl der Kerben ist der Lösungsraum zudem multimodal. Aufgrund der Genauigkeitsanforderungen und der Topologievariabilität ist der Einsatz von FEM-Modellen bei der Lösung der Optimierungsaufgabe unum-

gänglich. Herkömmliche rein lokale Suchverfahren sind bei dieser Problemstellung in der Regel überfordert. Bereits die Multimodalität und die hohe Parameterzahl bilden für viele lokale Suchverfahren eine große Hürde. Unter diesen Voraussetzungen sind globale Suchverfahren, wie die hier eingesetzten evolutionären Suchstrategien, besser geeignet.

Für die Strukturoptimierung der Aktorplatte wurde zunächst das in unserem Institut entwickelte Optimierungswerkzeug GADO [2] und später die weiterentwickelte Variante HyGADO [8,9] eingesetzt. GADO ist die auf Designoptimierung zugeschnittene Variante des GLEAM-Verfahrens [3], einem eigenständigen Evolutionären Algorithmus, der Elemente der Evolutionsstrategie [4] und der Genetischen Algorithmen [5] mit Konzepten der Informatik verbindet. Obwohl Evolutionäre Verfahren generell dazu geeignet sind, mit Parametervektoren dynamischer Länge umgehen zu können, erlauben dies die meisten Implementierungen im Gegensatz zu GLEAM nicht, so dass GADO ohne Modifikationen für die vorliegende Aufgabe eingesetzt werden konnte.

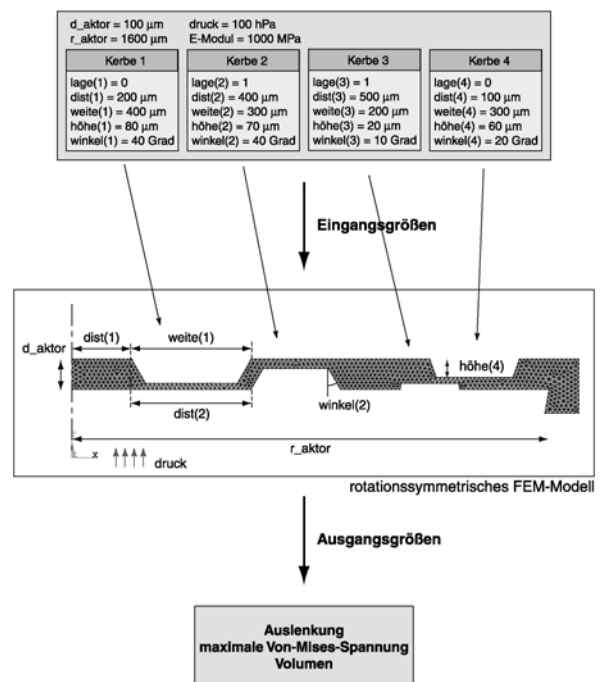
## 2 Strukturmechanisches Modell der Aktorplatte

Für die Aktorplatte wurde ein in seiner Geometrie frei parametrisierbares FEM-Modell für das kommerzielle FEM-Tool ANSYS erstellt (siehe Abbildung 2), mit dem das statische strukturmechanische Verhalten bei einer vorgegebenen homogenen Drucklast berechnet werden kann.

Wegen seiner Rotationssymmetrie ist das FEM-Modell zweidimensional aufgebaut.

Gegenüber einem dreidimensionalem FEM-Modell kann dadurch der Modellierungs- und Simulationsaufwand deutlich kleiner gehalten werden. Für die räumliche Diskretisierung wird ein unstrukturiertes Gitter verwendet, wodurch die automatisierte Geometriemodellierung erleichtert wird. Das FEM-Modell ist „Batch-Job“-fähig, d.h. unter Vorgabe eines Parametervektors kann der komplette Ablauf einer FEM-Simulation einschließlich der Geometriemodellierung, Gittergenerie-

rung und Auswertung der Analyseergebnisse vollautomatisch durchgeführt werden. Dies ist eine notwendige Voraussetzung für die automatisierte Durchführung umfangreicher Parameterstudien oder wie in diesem Fall für die rechnergestützten Optimierungen [6]. Der Parametervektor enthält neben den Basisparametern, wie z.B. Radius und Dicke der Aktorplatte zusätzlich für jede Kerbe einen eigenen Parametersatz, welcher die Lage, Position, Weite, Höhe und Wandwinkel der jeweiligen Kerbe beschreibt.



**Abbildung 2: Parametrisierbares FEM-Modell der Aktorplatte**

Die Anzahl der Kerben ist im FEM-Modell nicht fest vorgegeben, sie wird von dem Parametervektor bestimmt. Die Abbildung 2 zeigt ein Beispiel für ein parametrisiertes Modell mit vier Kerben, das gleichzeitig auch die Gestaltungsmöglichkeiten für die Strukturen verdeutlicht.

Für eine vorgegebene Drucklast liefert die nichtlineare statische Strukturanalyse als Ergebnis unter anderen die skalaren Kenngrößen, die Auslenkung der Aktorplatte, den Spitzenwert für die Von-Mises-Spannung, die lokal in der Struktur unter Last auftritt und das Volumen des Raumes, welches bei der Auslenkung der Aktorplatte unter der Platte

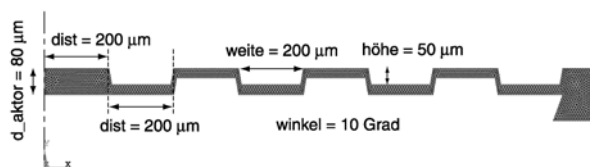
entsteht.

Die maximale Von-Mises-Spannung liefert eine Aussage über die mechanische Belastung der Aktorplatte. Das Volumen des Raumes ist eine Kenngröße, die später bei der Systemmodellierung für die Ermittlung einer fluidischen Kapazität benötigt wird.

Insgesamt werden für die Simulation einer Aktorplatte ca. drei bis zehn Minuten auf einer Sun-Workstation benötigt. Für Modalanalysen und für die Untersuchung des dynamischen Verhaltens der Aktorplatte wurden entsprechende FEM-Modelle erstellt.

### 3 Erster Entwurf der Aktorplatte

Ein erster manueller Entwurf für eine strukturierte Aktorplatte wurde von der Firma Bartels Mikrotechnik GmbH geliefert (siehe Abbildung 3), der vier oben liegende sowie drei unten liegende Kerben vorsieht, die so angeordnet sind, dass die Aktorplatte im Querschnitt betrachtet eine Mäanderform aufweist.



**Abbildung 3: Manueller Entwurf der Aktorplatte**

Von der Mäanderform werden vom Entwickler, ähnlich wie bei einem Faltenbalg, gute Eigenschaften hinsichtlich der erzielbaren Auslenkung bei gegebenem Arbeitsdruck erwartet. Der Druck wird von unten auf die gesamte Fläche der Aktorplatte beaufschlagt.

Die manuell entworfene Aktorplatte erreicht bei einem Arbeitsdruck von 100 hPa und einem E-Modul von 1000 MPa eine Auslenkung von 104,1 µm. Eine Modifizierung des manuellen Entwurfs, bei dem die Kerbenhöhe auf 60 µm erhöht wurde, bringt bereits eine Verbesserung der Auslenkung auf 160,3 µm.

Durch die größere Kerbenhöhe verbleibt für die modifizierte Aktorplatte somit eine Wandstärke von 20 µm.

**Tabelle 1: Mit FEM berechnete Auslenkungen von Aktorplatten im Vergleich**

Aktorplatte	Auslenkung (µm)	
	E-Modul = 1GPa, Druck = 100hPa	
manuell	104,1	451%
manuell mod.	160,3	694%
unstrukturiert (d_aktor = 80µm)	23,1	100%
unstrukturiert (d_aktor = 20µm)	95,5	413%

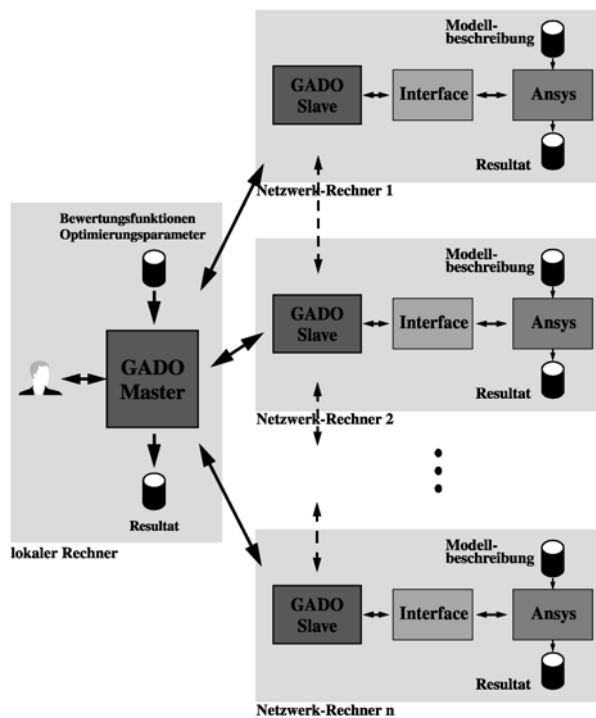
Im Vergleich dazu erreicht eine unstrukturierte Aktorplatte mit einer Dicke von 20 µm eine Auslenkung von 95,5 µm. Hier zeigt sich deutlich der Einfluss der Strukturierung auf die erzielbare Auslenkung.

### 4 Strukturoptimierung der Aktorplatte

Das Ziel im Einsatz der FEM besteht wie bereits erwähnt unter anderem darin, eine geeignete Strukturierung für die Aktorplatte zu finden, mit der sich bei vorgegebenem Arbeitsdruck eine möglichst hohe Auslenkung erzielen lässt. Das zur Fertigung der Aktorplatte eingesetzte Herstellungsverfahren lässt trotz fertigungs- und materialbedingter Restriktionen eine sehr große Gestaltungsfreiheit für die Aktorplattenstrukturierung zu, so dass der Suchraum für eine Optimierung sehr umfangreich werden kann.

Auf der einen Seite ist eine gezielte Suche nach einer optimalen Strukturierung ohne fundierte Kenntnisse in der Strukturmechanik und ohne eine gehörige Portion Erfahrung kaum machbar und auf der anderen Seite ist eine systematische Suche wegen der Größe des Suchraumes, wenn überhaupt, nur unter großem Zeitaufwand durchführbar.

Zur Lösung der vorliegenden Optimierungsaufgabe wurde zunächst eine auf ParallelGLEAM basierende Version von GADO eingesetzt [7] (siehe Abbildung 4). Durch die auf mehreren Netzwerkrechnern verteilte Ausführung der Optimierung lässt sich die Dauer eines kompletten Optimierungslaufes deutlich senken.

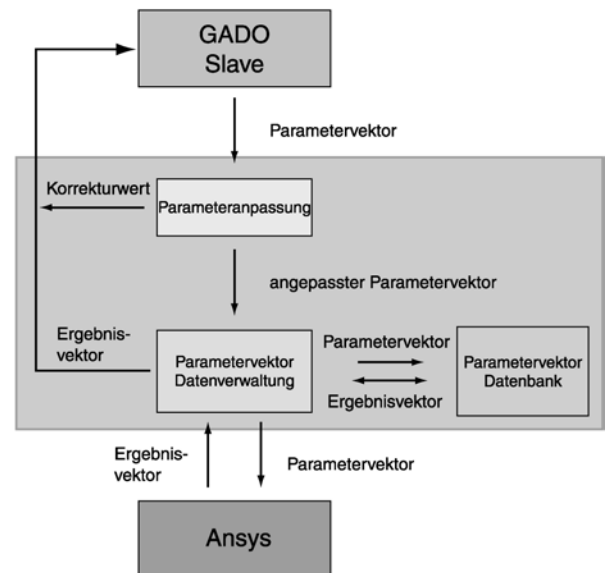


**Abbildung 4: Schematische Darstellung der Parallelversion von GADO mit dem Simulator ANSYS**

Die Parallelversion von GADO besteht aus zwei Komponenten. Die eine Komponente läuft als Masterprozess auf dem lokalen Rechner des Anwenders. Die andere Komponente besteht aus den Slaveprozessen, die vom Masterprozess auf jedem der beteiligten Netzwerkrechner gestartet werden. Der Masterprozess bildet die Schnittstelle zum Anwender und er initiiert und überwacht die Abläufe für die Durchführung einer Optimierung. Die eigentliche Optimierung wird von den Slaveprozessen durchgeführt, die untereinander kommunizieren, wobei jeder Slaveprozess für einen definierten Teil der Gesamtpopulation zuständig ist.

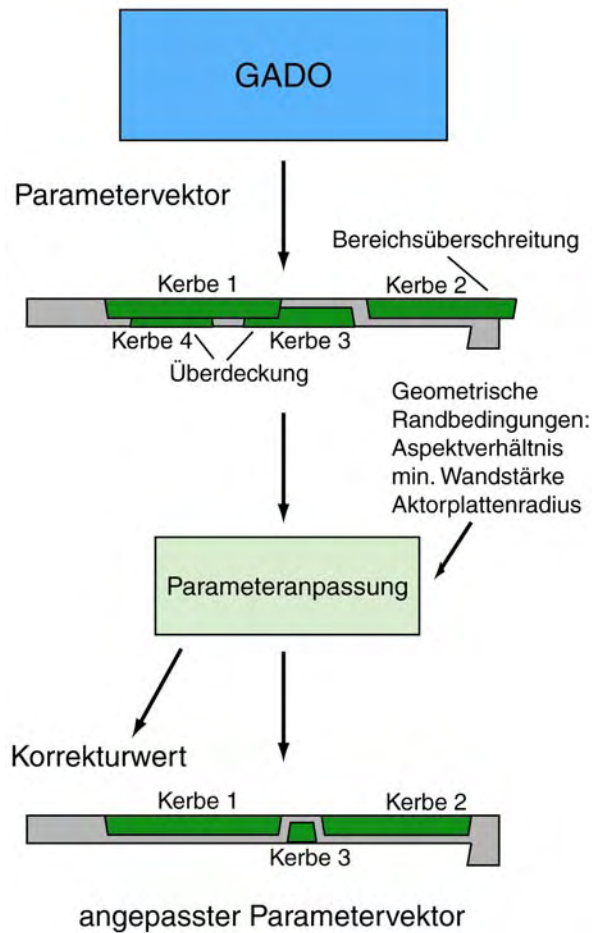
Die Kopplung zwischen einem GADO Slave und ANSYS erfolgt über ein spezielles Interfaceprogramm (siehe Abbildung 5), das die Ablaufsteuerung einer FEM-Simulation und die Aufbereitung der ausgetauschten Daten übernimmt. Bei der Aufbereitung werden die Kerben in der Reihenfolge, wie sie von GADO geliefert werden, auf Einhaltung der geometrischen Randbedingungen (Bereichsüberschreitung bzw. Überdeckung mit bereits platzierten Kerben) überprüft. Bei Auftreten einer Überdeckung oder Bereichsüberschrei-

ung wird eine Parameteranpassung durchgeführt, d.h. die Kerbe wird nach festgelegten Regeln entsprechend in ihrer Position verschoben und gegebenenfalls wird ihre Weite reduziert. Wenn dennoch kein Platz mehr vorhanden sein sollte, wird eine Kerbe auch ganz aus dem Parametervektor entfernt (siehe Abbildung 6).



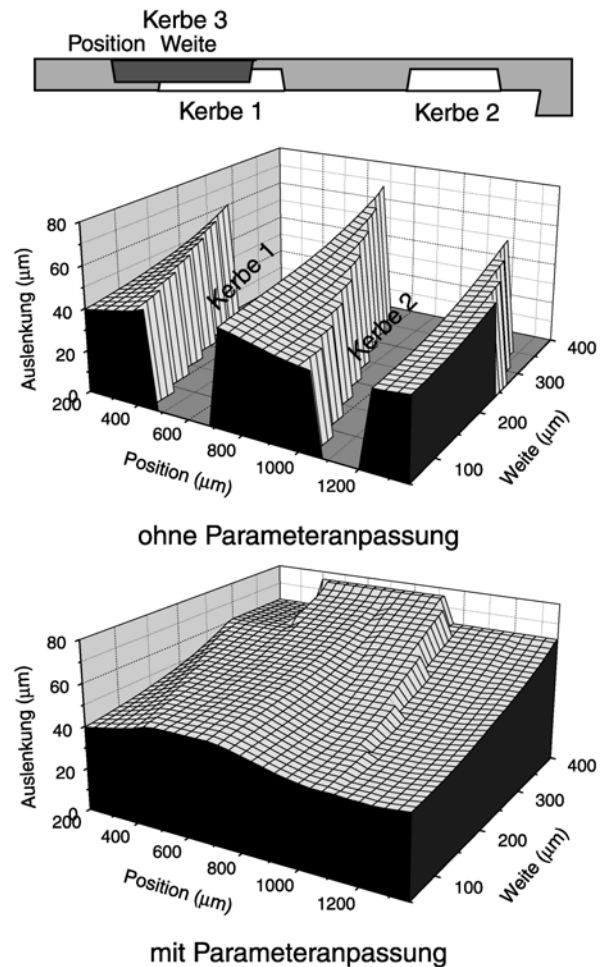
**Abbildung 5: Schematischer Aufbau des Interfaceprogrammes für die Kopplung zwischen dem GADO Slave und dem Simulator ANSYS**

Zusätzlich liefert die Anpassungsroutine als Ergebnis einen Korrekturwert, der eine Aussage über den Umfang der vorgenommenen Anpassungen des Parametervektors macht und zur Verbesserung der evolutionär erzeugten Parametervektoren benutzt wird. Allerdings kann die Parameteranpassung auch dazu führen, dass aus ähnlichen Parametervektoren, die von GADO geliefert werden, identische Parametervektoren für die Simulation entstehen. Daher wurde, um unnötige Mehrfachsimulationen zu vermeiden, im Schnittstellenprogramm zusätzlich eine Parametervektordatenbank implementiert, in der bei einem Optimierungslauf alle bereits simulierten Parametervektoren zusammen mit den Ergebnissen abgelegt werden.



**Abbildung 6: Korrektur eines vom GADO gelieferten Parametervektors**

Die Auswirkung der Parameteranpassung auf den Lösungsraum verdeutlicht das Beispiel in Abbildung 7. In diesem Beispiel wurden die Lösungsräume mithilfe von Variationsimulationen generiert, bei dem bei zwei fest vorgegebenen Kerben (Kerbe 1 und 2) die Position und die Weite der dritten Kerbe (Kerbe 3) variiert wurde. Deutlich ist die Zerklüftung des Lösungsraumes durch die Überdeckung der Kerbe 3 mit den anderen beiden fixen Kerben ohne die Parameteranpassung zu erkennen. Durch die Parameteranpassung ist der Lösungsraum dagegen die Zerklüftung deutlich weniger ausgeprägt, wodurch eine bessere Ausgangslage für die Optimierung erreicht wird. Unstetigkeiten im Lösungsraum, die bei bestimmten Topologieänderungen, wie z.B. Hinzufügen oder Entfernen von Kerben auftreten, lassen sich mit der Parameteranpassung allerdings dadurch nicht verhindern.



**Abbildung 7: Der Lösungsraum im Vergleich mit bzw. ohne Parameteranpassung**

Mit dieser Konfiguration wurden auf Basis des strukturellen FEM-Modells Optimierungsläufe durchgeführt, bei denen auch verschiedene fertigungs- und materialbedingte Randbedingungen mit berücksichtigt wurden. Im folgenden wird das Ergebnis eines dieser Optimierungsläufe vorgestellt.

## 5 Optimierungsergebnisse mit GADO

Der hier vorgestellte Optimierungslauf lief über ein Wochenende und dauerte insgesamt 2 Tage und 15 Stunden. Dabei waren zusätzlich zum lokalen Rechner insgesamt acht Netzwerkrechner an der Optimierung beteiligt.

Die Evolution lief über 61 Generationen bei denen von GADO insgesamt 28978 Parametervektoren generiert wurden. Von denen

wurden wegen der Parameteranpassung letztlich mit 11302 weniger als die Hälfte dann auch tatsächlich simuliert. Durch den Einsatz der Parametervektordatenbank konnte somit erheblich Rechenzeit eingespart werden.

Als Bewertungskriterien wurden die Auslenkung der Aktorplatte, die maximale Von-Mises-Spannung und der Korrekturwert herangezogen. Randbedingungen für die Optimierung bildeten unter anderem die Einhaltung eines maximal zulässigen Aspektverhältnisses von 4:1 für die Strukturen, die Einhaltung einer minimalen Wandstärke von 20  $\mu\text{m}$  und ein kerbenfreier Innenbereich. Weitere feste Vorgaben waren der E-Modul (1 GPa), der Arbeitsdruck (100 hPa), der Aktorplattenradius (1600  $\mu\text{m}$ ), der Wandwinkel ( $10^\circ$ ) und eine maximale Kerbenweite (1000  $\mu\text{m}$ ). Die Optimierungsparameter bildeten die Aktorplattendicke (in 20  $\mu\text{m}$  Schritten), sowie jeweils die Höhe, Weite, Position und Lage der Kerben. Um den Herstellungsaufwand für die Prototypen gering zu halten, wurde zunächst eine einheitliche Kerbenhöhe für alle Kerben verwendet.

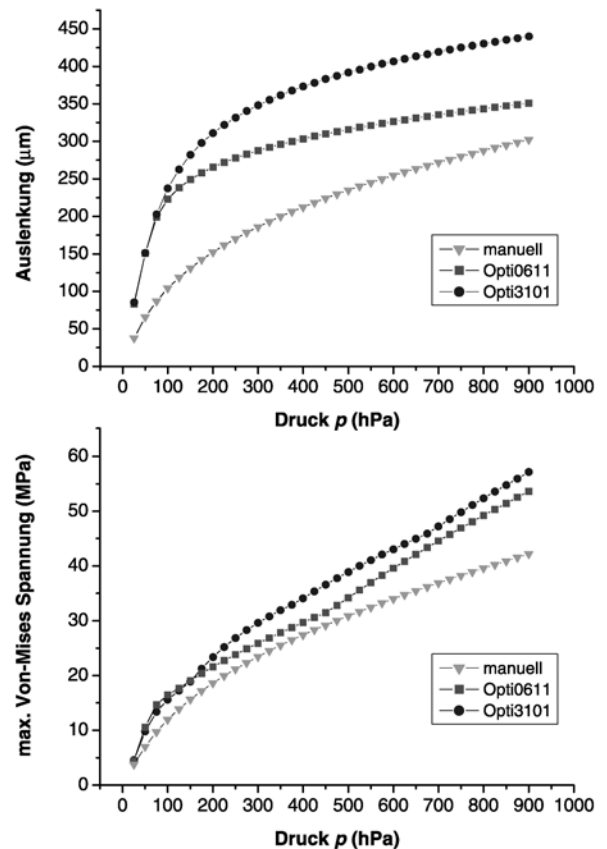
Die Anfangspopulation für die Optimierung wurde mit einem Zufallsgenerator generiert, wobei der gesamte gültige Suchraum zur Verfügung stand. D.h. es wurde keinerlei Vorwissen für die Optimierung eingebracht.

Die Optimierung lieferte unter den zuvor genannten Vorgaben eine Aktorplatte mit drei Kerben und der Aktorplattendicke von 120  $\mu\text{m}$  als Ergebnis (siehe Abbildung 8).



**Abbildung 8: Ergebnis einer Optimierung mit Parameteranpassung (Opti3101)**

Gegenüber dem ersten manuellen Entwurf erreicht die mit GADO optimierte Aktorplatte mit 238,3  $\mu\text{m}$  eine mehr als doppelt so große Auslenkung. Bei näherer Betrachtung der optimierten Aktorplatte wird deutlich, dass bei der Optimierung die vorgegebenen Restriktionen sehr gut ausgereizt wurden.



**Abbildung 9: Mit Simulationsrechnungen ermittelte Kennlinien der Aktorplatten**

In Abbildung 9 sind die Kennlinien für die Auslenkung (oben) und für die max. Von-Mises-Spannung (unten) der optimierten (Opti3101) und der manuell entworfenen Aktorplatte dargestellt. Zum Vergleich dazu ist die Kennlinie einer mit GADO optimierten Aktorplatte (Opti0611) dargestellt, bei dessen Optimierung keine Parameteranpassung verwendet wurde (siehe Abbildung 10).



**Abbildung 10: Ergebnis einer Optimierung ohne Parameteranpassung (Opti0611)**

In Tabelle 3 sind die bei der Simulation ermittelten Kenngrößen der Aktorplatten für den Arbeitsdruck 100 hPa und E-Modul 1-GPa aufgeführt.

## 6 Vergleich mit ersten Prototypen

Von Bartels Mikrotechnik wurden parallel zu weiteren Optimierungsläufen erste Prototypen hergestellt und Messungen durchgeführt, die hier im folgenden kurz angesprochen werden. Als Designvorlage für die Aktorplatte kamen in abgewandelter Form das manuelle Design nach Abbildung 3, das optimierte Design nach Abbildung 10 (Opti0611) sowie ein Design mit Einkerbungen in radialer Richtung zum Einsatz. Einige Abweichungen vom ursprünglichen Design waren deshalb notwendig geworden, weil für die Herstellung der ersten Prototypen zunächst keine Folien mit der gewünschten Dicke und Material zur Verfügung stand.

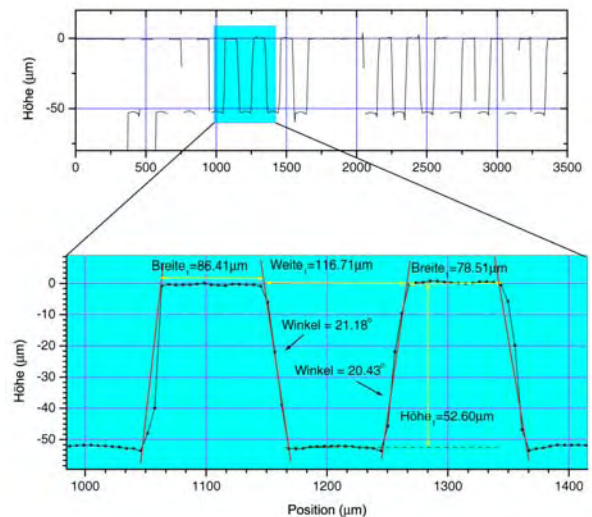
**Tabelle 2 Vergleich zwischen Entwurf und Realisierung.**

Aktorplatte	Entwurf		Prototyp	
	Man.	Opti0611	Man.	Opti0611
Dicke ( $\mu\text{m}$ )	80	100	125	
Kerbenhöhe ( $\mu\text{m}$ )	50	80	53	
Wandwinkel	10	10	21	
Simulierte Auslenkung ( $\mu\text{m}$ )	104,1	223	7,7	6,3
Gemesene Auslenkung ( $\mu\text{m}$ )	-	-	5	4

In Tabelle 2 sind einige der Designabweichungen zwischen Entwurf und Realisierung aufgelistet.

Weitere Abweichungen vom vorgegeben Design brachten Profilmessungen, die bei Bartels Mikrotechnik an einigen Prototypen vorgenommen wurden, ans Licht. Das in Abbildung 11 dargestellte Ergebnis einer Profilmessung zeigt unter anderem, dass die Kerbenweite, die für dieses Design mit 100  $\mu\text{m}$  vorgegeben war, mit 117  $\mu\text{m}$  deutlich größer geworden ist. Insgesamt führten die

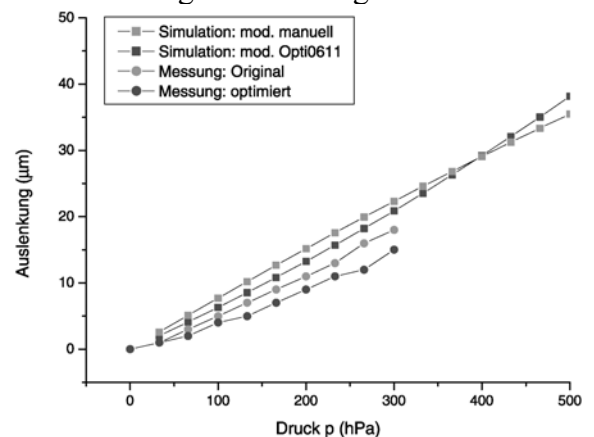
Abweichungen von den Designvorgaben dazu, dass mit den prototypischen Aktorplatten die laut Entwurf erwarteten Auslenkungen bei weitem nicht erreicht wurden (siehe Tabelle 2).



**Abbildung 11 Gemessener Profilverlauf eines nach einem manuellen Design hergestellten Prototyps (Unterseite) mit Ausschnittsvergrößerung.**

Um zumindest eine erste Evaluierung des Aktorplattenmodells durchführen zu können, wurden Simulationen mit den Designparametern der Prototypen durchgeführt. Dabei wurden auch die bei den Profilmessungen festgestellten Geometrieabweichungen mit berücksichtigt.

Gemessen wurde bei den Aktorplatten die statische Auslenkung in Abhängigkeit vom Druck. In Abbildung 12 sind die Messergebnisse zusammen mit den entsprechenden Simulationsergebnissen dargestellt.





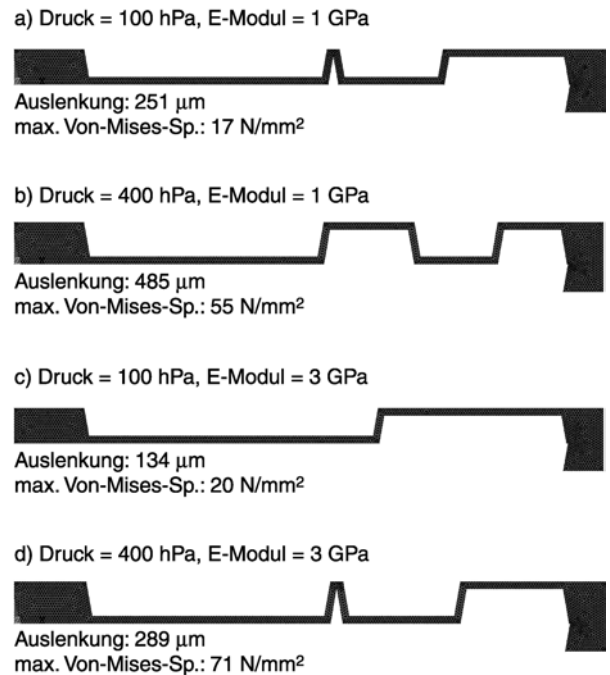
### Abbildung 12 Gemessene und simulierte Kennlinien zweier Aktorplattenprototypen im Vergleich

Obgleich die Abweichungen zwischen Messung und Simulation deutlich sichtbar sind, wird zumindest die qualitative Aussage der Simulation durch die Messung gut wiedergegeben.

## 7 Erste Optimierungsergebnisse mit HyGADO

Ein Nachteil globaler Suchverfahren, wie die hier bei GADO eingesetzten evolutionären Suchstrategien, besteht darin, dass sie sehr rechenzeitaufwendig sind. Dieser Nachteil soll bei HyGADO [8,9], einer Weiterentwicklung von GADO, durch eine Kombination von globalen und lokalen Suchverfahren vermieden werden. Mit HyGADO wurden vier Optimierungsläufe mit Parameteranpassung durchgeführt, deren Ergebnisse in Abbildung 13 dargestellt sind.

Bei ansonsten identischen Optimierungsvorgaben wurden die Optimierungsläufe mit zwei unterschiedlichen Arbeitsdrücken und E-Module durchgeführt. Da derzeit noch keine Parallelversion für HyGADO existiert benötigte jeder Optimierungslauf ca. 11 bis 12 Tage Rechenzeit. Für den Optimierungslauf des in Abbildung 13 a) dargestellten Optimierungsergebnisses, bei dem die gleichen Vorgaben verwendet wurden wie bei den Optimierungsläufen mit GADO, wurden von HyGADO insgesamt 122847 Parametervektoren erzeugt. Aufgrund der Parameteranpassung wurden von denen dann lediglich 5939 tatsächlich auch simuliert. Diese enorme Reduktion ist auf die Konvergenzphase des häufig aufgerufenen lokalen Suchverfahrens zurückzuführen, in der viele sehr ähnliche Parametervektoren erzeugt werden, die das zwischen HyGADO und dem Simulator geschaltete Interfaceprogramm dann als identisch ausfiltert.



**Abbildung 13: Optimierungsergebnisse mit HyGADO mit unterschiedlichen Vorgaben**

Im Vergleich zu den bisherigen Optimierungsergebnissen mit GADO konnte mit HyGADO die Auslenkung weiter erhöht werden (siehe Tabelle 3). Gleichzeitig waren dafür auch erheblich weniger Simulationen erforderlich.

**Tabelle 3: Optimierte Aktorplatten im Vergleich ( Druck = 100 hPa, E-Modul = 1 GPa)**

Aktorplatte	Auslenkung ( $\mu\text{m}$ )		Max. Von-Mises-Spannung (N/mm <sup>2</sup> )	
manuell	104,1	100%	12,0	100%
Opti0611	223,0	214%	16,5	137%
Opti3101	238,3	229%	15,9	133%
1-100	251,0	241%	17,0	142%

Die unterschiedlichen Optimierungsergebnisse in Abbildung 13 zeigen deutlich den Einfluss der verschiedenen Vorgaben auf das Resultat einer Optimierung. Aufgrund der wenigen Optimierungsergebnisse, die bislang zur Verfügung stehen, lassen sich noch keine allgemeingültigen Gesetzmäßigkeiten für die optimale Topologie und Dimensionierung erkennen bzw. ableiten. Lediglich die Annahmen, dass eine mäanderförmige Anordnung der Kerben, eine hohe Kerbenpackungsdichte und eine möglichst geringe

resultierende Wandstärke sich vorteilhaft auf die erzielbaren Auslenkungen auswirken, konnten dadurch bestätigt werden. Durch weitere Optimierungsläufe mit anderen Vorgaben lassen sich diese Wissenslücken weiter reduzieren. Wenngleich mit diesem Vorwissen sich wohl kaum für jede denkbare Vorgabe bereits das optimale Aktorplattendesign ableiten lässt, kann jedoch mit dem Vorwissen der Suchraum und damit der Zeitaufwand einer Optimierung deutlich reduziert werden.

**Tabelle 4: Die Auslenkung ( $\mu\text{m}$ ) der optimierten Aktorplatten bei unterschiedlichen Randbedingungen**

Optimierte Aktorplatte Abbildung 13	Randbedingung: E-Modul, Druck			
	1 GPa, 100 hPa	1 GPa, 400 hPa	3 GPa, 100 hPa	3 GPa, 400 hPa
a)	251	385,6	100,0	280,8
b)	167,1	485	67,9	210,9
c)	223,3	321,3	134	244,2
d)	228,7	443,8	79,0	289

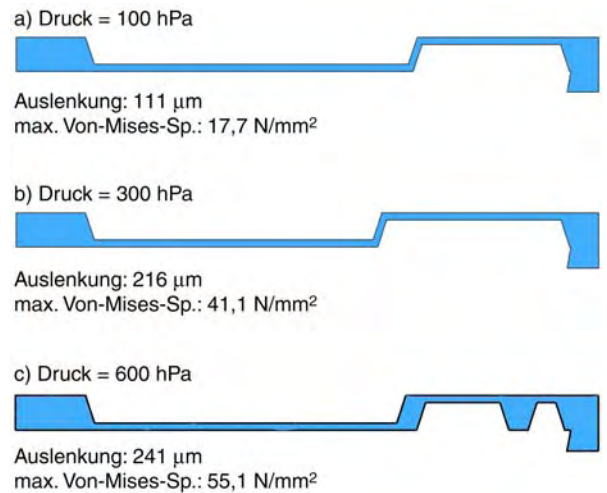
In Tabelle 4 sind zum Vergleich die erzielbaren Auslenkungen aufgelistet, die mit den von HyGADO optimierten Designs (Abbildung 13) unter Verwendung der vier dort eingesetzten Vorgaben berechnet wurden.

In allen Fällen liefert bei den Vorgaben jeweils die darauf optimierte Aktorplatte auch die höchste Auslenkung (in Tabelle 4 grau unterlegt), teilweise sogar mit großem Abstand im Vergleich zu den anderen Aktorplatten. Dies lässt den Schluss zu, dass es hier durchaus Sinn macht die Aktorplatte entsprechend der zu erwartenden Betriebsbedingungen und des eingesetzten Materials zu optimieren.

## 8 Optimierung mit HyGADO

Die Erfahrungen, die bislang mit den ersten Prototypen und Optimierungsläufen gesammelt wurden, führten zu der Entwicklung einer neuen Prototypenserie. Als Ausgangsmaterial für die Aktorplatte soll diesmal ein  $100\ \mu\text{m}$  dickes Polycarbonat-Folie mit einem E-Modul von  $3800\ \text{MPa}$  verwendet werden.

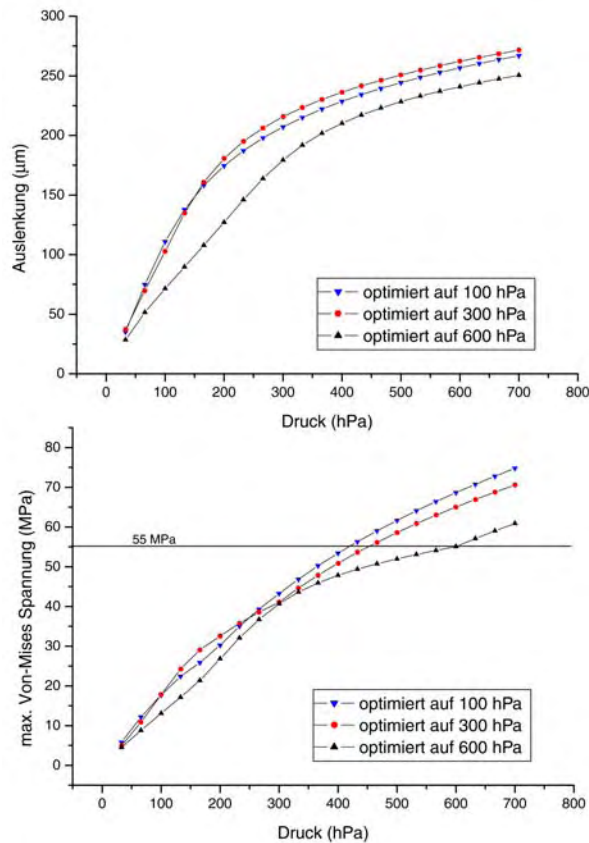
Vorversuche mit diesem Material bei Bartels-Mikrotechnik ergaben, dass mit dem Laserverfahren auch die Strukturen geometrietreuer hergestellt werden können als dies mit den bisher verwendeten Material der Fall war.



**Abbildung 14: Optimierungsergebnisse mit HyGADO für die Aktorplatten aus  $100\ \mu\text{m}$  dicken Polycarbonat-Folie (E-Modul:  $3800\ \text{MPa}$ )**

Auf Basis des neuen Materials wurden mit HyGADO insgesamt drei Optimierungsläufe mit drei verschiedenen Arbeitsdrücken ( $100\ \text{hPa}$ ,  $300\ \text{hPa}$  und  $600\ \text{hPa}$ ) durchgeführt, dessen Ergebnisse in Abbildung 14 zu sehen sind. In Abbildung 15 sind die entsprechenden Kennlinien für die Auslenkung (oben) und für die max. Von-Mises-Spannung (unten) der Aktorplatten aus Abbildung 14 dargestellt.

Für die Optimierung wurden auch deshalb höhere Arbeitsdrücke verwendet, hier um wegen dem hohen E-Moduls des Material mit den Aktorplatten dennoch ausreichend große Auslenkungen erzielen zu können. Die maximale Zugspannung für das Material wird vom Hersteller mit  $60\text{-}70\ \text{MPa}$  angegeben. Für die Optimierung wurde unter Berücksichtigung einer hier willkürlich gewählten Sicherheitsreserve  $55\ \text{MPa}$  als oberer Grenzwert für die maximale Von-Mises Spannung vorgegeben.



**Abbildung 15: Mit Simulationsrechnungen ermittelte Kennlinien der mit HyGADO optimierten Aktorplatten.**

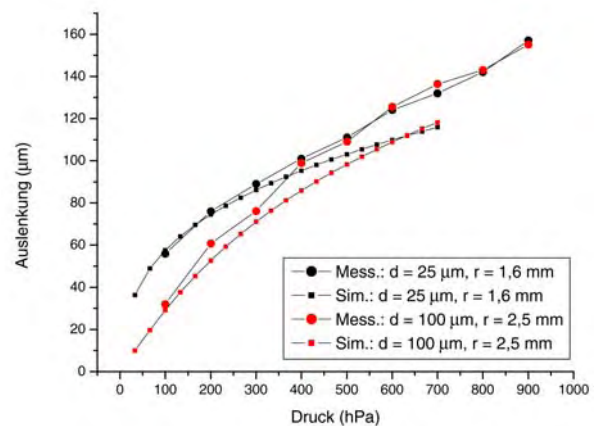
Für die Optimierungsläufe wurden als feste Größen die Dicke der Aktorplatte ( $100\ \mu\text{m}$ ), die Kerbenhöhe ( $80\ \mu\text{m}$ ) und der Wandwinkel ( $18^\circ$ ) verwendet um auch die Optimierung zu beschleunigen. Ansonsten wurden die gleichen Vorgaben verwendet wie bei den zuvor beschriebenen Optimierungsläufe. Die Ergebnisse in Abbildung 14 a) und b) ähneln dem Ergebnis in Abbildung 13 c). Etwas aus dem Rahmen der erzielten Optimierungsergebnisse fällt die Aktorplatte in Abbildung 14 c), die für einen Arbeitsdruck von  $600\ \text{hPa}$  optimiert wurde. Im Gegensatz zu den anderen Aktorplatten wurde hier bei der Optimierung das Design maßgeblich durch den vorgegebenen Grenzwert für die maximale Von-Mises Spannung beeinflusst. Für die Realisierung von Aktorplatten mit hoher Auslenkung ( $>250\ \mu\text{m}$ ), wie sie für taktile Arrays erforderlich wären, dürfte somit die hier zunächst vorgesehene Polycarbonat-Folie kein sehr geeignetes Material sein. Hier erweist sich die maximale Zugspannung als begrenzender Faktor, der auch nicht durch eine optimierte Strukturie-

rung und Erhöhung des Arbeitsdruckes aufheben lässt.

Für die Prototypenherstellung wurde dann doch eine andere Polycarbonat-Folie mit einem niedrigeren E-Modul von  $2,3\ \text{GPa}$  –  $2,4\ \text{GPa}$  (laut Hersteller) verwendet. Durch die höhere Auslenkungen, die mit dem weichen Material erreicht werden, kann unter anderem auch die Genauigkeit der Messungen gesteigert werden.

## 9 Vergleich mit Messungen

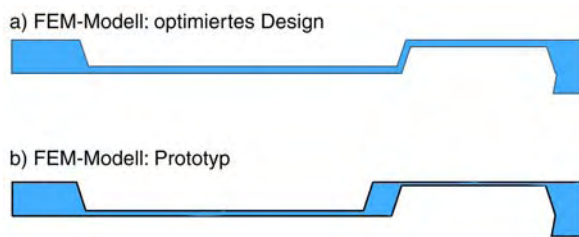
Zur Evaluierung des Simulationsmodells wurden auch Messungen an unstrukturierten Aktorplatten durchgeführt. Dadurch können unerwünschte Verfälschungen, die durch geometrische Abweichungen vom Design, die bei einer Strukturierung mit dem Excimer-Laserverfahren entstehen würden, vermieden werden. Vermessen wurden dabei zwei unstrukturierte Aktorplatten mit dem Radius  $1,6\ \text{mm}$  und Dicke  $25\ \mu\text{m}$  sowie mit dem Radius  $2,5\ \text{mm}$  und Dicke  $100\ \mu\text{m}$ . Die Messergebnisse sind zusammen mit den entsprechenden Simulationsergebnissen in Abbildung 16 zusammengefasst.



**Abbildung 16: Vergleich zwischen Simulation und Messung von zwei unstrukturierten Aktorplatten.**

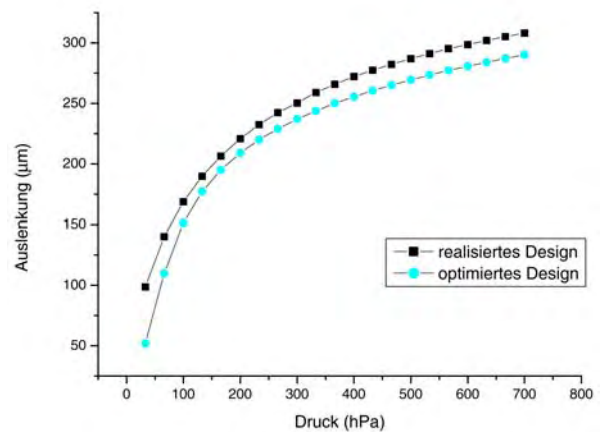
Die Abweichungen zwischen Simulation und Messungen sind bei kleinen Drücken relativ gering. Erst bei höheren Drücken nehmen sie sichtlich zu. Der Hersteller der Folie gibt für den E-Modul einen Bereich von  $2,3\ \text{GPa}$  bis  $2,4\ \text{GPa}$  an, simuliert wurde mit  $2,35\ \text{GPa}$ , so dass die Differenz zwischen Simulation und

Messung durchaus noch im Rahmen einer möglichen Parameterabweichung liegen können.



**Abbildung 17: Vergleich zwischen dem optimierten Design und dem realisierten Prototypen.**

Im Vergleich zu den ersten Prototypen konnten diesmal die neuen Prototypen deutlich geometrietreuer realisiert werden. In Abbildung 17 sind der optimierte Entwurf und das danach tatsächlich realisierte Design gegenübergestellt. Die Geometriedaten des Prototypen wurden ebenfalls aus den Ergebnissen von Profilmessungen ermittelt und im FEM-Modell mit berücksichtigt. In Hinblick auf eventuelle Justageungenauigkeiten und Aufgrund der bisher gewonnenen Erfahrungen wurde als zusätzliche Sicherheit die innere Kerbe um ca.  $50\ \mu\text{m}$  verkürzt, so dass zwischen der inneren und äußeren Kerbe die Seitenwand deutlich dicker ausfiel. Im Nachhinein erwies sich diese Sicherheitsmassnahme allerdings als überflüssig. Weitere deutliche Abweichungen vom optimierten Design sind bei der Kerbenhöhe feststellbar, die bei der inneren Kerbe  $84,6\ \mu\text{m}$  und bei der äußeren Kerbe sogar  $90,42\ \mu\text{m}$  anstatt wie vorgesehen  $80\ \mu\text{m}$  beträgt. Die mit FEM-Simulationen berechnete Kurven für die Auslenkung der beiden Designs sind in Abbildung 18 dargestellt.



**Abbildung 18: Simulationsergebnisse vom optimierten und realisierten Design der Aktorplatte (E-Modul: 2350 MPa).**

Bei Fertigstellung dieses Beitrages standen leider noch keine gültige Messdaten der strukturierten Aktorplatte zur Verfügung. Erste Messversuche deuten allerdings darauf hin, dass auch hier mit guten Ergebnissen gerechnet werden kann [10].

## 10 Zusammenfassung

GADO und später dann HyGADO hat sich als ein sehr stabiles und leistungsfähiges Werkzeug zur Lösung des hier vorliegenden Optimierungsproblems erwiesen. Die rechnergestützte Optimierung hat hier maßgeblich zum Entwurf der Aktorplatte beitragen können.

Das auch anspruchsvolle Topologieoptimierungen auf Basis von rechenzeitintensiven FEM-Modelle möglich sind konnte hier aufgezeigt werden. Das dies auch mit einem vertretbaren Rechenaufwand gelingt, zeigt die verteilte Optimierung mit der Parallelversion von GADO, bei der die Gesamtdauer einer Optimierung von 10 Tagen auf wenige Tage reduziert werden konnte.

## Literatur

- [1] Tahhan, I.: Die Aktorplatte – ein universeller fluidischer Mikroaktor. Statusseminar des BMBF-Verbundprojektes OMID, Bremen, 9.-10. Okt. 2001
- [2] Jakob, W., Meinzer, S., Quinte, A., Süß, W., Gorges-Schleuter, M., Eggert, H.: Partial

Automated Design Optimization Based on Adaptive Search Techniques. In: I. C. Parmee (ed.): Adaptive Computing in Engineering Design and Control '96, PEDC, University of Plymouth, (1996), p. 236-241

[3] Blume, C.: GLEAM - A System for Simulated "Intuitive Learning". In: H.-P. Schwefel, R. Männer (eds.): Proc. of PPSN I, LNCS 496, Springer-Verlag, (1991), S.48-54.

[4] Rechenberg, I.: Evolutionsstrategie - Optimierung technischer Systeme nach Prinzipien der biologischen Evolution. Frommann-Holzboog, Stuttgart-Bad Cannstatt. 1973

[5] Holland, H. J.: Adaptation in Natural and Artificial Systems. The University of Michigan Press, Ann Arbor. 1975

[6] Quinte, A., Eggert, H., Meinzer, S., Süß, W.: Entwicklung von Verfahren für den Einsatz der Finite-Elemente-Methode in der Systemoptimierung von Mikrosystemkomponenten,

Dissertation an der Universität Bremen, Wissenschaftliche Berichte FZKA 5951, Forschungszentrum Karlsruhe, April, 1998

[7] Jakob, W., Quinte, A., Scherer, K.-P., Eggert, H.: Optimization of a micro fluidic component using a parallel evolutionary algorithm and simulation based on discrete element method. Proceed. of the 6th int. Conf. on Computer Aided Optimum Design of Structures, WIT Press, (2001), S. 337-346

[8] Jakob, W., Peters, D.: Designoptimierung mit HyGADO. Statusseminar des BMBF-Verbundprojektes OMID, Bremen, 9.-10. Okt. 2001

[9] Jakob, W., Peters, D., Reiffer, A.: Internetbasierte Designoptimierung mit HyGADO. Abschlußseminar des BMBF-Verbundprojektes OMID, Karlsruhe, 25.-26. Nov. 2002

[10] Tahhan, I.: Die Aktorplatte - ein universeller fluidischer Mikroaktor. Abschlußseminar des BMBF-Verbundprojektes OMID, Karlsruhe, 25.-26. Nov. 2002

## EINE LASERSTEUERUNG FÜR OPTISCHE MESSGERÄTE

G. Gärtner, T. Dittrich VacuTec Meßtechnik GmbH

Email: Gerhard.Gaertner@vacutec-gmbh.de

### Kurzfassung

Zum Erreichen der gewünschten Messgenauigkeit müssen Laserquellen für optische Messgeräte eine hohe Stabilität aufweisen. Angepasst an die jeweilige Applikation sind spezielle Leistungsparameter einzustellen und stabil zu halten. Das Messsystem unterliegt vielfältigen äußeren und inneren Beeinflussungen, welche ausgeregelt werden müssen. Die Zusammenhänge zwischen Einflussgrößen und Auswirkung auf die Laserleistung sind oft nichtlinear und schwer zu beschreiben. Die Ansteuerschaltungen für die Laserquellen sollen die relevanten Einflussfaktoren umsetzen und ausgleichen. Die Entwicklung dieser Schaltungen ist kompliziert und aufwändig.

Durch die Modellierung des Messsystems und der Zusammenhänge der Systemkomponenten soll die Basis für die Simulation von Schaltungsvarianten bei der Weiterentwicklung der Lasersteuerung gelegt werden.

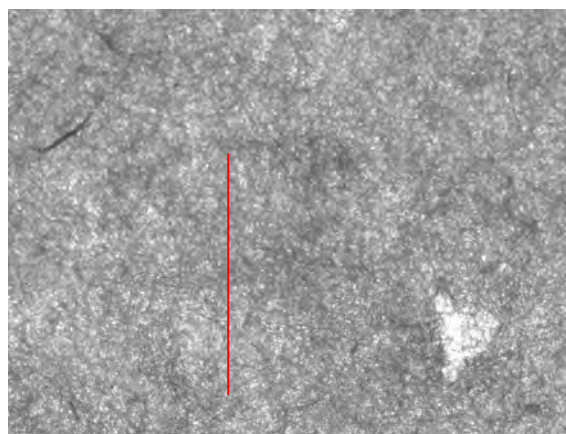
### 1 Einleitung

Optische Messgeräte der VacuTec Meßtechnik GmbH werden zur Erfassung der Struktur und der Oberfläche von Materialien verwendet. Neben der traditionellen Offline-Messung im Labor erfolgt zunehmend die Online-Messung während der Herstellung der Materialien.

Die optische, berührungslose Methode ermöglicht die kontinuierliche Kontrolle des Fertigungsprozesses. Durch hochgenaue Messung der Transmission bzw. Absorption an laser- bzw. lichtdurchlässigen Stoffen mittels eines Lasersensors wird die Homogenität, die Dicke und Materialstruktur bandförmiger Materialien bestimmt (Abb.1).

Mit einem Sensor auf Basis der Laseroptik kann mittels transmittierendem Licht die For-

mation von Papier (20g/m<sup>2</sup> ... 300g/m<sup>2</sup>) ermittelt werden. Bei der sich mit bis zu 2000m/min schnell bewegenden Papierbahn werden auf langen Spuren jeweils einige Tausend Messwerte aufgenommen.



**Abbildung 1: Feinstruktur der Formation einer Papierbahn**

Die analogen Messwerte werden mit einer Genauigkeit von besser als ein Promille digitalisiert. Verschiedene Kennzahlen, die aus den Messwerten berechnet werden, kennzeichnen die Struktur des Papiers.

Eine weitere Anwendung von optischen Messgeräten der VacuTec Meßtechnik GmbH liegt in der Vermessung der Oberfläche von verschiedensten Materialien mit einem speziellen Laser-Reflexionssensor. Die Höhenauflösung liegt im Bereich 0,2µm ... 1µm. Aus den Messdaten werden das Oberflächenprofil (Abb.2) und Kennzahlen wie zum Beispiel die Rauheit oder die Kreppung ermittelt.

Diese Auswertungsergebnisse erlauben eine höhere Stufe der Bewertung der Produktqualität und eine Verbesserung der Produktionsregelung.

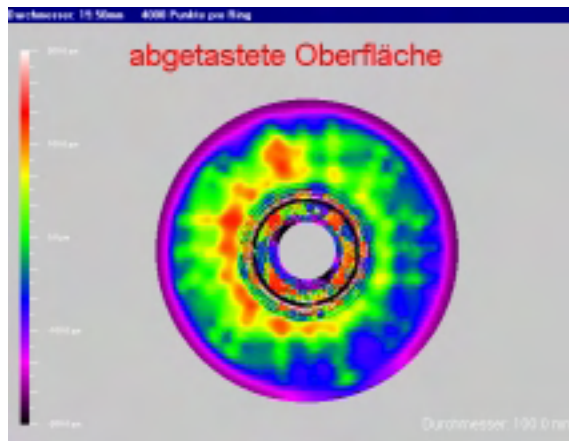


Abbildung 2: Oberflächenprofil einer CD

## 2 Überblick über den Systemaufbau

Das Messsystem besteht aus Sender und Empfänger. Als Sender dient eine Laserstrahlenquelle und als Empfänger dienen Photodioden. Durch die Eigenschaften des Materials in der Messstrecke wird die Transmission bzw. Reflexion des Laserlichts beeinflusst.

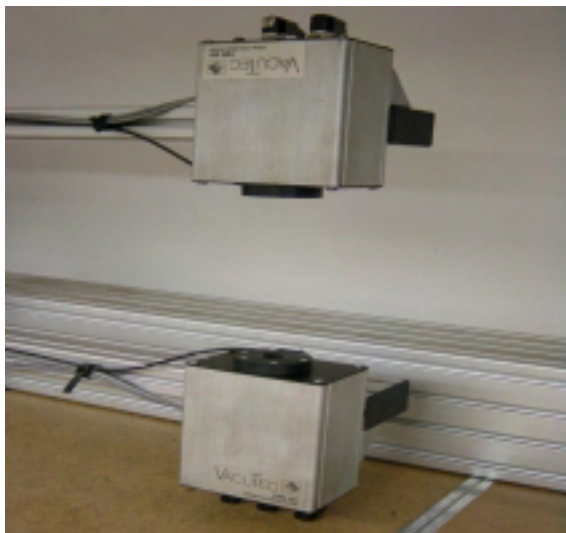


Abbildung 3: Anordnung für Transmissionsmessungen

Die Komponenten des Systems werden fest in sogenannten C-Bügeln (Abb.3) an den Messpositionen oder in Traversierrahmen gemeinsam mit anderen Sensoren in den Produktionsanlagen installiert.

### 2.1 Komponenten des Systems

Das Messsystem besteht aus einer Reihe von elektrischen, optischen und mechanischen Komponenten. Das Blockbild (Abb.4) zeigt die wesentlichen Bestandteile. Mit der Laserdiode wird ein feiner Laserstrahl (Durchmesser 10...50µm) erzeugt und auf das zu vermessende Material gerichtet. Die Laserdiode wird von einer elektronischen Baugruppe angesteuert. Das optische System erfasst das reflektierte bzw. transmittierte Licht und teilt es in die charakteristischen Bestandteile auf. Diese Bestandteile werden mit hochempfindlichen, sehr schnellen Photodioden und dazugehörigen Verstärkerschaltungen aufgenommen und in analoge Spannungssignale gewandelt. Die Spannungen repräsentieren zum Einen den eigentlichen Messwert und sind zum Anderen als bewertete Referenzspannungen Eingangswerte für die Steuerbaugruppe.

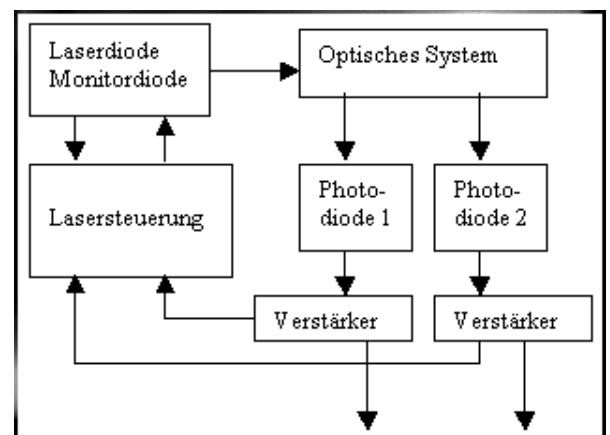


Abbildung 4: Hauptkomponenten der Lasersteuerung

Die Systemkomponenten werden je nach Applikation dimensioniert und in verschiedenen geometrischen Anordnungen verwendet. Somit erhalten die Messsysteme unterschiedliche äußere Form. Die Bestandteile und deren Funktion sind jedoch in wesentlichen Zügen immer wieder gleich. Das trifft in besonderem Maße für die Erzeugung des Laserstrahls und die Ansteuerung zu.



**Abbildung 5: Formationssensor**

Bei der Formationsmessung an Papier zum Beispiel sind die Komponenten des Sensors (Abb.5) in zwei robusten Gehäusen beidseitig der Materialbahn untergebracht. Je nach Messaufgabe wird ein bestimmtes Grundkonzept und damit auch eine festgelegte Grundkonfiguration ausgewählt. Dazu gehört zum Beispiel auch die Auswahl eines Lasers mit einer entsprechenden Laserleistung.

## 2.2 Beabsichtigte Einstellmöglichkeiten

Da die Materialien, die auf einer Produktionsanlage hergestellt werden und deshalb auch mit ein und demselben System vermessenbar sein müssen, in ihren optischen Eigenschaften variieren, muss die Helligkeit des Lasers im Arbeitsbereich auf einen optimalen Arbeitspunkt einstellbar sein. Das heißt, es müssen verschiedene Helligkeitsstufen vorwählbar sein.

Das ist anders als bei Anwendungen zum Beispiel in der digitalen Datenübertragung, bei denen eine Universaleinstellung allzeit verwendbar ist.

Weiterhin gibt es Sonderzustände, die stabil einstellbar sein müssen. Das trifft unter anderem für die Justage und für Sicherheitszustände bei der Wartung zu.

## 2.3 Unbeabsichtigte Einflussfaktoren

Neben den gewollten Einstellmöglichkeiten gibt es eine Reihe von Einflussfaktoren, welche von außen kommend, das Messsystem stören. Dazu gehören Fremdlicht, Temperaturschwankungen, mechanische Schwingungen

und EMV-Einflüsse auf die Spannungsversorgung. Das ist um so mehr zu beachten, da die Messgeräte für den Einsatz in Produktionslinien mit ausgesprochen rauen Umgebungsbedingungen vorgesehen sind.

Außerdem können die aktiven Komponenten des Systems selbst Quelle unbeabsichtigter Beeinflussungen sein. Vor Allem erzeugen die elektronischen Baugruppen Wärme. Damit unterliegen sie während des Betriebs Veränderungen. Das spiegelt sich teilweise in den Kennlinien der Bauelemente wider. Ebenso können sich die Komponenten des Systems gegenseitig beeinflussen. So kann Wärme an den hochempfindlichen mechanischen Aufbauten zu geringfügigen Verformungen und damit zur Dejustage führen.

Die gewollten und ungewollten Einflüsse und Verknüpfungen der Komponenten untereinander sind sehr vielfältig und komplex.

Ein Ziel des Projektes ist es, diese Einflüsse und Verknüpfungen zu beschreiben. Anhand geeigneter Versuchsreihen werden bei VacuTec die wesentlichen Einflüsse herausgearbeitet. Vom Projektpartner Fhg EAS Dresden wird das Messsystem modelliert [1]. Anhand von Simulationen sollen geeignete Maßnahmen erarbeitet werden, um einen hochstabilen Zustand der Laserquelle zu erreichen und damit die Basis für die notwendige Messgenauigkeit zu legen.

## 3 Komponenten und ihre Eigenschaften

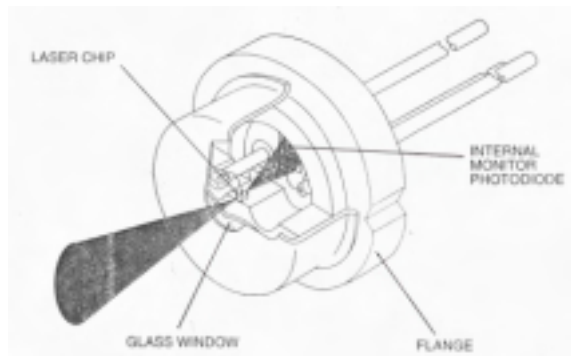
Zum Verständnis der Lasersteuerung werden die Systemkomponenten, ihre Funktion und die speziellen Einflüsse näher beschrieben.

### 3.1 Laser

Die Laserquelle ist ein kompaktes Bauelement. Sie beinhaltet die Halbleiterlaserdiode und die Monitordiode. Beide sind von einem Metallgehäuse mit Austrittsfenster umgeben (Abb.6). Das Bauelement wird durch Linsen zur Strahlkollimierung und ein robustes Außengehäuse ergänzt. Als Beispiel für den

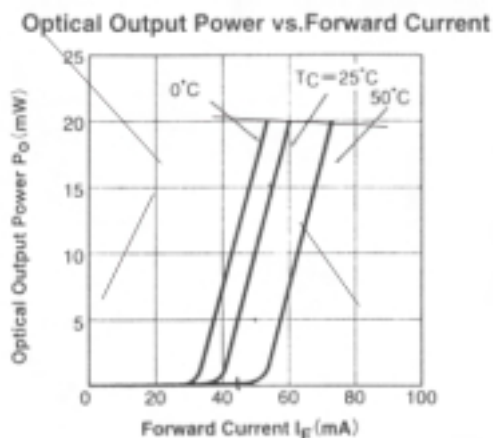


Aufbau und die Bauelementedaten wird die TOLD 9140 verwendet. [2]



**Abbildung 6: Aufbau einer Laserdiode**

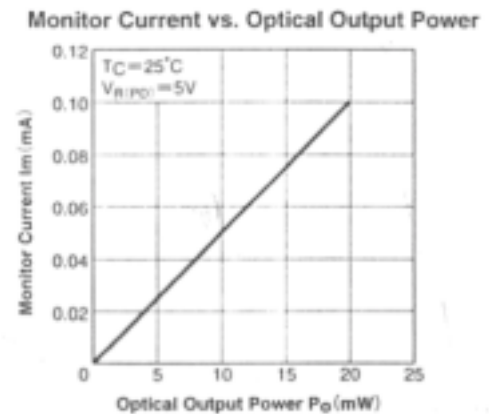
Die erforderliche Laserleistung wird durch Einspeisung eines entsprechenden Stromes erzeugt. Dieser Strom ist damit der entscheidende Faktor für die Stabilisierung der abgegebenen Laserleistung.



**Abbildung 7: Kennlinie Laserleistung**

Das Laserlicht wird erst nach Erreichen eines Schwellstromes erzeugt. Die Kennlinie zwischen Laserleistung und Laserstrom (Abb.7) ist vor allem im Knickbereich stark nichtlinear. Weiterhin ist zu erkennen, dass der Zusammenhang stark temperaturabhängig ist.

Die interne Monitordiode dient der Überwachung der Laserdiode. Ein Teil des abgestrahlten Laserlichtes erreicht die entgegengesetzt zur Strahlrichtung liegende Diode und erzeugt dort einen Photostrom, den sogenannten Monitorstrom (Abb.8).



**Abbildung 8: Kennlinie Monitorstrom**

Der Monitorstrom wird als Äquivalent für die abgegebene Laserleistung zur Ansteuerung der Schutzschaltung und für die Regelung der Laserleistung verwendet. Eine häufige oder dauerhafte Überschreitung der maximalen Laserleistung durch Einspeisung eines zu hohen Stromes führt zwangsläufig zur Zerstörung der Laserdiode und muss deshalb verhindert werden.

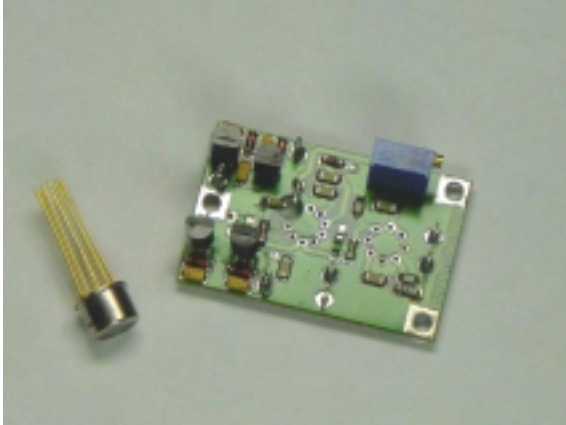
Es ist zu beachten, dass bei einer ungünstigen Anordnung auch vom Messgut oder vom optischen System reflektiertes Licht die Monitordiode erreichen kann. Dadurch wird eine höhere Laserleistung vorgetäuscht als real abgegeben wird.

### 3.2 Photodiode mit Verstärkerschaltung

Als Sensor für das Laserlicht wird eine großflächige Photodiode (Abb.9) verwendet. Die Diode wandelt das Licht in einen sehr kleinen Photostrom, welcher im Nano- oder Picoampere-Bereich liegt. Der Strom wird durch eine hochohmige Verstärkerschaltung in eine Spannung gewandelt. Hochohmige Verstärker unterliegen einem starken Einfluss hinsichtlich Temperatur und Instabilität der Spannungsversorgung. Weiterhin sind die Photodiode und der Verstärker die hauptsächlichen Rauschquellen im System. Die kritischsten Bauelemente werden in speziellen Ausführungen in Hybridbauweise mit der Photodiode integriert.

Außerdem beeinflusst die hohe Verstärkung das Zeitverhalten negativ. Zum Erreichen der

erforderlichen Messfrequenzen werden spezielle Typen mit einem großen Verstärkungs-Bandbreiten-Produkt verwendet.

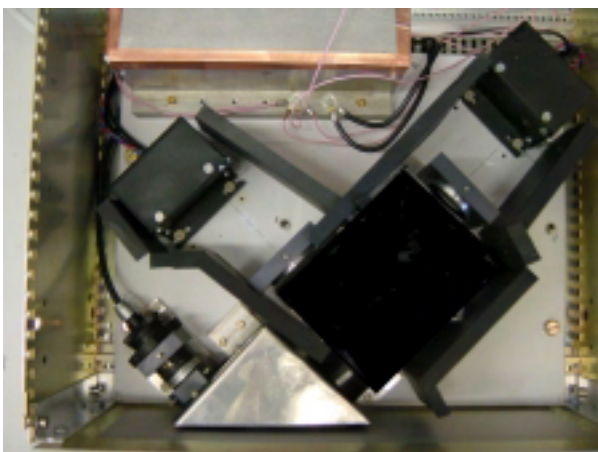


**Abbildung 9: Photodiode mit Verstärker**

In den Messsystemen werden zwei Photodiodenbaugruppen verwendet. Das Verhalten dieser Komponenten kann relativ gut beschrieben werden.

### 3.3 Optisches System

Das optische System besteht aus einer der jeweiligen Applikation angepassten Anordnung von optischen Bauteilen wie Linsen und Prismen (Abb.10). Die Bauteile müssen exakt gelagert und mit speziellen Justageeinrichtungen versehen sein.



**Abbildung 10: Optisches System**

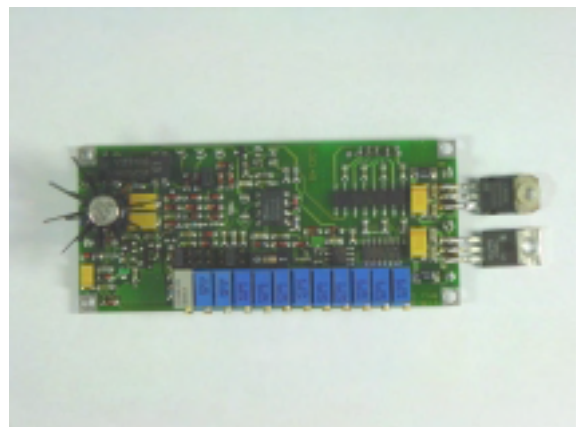
In das optische System ist die Messstrecke, in der sich das Messgut befindet, mit einzubeziehen. Die Übertragungsfunktion wird damit primär durch den eigentlichen Messeffekt bestimmt. Obwohl ein hoher

mechanischer Aufwand betrieben wird, sind Wärmeausdehnung und mechanische Schwingungen ernstzunehmende störende Einflussfaktoren.

Große Aufmerksamkeit muss auf parasitäre Lichtwege gelegt werden, welche zusätzliche, dem eigentlichen Messeffekt nicht zuordenbare Lichtmengen zu den Sensoren bringen können und damit die empfindlichen Messungen stark stören.

### 3.4 Steuerbaugruppe

Die bisher verwendete Baugruppe ist vorrangig aus diskreten Bauelementen aufgebaut (Abb.11). Für die einzelnen Applikationen werden einzelne Bauelemente speziell dimensioniert bzw. Schaltungsteile mit Jumpers zugeschaltet. Eine Kompensation des Temperaturverhaltens der Laserdiode wurde mit einer Reihenschaltung mehrerer Dioden unter Ausnutzung entgegengesetzt gerichteter Temperaturkoeffizienten vorgesehen. Das bedingt eine annähernd gleiche Temperaturbeeinflussung von Laserquelle und Steuerbaugruppe. In der vorliegenden Version funktioniert dieses ansatzweise.



**Abbildung 11: Steuerbaugruppe**

Mittlerweile sind für die Laseransteuerung integrierte Schaltungen verfügbar. Diese sind jedoch vorrangig für Massenanwendungen wie CD-Player oder Barcode-Leser ausgelegt. Demzufolge ist zu prüfen, ob sie funktionell und mit ihrer Einstellgenauigkeit den Anforderungen eines hochempfindlichen Messgerätes entsprechen.

## 4 Regelungproblematik

Wie bereits oben angeführt, besteht die Notwendigkeit die abgegebene Laserleistung zu stabilisieren, um den Einfluss auf die Messwerte zu minimieren. Ziel ist es, gleichbleibende optische Bedingungen für die Messung zu garantieren.

### 4.1 Betriebsart Transmission

Bei den Transmissionsmesssystemen wird entsprechend den optischen Eigenschaften des zu durchstrahlenden Materials eine Helligkeitsstufe ausgewählt, so dass in der Photodiode auf der Empfängerseite ein optimaler Aussteuerbereich für das Messsignal erreicht wird. Durch die Auswertesoftware wird diese Bedingung überwacht und bei Bedarf ein Wechsel der Helligkeitsstufe vorgenommen. Die Helligkeitsstufe auswählen, heißt einen Arbeitspunkt auf der Kennlinie (Abb.7) einzustellen und mit absoluter Konstanz einzuhalten. Zu diesem Zweck wird nahe der Strahlenquelle ein Teil des Laserlichtes ausgeblendet und mit einer zweiten Photodiode gemessen. In dieser Konfiguration kommt es also vorrangig darauf an, die Stabilität beim Umschalten zwischen den Arbeitspunkten zu garantieren und die bei den Komponenten beschriebenen Störeinflüsse auszuregeln. Als Beispiel sei die Abhängigkeit (Abb.12) des Schwellstroms, bei dem die Diode beginnt Laserlicht abzugeben, von der Gehäusetemperatur des Lasers genannt.

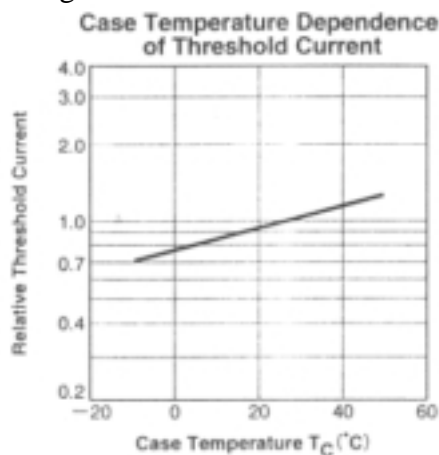


Abbildung 12: Temperaturabhängigkeit

### 4.2 Betriebsart Reflexion

Auch bei den Systemen, die nach dem Reflexionsprinzip arbeiten, ist es notwendig, abhängig vom Absorptionsverhalten des Messgutes, Helligkeitsstufen einzustellen. Außerdem werden aber auch weitere Effekte ausgeregelt. Zum Beispiel dürfen bei der Bestimmung der Oberflächenrauheit langwellige Höhenänderungen oder Abstandsänderungen des Messgutes keinen Einfluss haben. Diese Änderungen erfolgen mit geringer Frequenz im Bereich einiger Hertz. Der eigentliche Messeffekt ändert sich im Bereich von 100 bis 200 kHz. Dieses Zeitverhalten muss bei der Regelung beachtet werden. Vom optischen System werden für die Messung Lichtanteile gebildet. Um optimale Messbedingungen zu erreichen, wird bei der Lasersteuerung die Summe dieser Lichtanteile konstant gehalten. Auch in diesem Einsatzfall müssen die unbeabsichtigten Einflüsse auf die Systemkomponenten ausgeregelt werden.

### 4.3 Realisierung

Die Lasersteuerung hat folgende Aufgaben:

- Ansteuerung der Laserdiode im gewünschten Arbeitspunkt
- Regelung im Arbeitspunkt unter Beachtung der möglichen Einflüsse der Komponenten des Systems und ihrer Störungen
- Regelung im Arbeitspunkt unter Beachtung der Einflüsse der optischen Eigenschaften des Materials

Ausgangspunkt für die Untersuchungen im Projekt ist eine vorliegende Grundschaltung der Steuerbaugruppe. Diese realisiert zwei Regelkreise.

Zum Einen wird der Strom der internen Monitordiode rückgekoppelt. Im Idealfall sollte das im Normalbetrieb keinen Einfluss bringen und nur im Übersteuerungsfall den Laserstrom zum Schutz der Diode drastisch zurückregeln. Im Realfall beeinflusst dieser Schaltungsteil auch bei Normalbetrieb den Laserstrom.

Zum Anderen werden die Referenzspannungen der externen Photodioden rückgekoppelt. Eine Reduzierung der Referenzspannung durch die Verschlechterung der optischen Eigenschaften des Messguts soll durch eine Erhöhung des Laserstromes kompensiert werden. In diese Rückführung wird eine konstante, wählbare Offsetspannung einbezogen, die die Einstellung der Helligkeitsstufen realisiert.

Der Strom zur Ansteuerung der Laserdiode wird in einem Summationspunkt mit beiden Einflüssen beaufschlagt.

Für Justagezwecke ist es sinnvoll, die Regelung außer Betrieb zu setzen und mit der maximalen Laserleistung zu strahlen. Die Schutzfunktion für die Laserdiode muss auch in dieser Betriebsart gewährleistet sein.

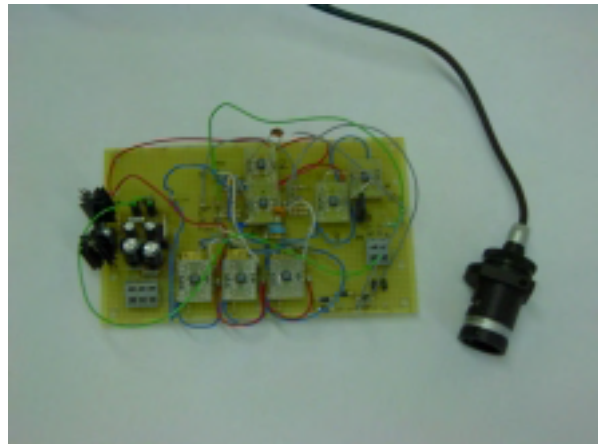
Weiterhin ist es möglich, durch Rückregelung des Laserstroms unter die Laserschwelle einen Betriebszustand zu erreichen, bei dem kein Laserlicht abgegeben wird. Somit kann in Messpausen die Laserdiode geschont werden, ohne durch Ausschalten der gesamten Steuerung das Temperaturgleichgewicht wesentlich zu beeinflussen.

## 5 Modellierung, Simulation, Verifizierung anhand eines Funktionsmusters

Das vorliegende Modell der Lasersteuerung [3] basiert auf der Beschreibung der Schaltungsblöcke der bisher verwendeten Baugruppe sowie der Beschreibung der Anforderungen an die Funktionalität. Die Analyse der verschiedenen Betriebszustände des Messsystems liefert weitere Basisdaten für die Modellierung.

Im ersten Modellansatz sind umfassend alle potentiellen Einflussfaktoren berücksichtigt. Den Schwerpunkt bilden dabei eindeutig die elektrischen Eigenschaften des Systems. Entsprechend liefert das Modell die Grundlage für den Aufbau einer neuer Version der Steuerbaugruppe.

Die Dimensionierung der einzelnen Schaltungselemente ist anhand der Ergebnisse einer ersten Simulation festgelegt. Der abgebildete Laboraufbau dient zur Verifizierung der Simulationsergebnisse.



**Abbildung 13: Laboraufbau**

In den nachfolgenden Entwicklungstufen wird das Modell verfeinert. Die Kontrollmessungen zeigen eine zunehmend bessere Übereinstimmung der erzielten Simulationsergebnisse mit dem realen Verhalten der Schaltung [4]. Weiterhin liefern die Simulationsrechnungen gute Aussagen für die Realisierbarkeit von Schaltungsauslegung im Zusammenhang mit Sensorvarianten, für die die entwickelte Baugruppe ebenfalls verwendbar ist. Das neu entwickelte Layout ist das Ergebnis der Produktoptimierung in den vorangegangenen Entwicklungsschritten.



**Abbildung 14: Neues Layout**

Wesentlich schwieriger ist die Modellierung und Simulation der nichtelektrischen Einflüsse auf die Stabilität der Laserregelung, wie sie bereits oben beschrieben wurden. Im vorliegenden Modell sind als erster Ansatz einfache Übertragungsfunktionen für die mechanischen und thermischen Einflüsse verwendet worden. Die Variation der Parameter wurde grob abgeschätzt.

Genauere Beschreibungen müssen anhand komplizierter Messungen ermittelt werden, da von den Herstellern der Komponenten keine geeigneten Beschreibungen vorliegen. Außerdem wird das Verhalten wesentlich von den sehr speziellen Systemaufbauten bestimmt.

Bei diesen Messungen kommt es darauf an, jeweils nur den zu untersuchenden Parameter zu verändern und alle anderen konstant zu halten.

Im Zuge der theoretischen und praktischen Entwicklungsschritte ist die Relevanz der Einflüsse und damit die Notwendigkeit der Einbeziehung in das Modell zu prüfen.

Die genannten Untersuchungen und die Einbeziehung der Ergebnisse in das Modell sind Gegenstand der Schlussphase des Projektes und werden darüber hinaus fortgeführt werden müssen.

## 6 Schlussfolgerung und Ausblick

Als Ergebnis des Entwicklungsprojektes OMID liegt eine optimierte Lasersteuerung für hochpräzise optische Messgeräte vor. Die Entwicklung basiert auf enger Zusammenarbeit von Modellierung, Simulation, Optimierung und Musteraufbau.

Die entwickelten Werkzeuge können zur Optimierung der Dimensionierung der Schaltungsvarianten auf der Steuerbaugruppe verwendet werden.

Weiterhin gibt es Ansatzpunkte für die Verwendung der Projektergebnisse bei vergleichbaren Aufgabenstellungen in anderen Messtechniken. In einem weiteren Arbeitsgebiet der VacuTec Meßtechnik GmbH werden andere hochstabile Strahlungsquellen wie Röntgenstrahler und IR-Strahler benötigt. Für diese Einsatzfälle gilt es die Projektergebnisse und Projekterfahrungen anzuwenden und zu verallgemeinern.

Wir danken den Kollegen des Fraunhofer-Instituts für Integrierte Schaltungen, Außenstelle EAS Dresden, insbesondere Herrn Dr. Schwarz und Herrn Dr. Trappe, für die gute Zusammenarbeit.

## Literatur

- [1] Schwarz,P.;Trappe,P.: Modellierung und Simulation einer Lasersteuerung (November 2001)
- [2] Schäfer&Kirchhoff.: Datenblatt der Laserdiode TOLD9140(S), 1997
- [3] Trappe,P.: Bericht zur Laserregelung – OMID-Seminar (Mai 2002)
- [4] Trappe,P.: Laserregelung – Simulation nach Schaltungsaufbau (Juli 2002)

*Das diesem Beitrag zugrunde liegende Vorhaben „OMID-Optimierung von Mikrosystemen für Diagnose- und Überwachungseinrichtungen“ wird mit Mitteln des Bundesministeriums für Bildung, Wissenschaft, Forschung und Technologie (BMBF) unter dem Kennzeichen 16 SV 988/2 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt allein bei den Autoren.*

# MODELLIERUNG, SIMULATION UND OPTIMIERUNG EINER LASERSTEUERUNG

P. Schwarz, P. Trappe, P. Schneider  
Fraunhofer-Institut für Integrierte Schaltungen,  
Außenstelle Entwurfsautomatisierung EAS Dresden  
E-Mail: peter.schwarz@eas.iis.fhg.de

## Kurzfassung

Für den rechnergestützten Entwurf und die optimale Auslegung einer Lasersteuerung in einem optischen Meßsystem ist die Modellierung der optischen, elektrischen und thermischen Eigenschaften aller Teilsysteme und ihrer Wechselwirkungen erforderlich. Zwei vermaschte nichtlineare Regelkreise sind zu modellieren und ihre Parameter modellgestützt einzustellen. Es werden Modellierungsansätze vorgestellt, die eine Gesamtsystem-Simulation und -Optimierung gestatten. Als Werkzeuge werden dabei der Schaltungs- und Systemsimulator Saber sowie das Internet-basierte Optimierungssystem MOSCITO eingesetzt.

## 1 Einleitung

In der Firma VacuTec Meßtechnik GmbH Dresden wird ein Meßgerät zur optischen Messung von Oberflächenqualität und Dicke bzw. Dichte von Stoff- und Papierbahnen entwickelt [1]. Der Einsatz rechnergestützter Entwurfsmethoden, vor allem der Optimierung, soll den Entwurfsprozeß effektiveren, die Eigenschaften des Gerätes verbessern und die weitere Miniaturisierung unterstützen [16]. Wie häufig in der Mikrosystemtechnik und Mechatronik müssen hierbei elektrische, thermische, optische und künftig auch mechanische Effekte in ihrem engen Zusammenwirken modelliert werden. Ein wichtiges Teilsystem ist die Lasersteuerung, bei deren rechnergestütztem Entwurf für die Optimierung zunächst die Modellierung und Simulation des gesamten, sehr heterogenen Systems nötig ist. Die Modellierung ist kritisch, da es sich um ein stark nichtlineares dynamisches System handelt. Die

Modelle müssen sehr genau sein, da die Regelkreise im Meßsystem modellbasiert optimiert werden. Auch die Formulierung der Gütefunktion (Zielfunktion) für die Optimierung ist nicht einfach, weil mehrere, sich z.T. widersprechende Entwurfsziele berücksichtigt werden müssen.

## 2 Meßprinzip und Geräteaufbau

In Bild 1 und Bild 2 sind die im Meßgerät verwendeten Prinzipien (*Durchstrahlung* und *Reflexion*) dargestellt. Sie sind als Varianten in der Lasersteuerungs-Schaltung (und daher auch bei deren Optimierung) zu berücksichtigen. Der vom Laser erzeugte Lichtstrahl wird von zwei Fotodioden FD1 und FD2 aufgenommen. Aus der Differenz der beiden Signale wird der Meßwert abgeleitet. Zusammen mit einem weiteren Signal aus der Monitordiode, s. Bild 3, dient er der Steuerung der Laserdiode. Probleme bereiten beim Entwurf vor allem die sehr enge Wechselwirkung zwischen den elektrischen, optischen und thermischen Teilsystemen, die die Meßgenauigkeit verringern und Stabilitätsprobleme hervorrufen können. Die Entwurfsziele sind:

- hohe Stabilität der optischen Leistung bezüglich Temperatur und Alterung der Laserdiode,
- gesicherte Begrenzung der optischen Leistung in der Laserdiode,
- Reduzierung des Rauschens,
- optimales Einschwingverhalten des Regler für eine hohe Meßgeschwindigkeit,
- Miniaturisierung.

Zu den Entwurfszielen gehört weiterhin die Vorbereitung des Einsatzes eines integrierten Schaltkreises (AD 9660) als Ersatz für eine eigene Leiterplattenschaltung. Da dieser Schaltkreis nur sehr unvollständig dokumentiert ist, bereitet seine Modellierung Schwierigkeiten - eine typische Situation beim Entwurf technischer Systeme.

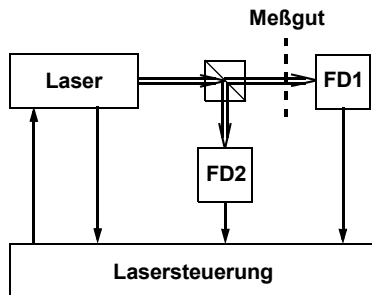


Bild 1: Transmission

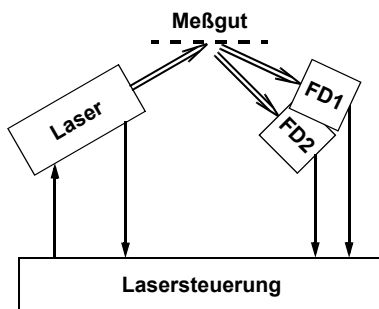


Bild 2: Reflexion

### 3 Modellierung und Simulation

Für die Modellierung und Simulation haben sich folgende Probleme herausgestellt:

- Einige Systemkomponenten sind extrem nichtlinear.
- Die Regelung erfolgt durch geeignet dimensionierte *Schaltungen*, deren nichtlineares und dynamisches Verhalten funktionsbestimmend ist. Da Teile der Schaltung auch auf *Transistorebene* simuliert werden müssen, scheidet Matlab/Simlink als regelungstechnischer Simulator aus. Stattdessen wird Saber als Schaltungs- und System Simulator eingesetzt.
- Die zu optimierenden Regler arbeiten aus Gründen der Meßdynamik sehr dicht an ihrer

Stabilitätsgrenze. Daraus ergeben sich hohe Anforderungen an die Modellierungsgenauigkeit der Reglerkomponenten und des Gesamtsystems, da eine simulationsbasierte Optimierung sonst sinnlos wäre.

- Für einige elektrische und thermische Parameter stehen nur Kennangaben zur Verfügung und müssen durch Messungen ergänzt werden.

Weiterhin ist für den auf seine Eignung hin zu untersuchenden Schaltkreis AD 9660 (er wird vorzugsweise in Abtastsystemen für CD-Player eingesetzt) ein *Makromodell* zu entwickeln. Dieses soll sowohl der Simulation dienen als auch zum Verständnis der Schaltung beitragen.

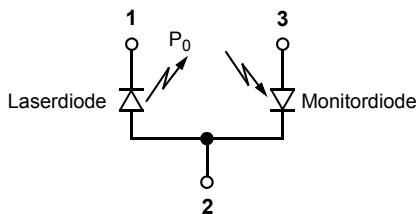
Als *einheitlicher Modellierungsansatz* werden verallgemeinerte Kirchhoffsche Netzwerke [3], [4], [5], [6], [7], [8] verwendet, die sich besonders gut für Multi-Physics-Systeme (einschließlich Regelungssysteme) eignen und für die es leistungsfähige Simulatoren gibt. Vor allem bei Systemen, die auch nichtelektrische (z.B. thermische und mechanische) Teilsystemen einschließen, hat sich dieser einheitliche Modellierungs- und Simulationsansatz bewährt. Künftig wird dieses Vorgehen auch durch die Entwicklung leistungsfähiger Modellierungssprachen (z.B. VHDL-AMS und Modelica) und zugehöriger Simulatoren noch besser unterstützt.

Aufbauend auf den Arbeiten zur Modellierung und Simulation (sowie zum manuellen Reglerentwurf) wird eine simulationsgestützte Optimierung durchgeführt. Dabei wird das Optimierungssystem MOSCITO [9], [10] eingesetzt, das eine Schnittstelle zum Simulator Saber besitzt. In MOSCITO lassen sich die im Verbundprojekt OMID von den Partnern Forschungszentrum Karlsruhe, Universität Bremen und FhG EAS Dresden entwickelten Optimierungsverfahren nutzen [11], [12]. Die Untersuchungen zur Art des Optimierungsproblems (dazu gehören auch die Definition der Optimierungskriterien und der Zielfunktion, Nebenbedingungen, Multimodalität...) sind z.Z. noch nicht abgeschlossen.

### 3.1 Laserdiode

Die Baugruppe Laserdiode (Bild 3) [13] enthält zusätzlich zur Laserdiode noch eine Fotodiode („Monitordiode“), die zur Überwachung der Lichtleistung eingesetzt wird. Entsprechend besteht das Verhaltensmodell *vld* (s. Quelltext des Modells in der zum Simulator Saber gehörenden Modellbeschreibungssprache MAST) aus den beiden Teilen:

- Generierung der optischen Leistung  $P_o$  in Abhängigkeit von Ansteuerstrom und Temperatur. Das dynamische Verhalten wird durch einen Tiefpaß 1. Ordnung modelliert. Im MAST-Modell wird dafür im equations-Teil eine Differentialgleichung benutzt, am Operator *d\_by\_dt(...)* zu erkennen.
- Monitoring des optischen Signals.



**Bild 3:** Aufbau der Baugruppe Laserdiode

Das Modell besitzt daher neben den drei elektrischen (ungerichteten) Klemmen noch eine gerichtete Klemme, die die optische Leistung  $P_o$  trägt.

Für Simulationen im Frequenzbereich sind zur optischen Leistung und zum Monitorstrom Rauschquellen hinzugefügt. Bei den Transient-Simulationen werden diese Quellen nicht berücksichtigt.

Saber-Quelltext des Modells:

```

template vld pin1 pin2 Po pin3 = td, se, ithd, t1, ir1, sm, tr
electrical pin1, # Laser Diode Cathode
            pin2, # Laser Diode Anode, Photodiode Cathode
            pin3 # Photodiode Anode
output nu Po # Optical Output Power

number td = 25, # nominal temperature [C]
        se = 1, # slope efficiency [W/A]
        ithd = 40e-3, # threshold current [A]
        t1 = -20, # temperature for ir1 [C]
        ir1 = 0.63, # relative threshold
                    # current for t1 [-]
        sm = 5e-3, # slope of monitor diode [A/W]
        tr = 1e-6 # rise and fall time [s]
external number temp # temperature
    
```

```

{
number dse = 0, # slope efficiency temperature
        # variation [%/C]

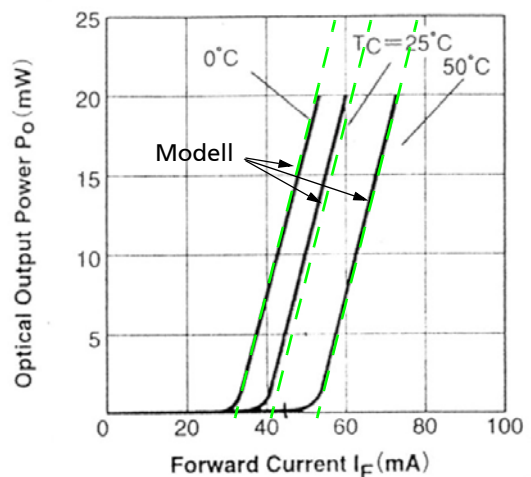
        ut3 = 3*0.02585,
        is = 35e-15,
        k = 1.38e-23,
        qe = 1.6e-19,
        kith

branch vIm = v(pin3, pin2), iIm = i(pin3, pin2)
val nu ith, popt0
var nu popt
var i iin
val v vin
val ni nsi_pd
val nu ns_po

parameters {
    kith = ln(ir1)/(t1-td)
}
values {
    vin = v(pin2, pin1)
    ith = ithd*limexp(kith*(temp-td))
    if (iin < ith) {
        popt0 = (iin-ith)/1g
    }
    else {
        popt0 =
            (iin-ith)*(se*(1+(dse*(temp-td)/100)))
    }
    nsi_pd = 1e6*sqrt(abs(2*qe*iIm))
    ns_po = 1u
}
equations {
    popt: popt0-popt = d_by_dt(tr*popt)
    i(pin2, pin1) += iin
    iin: iin = is*(limexp(vin/ut3)-1) + vin/1g
    Po = popt
    iIm = -sm * popt
}
control_section {
    noise_source (nsi_pd, pin3, pin2)
    noise_source (ns_po, popt)
}
}
    
```

Das Diagramm in Bild 4 zeigt die gute Übereinstimmung von Modell (optische Leistung als Funktion des Laserdioden-Stromes und der Temperatur) und Datenblatt-Kennlinien [13], die sich in der graphischen Darstellung kaum voneinander trennen lassen.

**Optical Output Power vs. Forward Current**

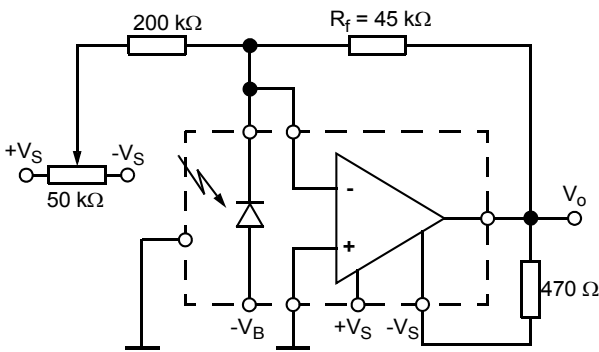


**Bild 4:** Vergleich von Modell und Datenblatt



### 3.2 Fotodiode

Für die Fotodiode [14] wurden zwei Modelle entwickelt. Im Vergleich zum dynamischen Verhalten der Laserdiode mit ihrer relativ hohen Verlustleistung müssen bei der Fotodiode keine dynamischen Effekte modelliert werden.



**Bild 5:** Fotodiode mit Verstärker

- Modell einer Fotodiode mit Verstärker (Bild 5)

$$V_o = r_f \cdot P_o \cdot R_f$$

$r_f$  - responsivity [A/W]  
 $P_o$  - optische Leistung [W]  
 $R_f$  - Meßwiderstand [ $\Omega$ ]

- Modell einer einfachen Fotodiode

$$I_f = r_f \cdot P_o$$

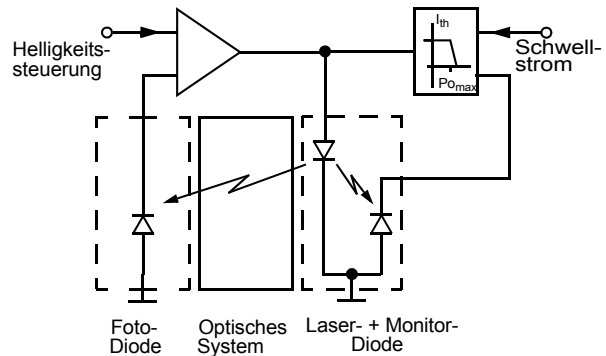
Der Parameter  $r_f$  wird aus Datenblattangaben entnommen.

### 3.3 Optisches System

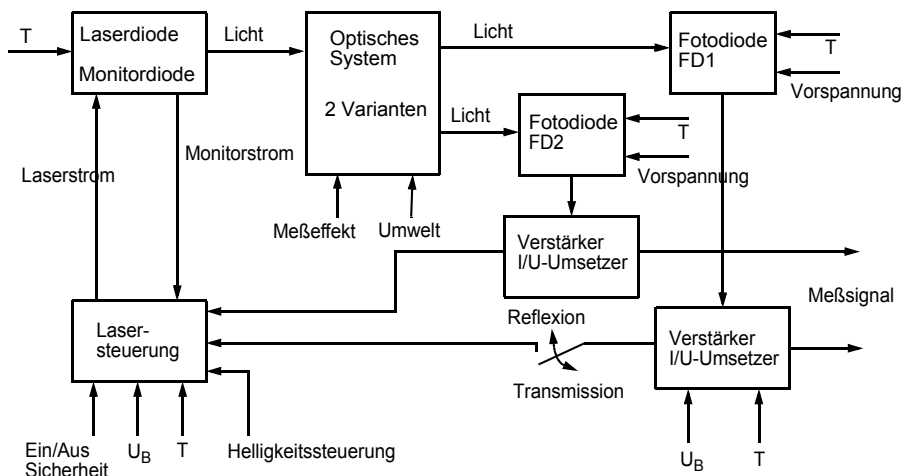
Das Modell des optischen Systems besteht gegenwärtig lediglich aus einer steuerbaren Dämpfung. Künftig werden hier aber die optischen Eigenschaften des Meßgutes berücksichtigt. In das Modell können dann auch einige mechanische Eigenschaften der Anlage einbezogen, da diese für die Dynamik und die Genauigkeit der Meßanordnung wichtig sind. Beispielsweise können dann mechanische Schwingungen in die Untersuchungen einbezogen werden.

### 3.4 Gesamtmodell

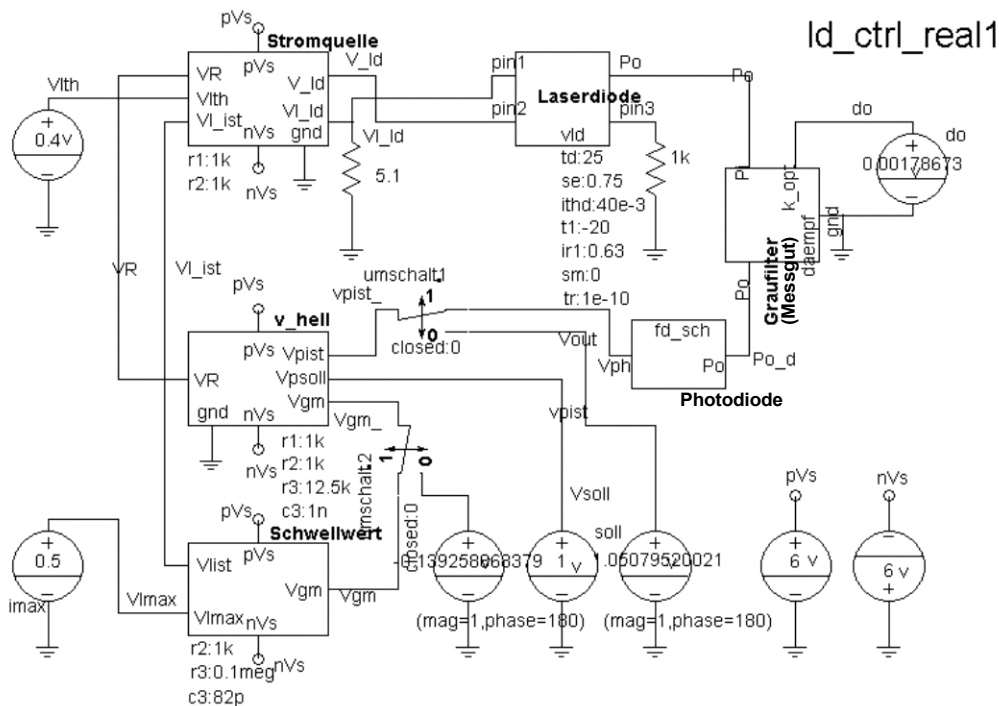
Ein wesentlicher Teil der Meßanordnung (Blockschaltbild: Bild 6) ist die *Laserregelung*, deren Prinzip in Bild 7 als Transistor-Schaltbild in der Form dargestellt ist, wie sie in den Simulator Saber mit einem Schematic-Entry-Tool graphisch eingegeben wird. Dabei ist nur



**Bild 7:** Laserregelung



**Bild 6:** Gesamtsystem: Prinzip-Blockschaltbild



**Bild 8:** Saber-Schematic der Transistorschaltung

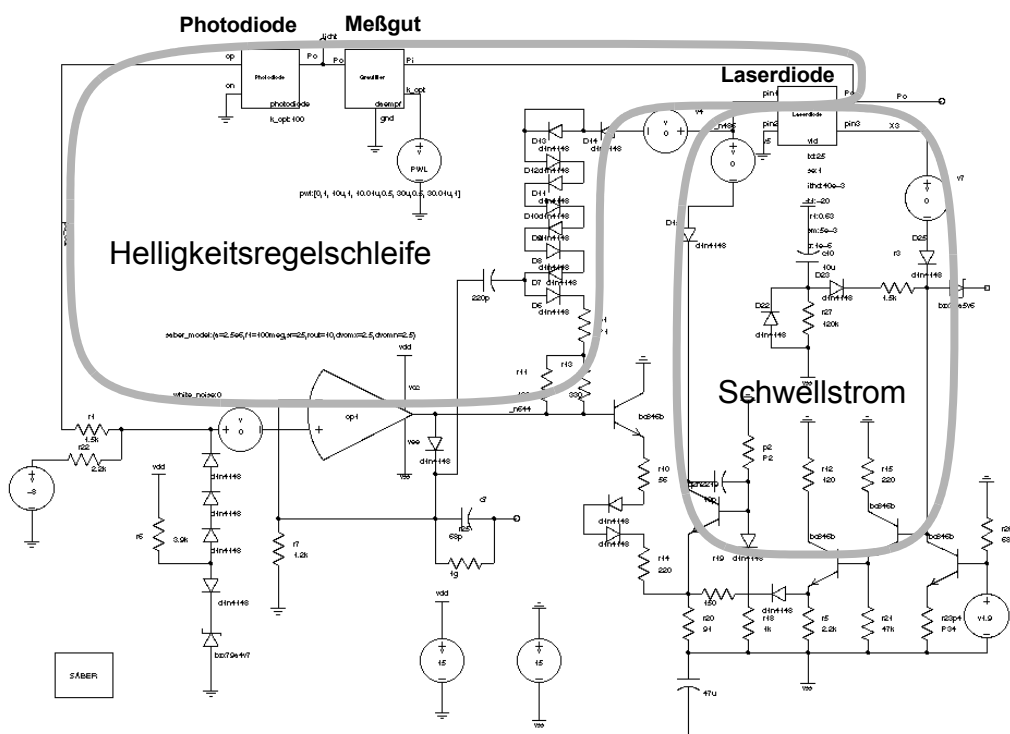
die oberste Hierarchieebene dargestellt; die einzelnen Blöcke sind durch Sub-Schaltungen weiter untersetzt.

Die Laserregelung besteht aus zwei Regelkreisen:

- *Helligkeitsregelschleife*, die für eine konstante Helligkeit am Ausgang des optischen Systems verantwortlich ist:

Laserdiode - Optisches System - Fotodiode - Verstärker

- *Schwellstromregelschleife*, die den Schwellstrom einstellt und die optische Leistung (Laserdiodenstrom) begrenzt, um eine Zerstörung der Laserdiode zu verhindern: Laserdiode - Monitordiode - Verstärker und Begrenzer



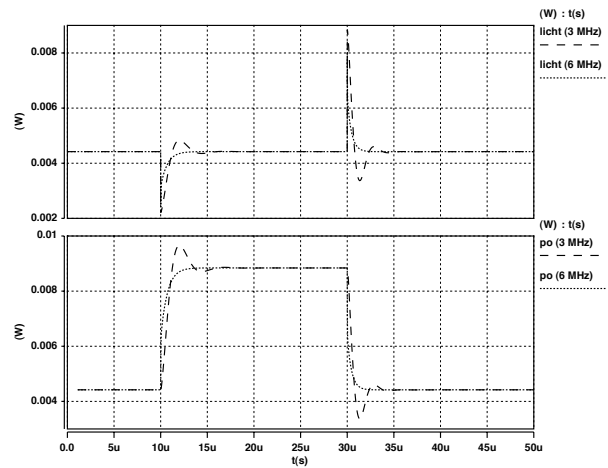
**Bild 9:** Saber-Modell mit gekennzeichneten Regelschleifen

Die Simulation auf einer noch abstrakteren Blockschaltbild-Ebene (z.B. Modellierung als rückwirkungsfreie Zusammenschaltung von Blöcken für den Simulator Matlab/Simulink) wäre nur für überschlägige Analysen sinnvoll. Die kritischen nichtlinearen Eigenschaften der Bauelemente lassen sich am effektivsten auf der Transistorschaltungs-Ebene berücksichtigen.

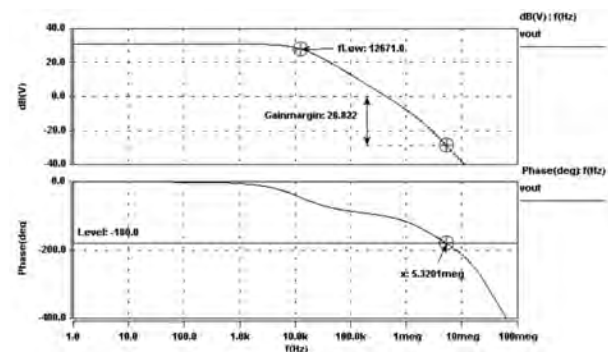
Daher wird der *Schaltungssimulator* Saber eingesetzt. Er besitzt zusätzlich den Vorteil, daß mit Hilfe der Modellierungssprache („Hardware-Beschreibungs-Sprache“) MAST das Verhalten nichtlinearer dynamischer Elemente beschrieben werden kann, was bereits im Abschnitt 3.1 bei der Modellierung der Laserdiode ausgenutzt wurde. In Ergänzung zum Transistorlevel-Modell der Regelung (Bild 8) ist eine (etwas vereinfachte) Transistorschaltung mit eingezeichneten Regelschleifen im Bild 9 dargestellt. Eine Besonderheit des Systems besteht in einer Verkopplung der nichtlinearen Regelschleifen, die beide die Baugruppe „Laserdiode“ einschließen.

Durch Simulation einiger typischer Betriebsfälle wurde die Brauchbarkeit der Modelle überprüft. Gleichzeitig ergeben sich daraus wesentliche Informationen zu einer manuellen Einstellung der Parameter der Regelkreise (als Vorstufe einer rechnergestützten Optimierung). Simulationsergebnisse mit dem Blockmodell und der Transistorschaltung sind in den folgenden Bildern wiedergegeben:

- Bild 10: Regelverhalten der Helligkeitsregelschleife nach einer sprunghaften Änderung der optischen Dämpfung bei zwei verschiedenen Verstärkungs-Bandbreite-Produkten des Operationsverstärkers.
- Bild 11: Frequenzgang der offenen Helligkeitsregelschleife der Transistorschaltung. Daraus sind auch die für die Stabilität des Reglers wichtigen Parameter Verstärkungsreserve (gain margin) und Phasenreserve (phase margin) zu ermitteln.

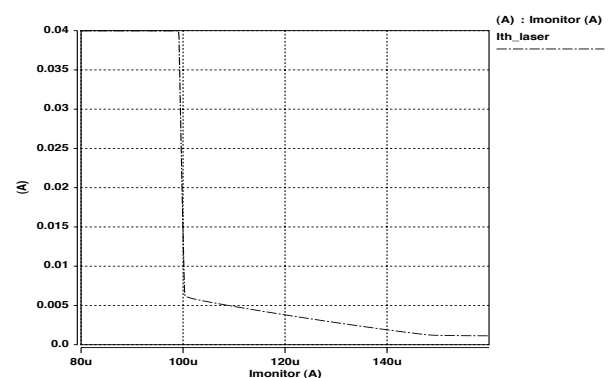


**Bild 10:** Zeitverhalten der Helligkeitsregelung ( $f_1 = 3$  bzw. 6 MHz)



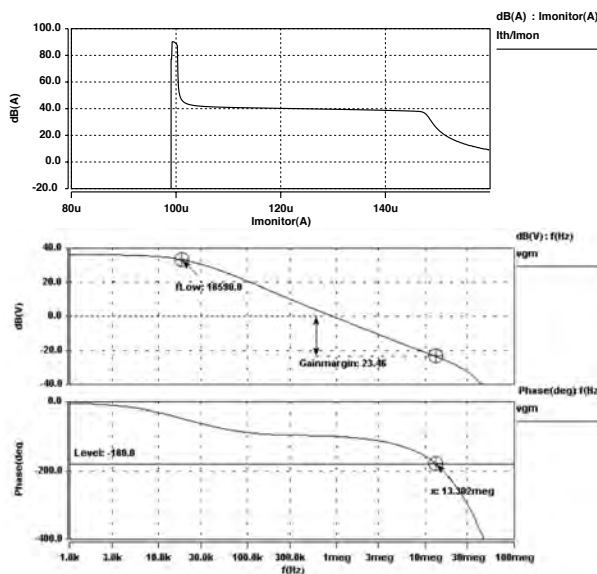
**Bild 11:** Frequenzgang der offenen Helligkeitsregelschleife

- Bild 12: Gleichstromverhalten (DC-Kennlinie) der Schwellstromregelschleife. Dazu wurde die Monitordiode durch eine Stromquelle  $I_{\text{monitor}}$  ersetzt und der Strom in der Laserdiode gemessen. Deutlich ist der steile Abfall des Schwellstromes von 40 auf 5 mA, wenn der Monitorstrom 100  $\mu\text{A}$  erreicht.



**Bild 12:** DC-Kennlinie der Schwellstromregelschleife

- Bild 13: zeigt die DC-Verstärkung des offenen Schwellstromregelkreises und darunter den Frequenzgang.



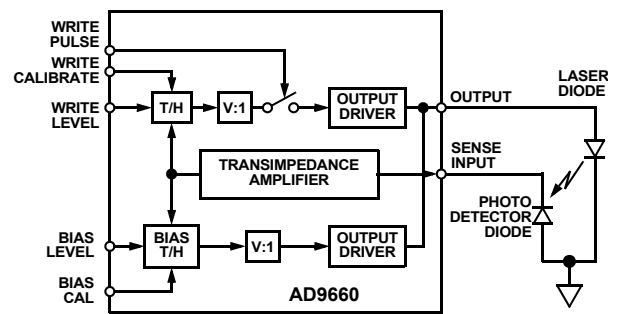
**Bild 13:** DC-Verstärkung und Frequenzgang der offenen Schwellstromregelschleife

Aus den statischen Kennlinien in Bild 12 und Bild 13 sind die starken Nichtlinearitäten zu erkennen, die sich aus den nichtlinearen Eigenschaften der Transistoren und Dioden ergeben. Sie lassen sich am einfachsten auf der Transistorschaltungs-Ebene erfassen.

### 3.5 Modellierung des AD9660

Der Ersatz einer gegenwärtig eingesetzten Leiterplatte durch einen IC AD9660 kann ein wesentlicher Schritt zur Qualitätssteigerung der Steuerung sein, unter der Voraussetzung, daß sich ein optimales Reglerverhalten einstellen läßt. Gleichzeitig ist es ein Schritt zur Miniaturisierung.

Für den Schaltkreis AD9660 liegen keine vollständigen Schaltungsunterlagen vor. Aus groben und etwas detaillierteren Angaben auf Blockschaltungs-Niveau (dargestellt in Bild 14 und Bild 15) werden gegenwärtig Modelle entwickelt, deren Parameterwerte durch Vergleich mit Messungen gewonnen werden müssen. Für die Parameteranpassung der Modelle an die Messwerte wird ebenfalls das Optimierungssystem MOSCITO eingesetzt.



**Bild 14:** AD9660 - Functional Block Diagram

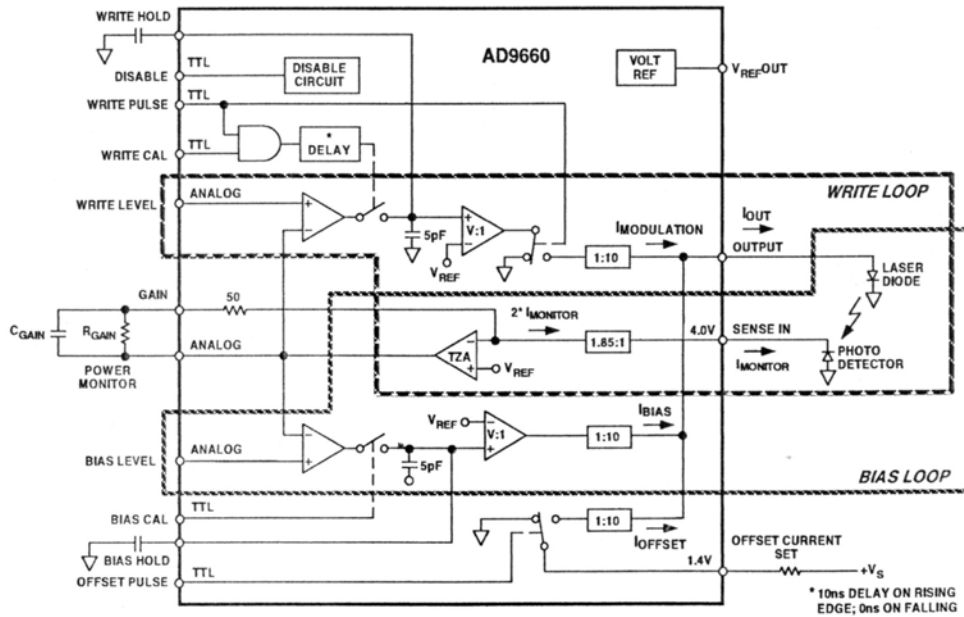
## 4 Optimierung

Der Einsatz von Optimierungsverfahren erfolgt sowohl für die Bestimmung von Parametern für die Modelle der eingesetzten Baugruppen als auch vor allem für die Dimensionierung der Regelkreise.

Als Optimierungsparameter für den Entwurf der Regelkreise stehen 4 Einflussgrößen zur Verfügung: die Verstärkungsfaktoren und die Zeitkonstanten der beiden Regler. Schwierigkeiten bereitet die Definition einer Gütefunktion: die *Verstärkungsreserven* und die *Grenzfrequenzen* beider Regelkreise sollen jeweils möglichst groß sein. Das erfordert eigentlich eine multikriterielle Optimierung, für die im System MOSCITO aber gegenwärtig kein Verfahren bereitsteht. Außerdem liegen Erfahrungswerte bei den Schaltungsentwicklern vor, die ebenfalls noch nicht in die Gütefunktion einbezogen werden können. Daher wurde ein zweistufiges Vorgehen gewählt:

1. Zunächst werden Simulationsexperimente mit verschiedenen Parameterwerten vorgenommen und der Entwickler entscheidet „nach seinem Gefühl“, dass die erreichten Werte „optimal“ sind. Auf diese Weise wurden die in Bild 11 und Bild 13 dargestellten Systemcharakteristiken gewonnen.

2. Mit den vier einzelnen Gütekriterien wird (z.B. durch gewichtete Mittelung) eine einzige Gütefunktion definiert und mit den in MOSCITO bereitstehenden Algorithmen eine mathematische Optimierung durchgeführt. Diese Arbeiten waren bei Redaktionsschluss dieses Beitrages noch nicht abgeschlossen.



**Bild 15:** AD9660 - Functional Block Diagram (2)

Bei dem unter 1. beschriebenen Vorgehen gab es bisher keine Anzeichen für eine besonders komplizierte Gestalt des mathematischen Optimierungsproblems, so dass Probleme zunächst nur bei dem Experimentieren mit der Definition der einheitlichen Gütefunktion erwartet werden.

Es sei an dieser Stelle ausdrücklich betont, dass das geschilderte Vorgehen keine Notlösung, sondern typisch für das Einführen mathematischer Optimierungsverfahren in den Entwurfsprozess technischer Systeme ist. Der Entwickler wird am ehesten durch erfolgreiche simulationsgestützte Experimente und das Einbeziehen seiner Erfahrung zu überzeugen sein, den nicht geringen Aufwand zur Modellbildung und zum Beschäftigen mit Optimierungssystemen zu erbringen.

## 5 Zusammenfassung

Die Modellierungsarbeiten mit verallgemeinerten Netzwerken und die Simulationen auf Schaltungs- und Systemebene zeigen, dass eine simulationsbasierte Optimierung der La-

sersteuerung möglich ist. Bereits bei einigen simulationsgestützten Variantenuntersuchungen haben sich wichtige Erkenntnisse über die Regeleigenschaften der Lasersteuerung ergeben. Die nächsten Schritte werden die Modellierung der mechanischen Komponenten, die Makromodellierung des Schaltkreises AD9660 und die flexible Definition von Gütefunktionen für die mathematische Optimierung des Gesamtsystems sein.

Wenn die mit der Modellierung und der Optimierung verbundenen Probleme gelöst werden können, ist für die Firma VacuTec Meßtechnik GmbH Dresden ein wichtiger Schritt bei der Einführung rechnergestützter Entwurfsverfahren für miniaturisierte Messgeräte gelungen.

Wir danken den Kollegen von VacuTec Dresden, insbesondere Herrn Gärtner und Herrn Dittrich, für die gute Zusammenarbeit.

*Das diesem Beitrag zugrunde liegende Vorhaben „OMID - Optimierung von Mikrosystemen für Diagnose- und Überwachungsanwendungen“ wird mit Mitteln des Bundesministeriums für Bildung, Wissenschaft, Forschung und Technologie (BMBF) unter dem Kennzeichen 16 SV 988 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt allein bei den Autoren.*

## 6 Literatur

- [1] Gärtner, G.; Dittrich, T.: Eine Lasersteuerung für optische Messgeräte. Diese Tagung.
- [2] G. Gerlach, W. Dötzel: Grundlagen der Mikrosystemtechnik. München: Hanser 1997.
- [3] Fischer, W.-J. (Hrg.): Mikrosystemtechnik. Würzburg: Vogel 2000.
- [4] Voelklein, F.; Zetterer, T.: Einführung in die Mikrosystemtechnik. Braunschweig: Vieweg 2000.
- [5] Kasper, M.: Mikrosystementwurf - Entwurf und Simulation von Mikrosystemen. Berlin: Springer, 2000.
- [6] Romanowicz, B. F.: Methodology for the Modeling and Simulation of Microsystems. Dordrecht: Kluwer 1998.
- [7] Wachutka, G.: Tailored modeling: a way to the 'virtual microtransducer fab' ? Sensor and Actuators A 46-47 (1995), pp. 603-612.
- [8] Schwarz, P.: Physically oriented modeling of heterogeneous systems. Mathematics and Computers in Simulation 53 (2000), 333-344.
- [9] Schneider, A.; Schneider, P.; Bastian, J.: Moscito - Ein modulares, Internetbasiertes Programmsystem für die Optimierung von Mikrosystemen. Diese Tagung.
- [10] Schneider, P.; Huck, E.; Reitz, S.; Parodat, S.; Schneider, A.; Schwarz, P.: A modular approach for simulation-based optimization of MEMS. Proc. ISMA2000, Singapore, December 2000, 71-82.
- [11] Peters, D.: Unterstützung des Entwurfsprozesses von Mikrosystemen mit direkten Optimierungsverfahren. Dissertation Universität Bremen, ITEM, 2001
- [12] Schneider, A.; Schneider, P.; Peters, D.; Jakob, W.; Quinte, A.: Simulationsgestützte Optimierung von heterogenen Systemen. Workshop "Multi-Nature-Systems 2001", Hamburg, 21.02.01, S. 43-50
- [13] Schäfer & Kirchhoff: Datenblatt der Toshiba Laserdiode TOLD 9140, 1997
- [14] Datenblatt: Schnelle Siliziumphotodiode HFD-Serie (1060, 1100)
- [15] Analog Devices: 200 MHz Laser Diode Driver with Light Power Control, Data sheet, 1995
- [16] Schwarz, P.; Trappe, P.: Modellierung und Simulation einer Lasersteuerung. Proc. OMID-Statusseminar, Bremen, Oktober 2001, S. 71-78

# "IR-GASSENSOR FÜR DIE UMWELT- UND GEBÄUDETECHNIK"

Karl-Heinz Suphan, Holger Urban, Achim Berger

Micro-Hybrid Electronic GmbH, Hermsdorf

Email: suphan@micro-hybrid.de

## Kurzfassung

Anhand eines CO<sub>2</sub>-NDIR-Gassensors werden verschiedene Aspekte der Signalaufbereitung optischer Low cost - Systeme und der Ergebnisberechnung in Sensorsystemen mit mehrfach verkoppelten Eingangsgrößen vorgestellt und diskutiert.

## 1 Einleitung

Die Überwachung bzw. Messung der Konzentration von verschiedenen Gasen gewinnt im Umweltbereich und in der Gebäudetechnik stetig an Bedeutung. Gilt es beim Einsatz in der Umwelttechnik zum Teil für Mensch und Umwelt kritische Gase hinsichtlich ihrer Konzentration zu registrieren und zu bewerten, so strebt die Gebäudetechnik durch Realisierung entsprechender Regelalgorithmen einen effektiven Energieeinsatz bei Beibehaltung eines optimalen Raumklimas an. Hierfür ist die Messung und Überwachung der wesentlichen Raumluftbestandteile eine Voraussetzung. Für diese Meßaufgaben haben sich neben chemischen Sensoren IR-optische Meßsysteme etabliert. Diese zeichnen sich im Anwendungsverhalten bei einem einfachen Aufbau durch eine hohe Systemstabilität und geringe Querempfindlichkeiten gegenüber anderen Gasen aus.

Im Rahmen des vom BMBF geförderten Verbundprojektes "OMID" (Förderkennzeichen 16SV994) wurden am Beispiel eines CO<sub>2</sub>-Sensorsystems wesentliche Fragen zur Modellierung und Optimierung von optischen Sensorsystemen untersucht. Die Optimierung von

Analysekammern mit gefalteten Strahlengängen bildete einen Schwerpunkt der Arbeiten. Ein modulare Modellierungsansatz für das Gesamtsystem unter Einbeziehung der gasphysikalischen und optischen Abhängigkeiten sichert die Übertragbarkeit der Ergebnisse auf andere Anwendungen (Gasarten und Konzentrationen).

## 2 Meßprinzip und Konzept des lowcost NDIR-Sensorsystems

Bei Einwirkung von elektromagnetischen Strahlungen können durch Resonanz bestimmte Moleküle eines Gases oder eines Gasgemisches zu mechanischen Schwingungen angeregt werden, was als Absorption der entsprechenden Strahlungsbestandteile erkennbar ist. Diese ist bei konstanter Weglänge des Strahls im Medium direkt von der Gaskonzentration abhängig. Dieser Zusammenhang wird durch das Lambert-Beer-Gesetz beschrieben.

$$I / I_0 = e^{-kcl}$$

$I_0$  = Ausgangsintensität

$I$  = Restintensität (nach Absorption)

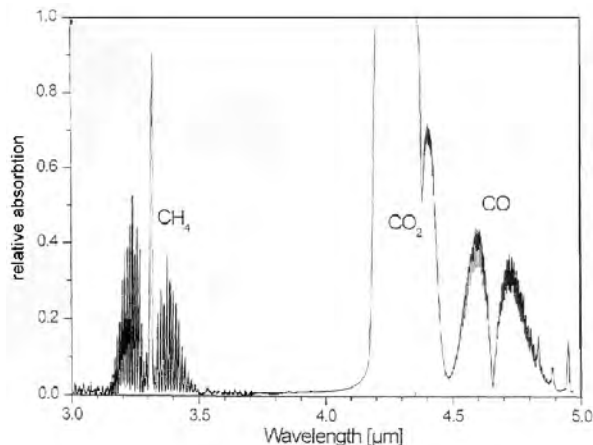
$k$  = Absorptionskoeffizient

$c$  = Konzentration

$l$  = optische Weglänge

Der darin enthaltene Absorptionskoeffizient  $k$  ist materialspezifisch und in der Praxis zudem druck- und temperaturabhängig. Das (nicht dispersive) NDIR-Meßverfahren basiert auf der Erkenntnis, daß die jeweiligen Gase im

IR-Wellenlängenbereich typische Absorptionscharakteristika aufweisen. Im Gegensatz zu den Spektrometersystemen erfolgt hier nicht die Betrachtung eines Wellenlängenbereiches sondern nur die Bewertung einer separaten spezifischen Wellenlänge.



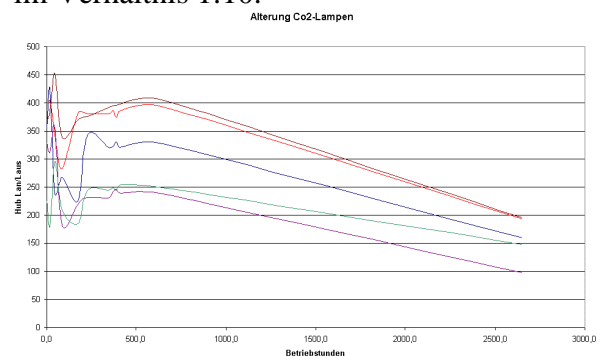
**Abb.1 Spektrale Verteilung**

Wesentliche Vereinfachungen in der Gerätetechnik werden möglich, die das NDIR-Meßverfahren auch für den Low cost - Bereich attraktiv erscheinen lassen. So werden ursprünglich breitbandige Systeme mittels Filter an das jeweils zu detektierende Meßmedium angepasst. Die Anpassung erfolgt aus Kostengründen vorwiegend durch Verwendung spezieller schmalbandiger Detektoren. Dadurch können im Wellenlängenbereich bis etwa 4...5 µm Standard-Glühlämpchen mit lediglich modifizierten Glaskolben als IR-Quelle Anwendung finden. Die Auslegung des IR-optischen Systems als Einstrahlensystem stellt einen weiteren Schritt zur Realisierung von Low cost - Systemen dar, bedingt aber Kompromisse hinsichtlich der Auflösung und Stabilität des Gesamtsystems.

Im einfachsten Fall reduziert sich ein NDIR-Meßsystem auf eine geometrische Lagefixierung einer IR-Quelle zu einem selektiven IR-Detektor (Linearkammer). Für Anwendungen auf Gase mit geringeren Absorptionskoeffizienten bzw. für Anwendungen mit geringen Aussenabmessungen ist eine Realisierung von optischen Systemen mit Strahlengangfaltung

zum Erreichen eines kompakten Gerätedesigns als sinnvoll. Dabei ist einerseits das Erreichen einer möglichst maximalen Abbildungsgüte der Quelle auf dem Detektor die Hauptforderung an die Systemkonstruktion. Zur Sicherung einer Serientauglichkeit der Konstruktion ist andererseits eine Resistenz des optischen Systems gegenüber mechanischen und Fertigungstoleranzen erforderlich. Die Ergebnisse der Arbeiten zur Optimierung des optischen Systems mit Strahlengangfaltung und zur Modellbildung des Gesamtsystems werden in einem gesonderten Beitrag des Statusseminars vorgestellt.

Das elektronische Systemkonzept der NDIR-Sensoren hat neben der Signalaufbereitung des IR-Detektors auch die Ansteuerung der IR-Quelle zu berücksichtigen. So wird durch eine Taktung der IR-Quelle einerseits eine Senkung des Strombedarfs realisiert. Andererseits können so Meßwertverfälschungen durch unbeabsichtigte Erwärmungen des Prüfgases vermieden und Alterungseffekte der IR-Quelle verzögert werden. Tastverhältnisse von 1:5 bis 1:10 bei Periodenzeiten von 1..2 sec erscheinen bei Einhaltung der Forderungen nach entsprechenden Ansprechgeschwindigkeiten als praktikabel. Die Verzögerung der Quellenalterung ist besonders wichtig in Einstrahlensystemen, wird hier doch bewußt auf eine Systemkorrektur über einen zusätzlichen Referenzkanal verzichtet. Praktische Vergleiche von Langzeitprüfungen einzelner IR-Quellen im Dauerbetrieb mit denen vom Dauerbetrieb von Komplettsensoren verdeutlichen die erreichbare Verbesserung der Quellenstabilität bei einer Systemtaktung im Verhältnis 1:10.



**Abb.2 Dauerbetrieb Lämpchen**



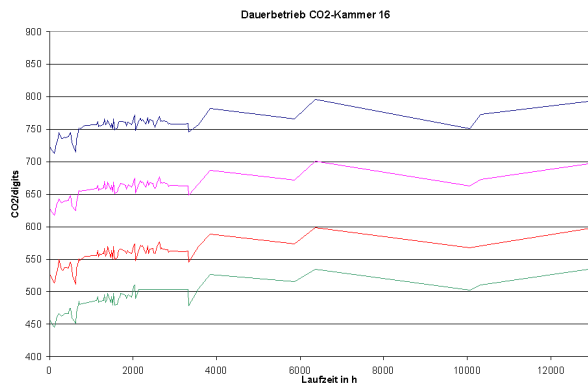


Abb.3 Dauerbetrieb Gassensor

Die Betrachtung der Signaldifferenz zwischen bestrahlten und unbestrahlten Detektor sichert zudem einen einfachen Driftausgleich der Analogstufen der Sensorelektronik. Eine zum IR-Detektorsignal synchrone Erfassung der Gastemperatur ist zum Erreichen einer maximalen Systemauflösung unumgänglich. Diese erfolgt am besten unmittelbar im Detektorbereich. Eine Berücksichtigung des geographischen Höhendrucks kann statisch erfolgen (Barometerprinzip). Der Einfluß klimatischer Luftdruckschwankungen liegt innerhalb der Systemauflösung. [1]

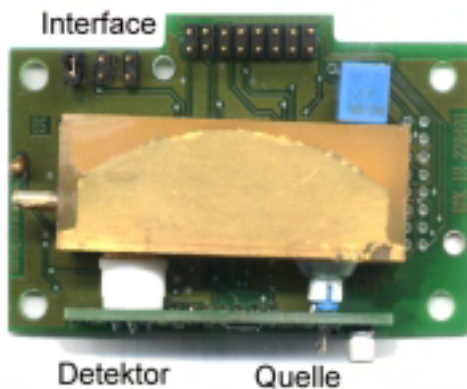


Abb.4 Bild Prototyp

Zur Systemsteuerung und zur Erfassung der physikalisch bedingten, nichtlinearen Übertragungscharakteristiken von Meßprinzip und Einzelkomponenten ist der Einsatz von Mikrocontrollern sinnvoll. Bereits 8Bit-Systeme stellen dafür ausreichend Rechenleistung zur Verfügung.

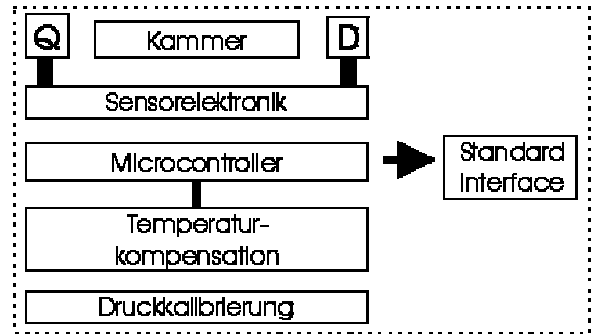


Abb.5 Systemaufbau Prototyp

### 3 Signalauswertung / Kennfeldberechnung

Anhand umfangreicher praktischer Untersuchungen erster Prototypen wurden die grundlegenden Abhängigkeiten des unkompensierten Ausgangssignals der NDIR-Gassensoren von Gaskonzentration und Temperatur ermittelt. Diese konnten unter Annahme der prinzipiellen Abhängigkeiten

$$CO_2\text{-Konzentration} = a \cdot X^3 + b \cdot X^2 + c \cdot X + d$$

mit  $X = CO_2MW$  ( $CO_2$ -Meßwert)

$$\text{Temperatur} = e \cdot Y^2 + f \cdot Y + g$$

mit  $Y = \text{TempMW}$  (Temperatur-Meßwert)

in einen Polynomansatz

$$CO_2\text{-Konz} (CO_2MW, \text{TempMW}) = a_1 X^3 Y^2 + a_2 X^3 Y + a_3 X^3 + a_4 X^2 Y^2 + a_5 X^2 Y + a_6 X^2 + a_7 X Y^2 + a_8 X Y + a_9 X + a_{10} Y^2 + a_{11} Y + a_{12}$$

zusammenfassen werden.

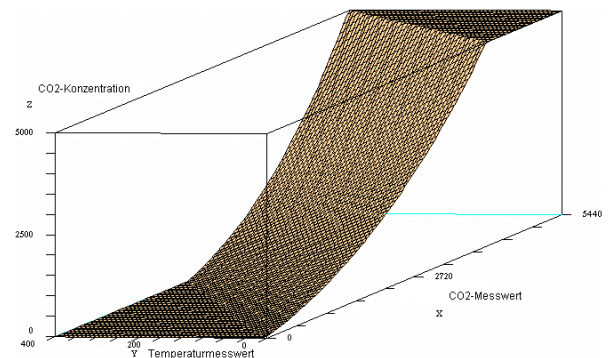


Abb.6 Kennlinienfeld

Zur Bestimmung der Koeffizienten der beschriebenen Polynomgleichung sind individuelle Kalibriermessungen der einzelnen Sensoren bei jeweils vier verschiedenen Gaskonzentrationen und bei je drei verschiedenen Temperaturen erforderlich. Die Berechnung der Koeffizienten  $a_1 \dots a_{12}$  aus den Meßwerten erfordert die Anwendung einer Hochsprache (SCILAB-Script oder Tabellenkalkulation) auf einem externen PC-System. Auf deren Basis kann anschließend die eigentliche Kennfeldinterpolation mit Hilfe des integrierten Mikrocontrollers unmittelbar im Sensor erfolgen.

## Kennfeldinterpolation

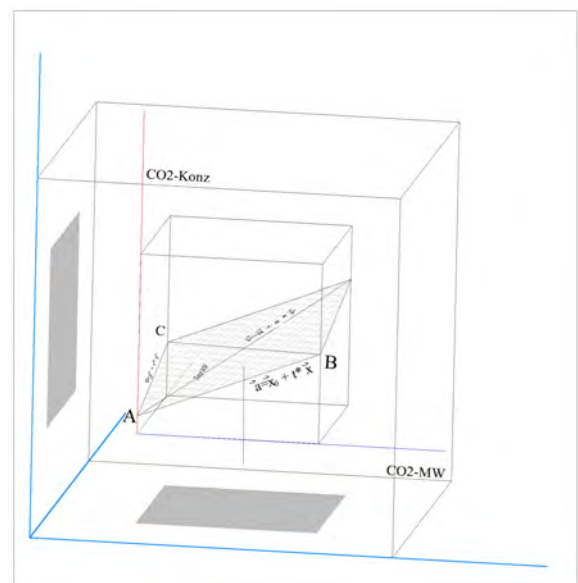
Zur Realisierung der Kennfeldinterpolation im Gassensor wurden zwei unterschiedliche Lösungsstrategien realisiert und auf ihre praktische Anwendbarkeit untersucht:

- die Berechnung der Gaskonzentration anhand der Approximationsgleichung
- die Berechnung der Gaskonzentration durch Polygonzerlegung des Kennlinienfeldes.

Die unmittelbare Berechnung der Gaskonzentration anhand der Approximationsgleichung erfordert zur Sicherung einer entsprechenden Systemauflösung die Realisierung einer 16Bit – Arithmetik für Meßwerte und Koeffizienten (Zahldarstellungen jeweils als 16Bit-Mantisse und 16Bit-Exponent) im Mikrocontroller des Systems. Daraus resultiert einerseits ein erheblicher Programmier- und Rechenaufwand (besonders bei Anwendung von kostengünstigen 8Bit-Mikrocontrollern). Zum anderen sind damit Mehraufwendungen hinsichtlich ggf. erforderlicher Erhöhung der Taktfrequenz zur Sicherung der Rechenzeiten bzw. ggf. Vergrößerung des RAM-Speichers verbunden, die sich unmittelbar auf die Hardwarekosten des Sensorsystems auswirken können. Andererseits überzeugen die Vorteile des Verfahrens wie die gleichmäßige Genauigkeit und Stetigkeit des Rechenergebnisses im Kennfeld. Ein weiterer

Vorteil in der praktischen Anwendung des Verfahrens liegt darin, daß bei stets gleichen Berechnungsformeln nur die Koeffizienten sensorindividuell sind.

Weiterhin wurden zur Reduzierung des Rechenaufwandes zur Kennfeldinterpolation alternative Verfahren auf Basis der Polygonzerlegung in Verbindung mit linearen Interpolationen auf ihre Anwendbarkeit geprüft. Dabei erfolgt nach der Koeffizientenberechnung eine Zerlegung des Kennlinienfeldes in Teilpolygone über eine Berechnung einer entsprechenden Zahl von Stützstellen. Das geschieht unmittelbar nach der Koeffizientenberechnung in der Hochsprachenprogrammierung auf einem externen PC-System. Die berechneten, sensorindividuellen Stützstellen werden im Sensor gespeichert und bilden die Basis für die Kennwertberechnung durch lineare Interpolationen innerhalb des entsprechenden Teilpolygons.



**Abb.7 Lineare Interpolation innerhalb des Kennfeldpolygons**

Das Verfahren besteht vom Gedankenansatz durch die Verteilung des Rechenaufwandes zwischen einem externen System und der Sensorelektronik. Im Vergleich zum direkten Berechnungsverfahren wird der Rechenaufwand für die lineare Interpolation zwar wesentlich reduziert, aber die Größe der Stützstellenmatrix wächst mit der geforderten

Systemauflösung. So kann bereits zum sicheren Erreichen einer Systemauflösung von 1% eine Matrix von 128 x 128 16Bit-Werten erforderlich sein. Schwierig gestalten sich in der praktischen Anwendung ggf. erforderliche Kennwertkorrekturen, die stets einen Austausch der kompletten Stützstellenmatrix erfordern.

## 4 Erprobungsergebnisse

Für einen praxisrelevanten Vergleich der unterschiedlichen Verfahren zur Kennfeldinterpolation wurden Prototypen jeweils auf Basis von 8051-Mikrocontrollern realisiert und befinden sich im Probelauf. Mit beiden Interpolationsverfahren sind Systemauflösungen im Bereich 1% erreichbar.



**Abb.8 Prototyp**

Für die Systeme mit realisierter Kennfeldberechnung erscheint nach der praktischen Inbetriebnahme eine Aufteilung des Programmspeichers in den klassischen EPROM-Bereich und einen Flash-ROM-Bereich für die Koeffizienten als äußerst praktisch. So wird eine einfache

Programmierung/Korrektur der individuellen Sensorkoeffizienten möglich. In der praktischen Erprobung mit einer Controllertaktfrequenz von 8MHz und einem Tastverhältnis für die IR-Quelle von 1:10 bei einer Periodenzeit von 2 sec steht ausreichend Rechenleistung für die 16Bit-Arithmetik zur Verfügung. Dabei genügt das zeitliche Ansprechverhalten des Komplettsensors den Systemforderungen.

Bei Anwendung der Polygonzerlegung wurden im praktischen Betrieb bei Grenzlagen im Kennlinienfeld teilweise Kennwertsprünge festgestellt. Durch Optimierungen hinsichtlich Dimensionierung der Stützstellenmatrix erscheinen auch hierbei die Einhaltung der geforderten Systemauflösung von 1% als realisierbar.

Der bisherige Umfang der Prototypenfertigung läßt derzeit noch keine Aussagen zur Übertragbarkeit der Ergebnisse bzw. zur Toleranzverteilung innerhalb eines Musterloses zu.

## 5 Ausblick

Die vorgestellten zwei Verfahren zur Kennwertberechnung des NDIR-Gassensors können nachfolgenden Produktentwicklungen als Basis dienen. Dabei kann die vergleichende Bewertung bzw. der Nachweis der Praxisrelevanz der beiden Verfahren anhand entsprechende Mustererien und Erprobungen erfolgen.

## Literatur

[1] VDI/VDE-IT (Herausgeber): Kombiniertes Einsatz von Gassensoren verschiedener Technologien (Verbundprojekt 1993-1996 Abschlußbericht) Teltow:1998

[2] Junger, S. ua. : High Path Length-to-Volume Ratio Optical Cell for Biomedical In-Line Gas Monitoring in Real Time Using Multiple Wavelength Infrared Spectroscopy. IRS2 2000 Proceedings (2000), S. 191-194

[3] Staab, J.: Industrielle Gasanalyse  
Oldenburg,

[4] N.N.: Applications of thermoelectric  
infrared Sensors (Thermopiles): Gas detection  
by infrared absorption; NDIR, Firmenschrift  
PerkinElmer, 2000

[5] Firmenschriften Madur electronics  
GmbH, A-Wien

[6] Günzler, H. IR-Spektroskopie VCH  
Verlagsgesellschaft mbH Weinheim, 1995

# MODELLBASIERTE OPTIMIERUNG EINES IR-GASSENSORS

Ingo Sieber, Institut für Angewandte Informatik, Forschungszentrum Karlsruhe  
 Oliver Nüssen, ITEM, Universität Bremen  
 sieber@iai.fzk.de;

## Kurzfassung

Die Herstellung von Teststrukturen von Mikrosensoren und Mikroaktuatoren ist sehr zeit- und kostenintensiv. Dies begrenzt im konventionellen Entwurfsprozess die Anzahl der Designvarianten. Aus diesen Gründen sind rechnergestützte Entwurfstechniken aus der Mikrosystemtechnik nicht mehr wegzudenken. Diese Entwurfstechniken können dazu eingesetzt werden, die Systemperformance in Abhängigkeit vielfältiger Einflussfaktoren zu optimieren.

## 1 Einleitung

In diesem Artikel wird der grundlegende Ansatz eines Systemmodells eines IR-Gassensorsystems vorgestellt. Das Ziel ist, den CO<sub>2</sub>-Gassensor der Firma Micro-Hybrid Electronic GmbH auf Verhaltensbasis zu beschreiben, um eine Gesamtsystemsimulation durchführen zu können. Die wesentliche Eingangsgröße für das Systemmodell ist die Verteilung der Laufwege innerhalb der Analyse- und Absorptionskammer. Zur Gewinnung dieser Information wurde ein optisches Modell der Gasanalyse- und Absorptionskammer erstellt. Dieses optische Modell dient als Grundlage für eine Optimierung der Analyse- und Absorptionskammer. Die Optimierung fand dabei hinsichtlich eines robusten Systemdesigns statt. Robustheit bedeutet in diesem Zusammenhang, dass die Funktionalität des Systems auch in Abhängigkeit der Toleranzen der einzelnen Funktionsmodule erhalten bleibt.

Die Funktionsweise von IR-Analysatoren beruht auf der Absorption bestimmter Wellenlängen von Infrarotstrahlung für spezifische Gase. Im Wellenlängenbereich von 3,0-5,2µm bieten die Absorptionsbanden vieler Gase und Dämpfe die Möglichkeit Messsysteme aufzubauen, die aus der Kombination von IR-Quelle, Absorptions-

zelle, Wellenlängenfilter (bzw. Spektrometer) und Detektor bestehen (siehe Abb.1). Zur Quantifizierung der IR-Gasanalyse findet das Lambert-Beer Gesetz Anwendung, in welchem die Absorption beschrieben wird als Funktion der Schichtdicke (bzw. des Laufweges des Strahls im zu messenden Medium), dem Molekültyp und der Konzentration einer gelösten Substanz, die bei konstanter Temperatur proportional zum Partialdruck ist. Die IR-Strahlung durchläuft die Absorptionszelle auf einem Weg der Länge  $d$ . Dabei wird die Eingangsstrahlungsintensität  $I_{0\lambda}$  auf die Strahlungsintensität  $I_{\lambda}$  nach Durchlaufen der Probe reduziert. Mit diesen Größen kann das Lambert-Beer Gesetz folgendermaßen formuliert werden [1]:

$$\log \frac{I_{\lambda}}{I_{0\lambda}} = -\varepsilon \cdot c \cdot d = -E \text{ mit:}$$

$E$  = Extinktion

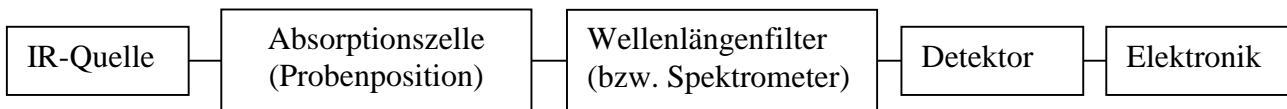
$\varepsilon$  = Extinktionskoeffizient

$c$  = Konzentration der Substanz

$d$  = Strahlweg in der Substanz

Auf Grundlage dieser Gleichung kann jede Komponente eines Gemischs aus dem IR-Spektrum quantitativ bestimmt werden, unter der Voraussetzung, dass ein genügend intensives Absorptionsband gefunden werden kann, das durch die anderen Komponenten nicht, oder nur zu einem bekannten Ausmaß gestört wird.

Der Artikel gliedert sich folgendermaßen: zuerst wird eine allgemeine Beschreibung des IR-Gassensors gegeben. Die darauf folgenden Abschnitte beschäftigen sich mit der Optimierung des optischen Aufbaus des Gassensors. Auf der Grundlage von Einzelmodellen wird das Gesamtmodell des IR-Gassensors vorgestellt.

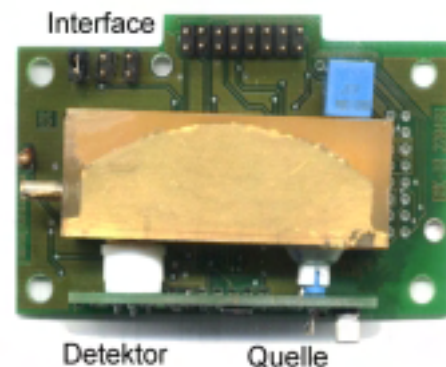


**Abb. 1: Schematische Darstellung eines IR-Gassensors**

## 2 Der IR-Gassensor

Die im folgenden betrachteten nicht dispersiven IR-Gassensoren (NDIR) erlauben die Konzentrationsmessung von Gasen unter Nutzung der IR-Absorption als eigentliche Meßgröße. In der heute am meisten genutzten Form sind die Absorptionskammern als lineare Anordnung mit gegenüberliegenden IR-Strahler und IR-Detektoren ausgeführt. Der lineare, meist relativ kurze Strahlungsverlauf stellt jedoch eine wesentliche Begrenzung für die praktische Anwendung der Sensoren dar, da die nachzuweisenden Gase eine gewisse Mindestabsorption besitzen müssen. Für die Detektion schwach absorbierender Gase muß deshalb der optische Strahlweg in der Analyseammer vergrößert werden. Dieses läßt sich bei Beibehaltung der oft begrenzten geometrischen Abmessungen nur durch eine Faltung des Strahlenganges unter Einfügung zusätzlicher reflektierender Elemente erreichen. Dabei gewinnt die Frage der Integration von strahlfokussierenden Elementen zunehmend an Bedeutung, um einerseits eine ausreichende Aussteuerung des Detektors zu erreichen und andererseits den Anteil von Strahlen, die aufgrund unbeabsichtigter Reflexionen eine undefinierte Weglänge beim Auftreffen auf den Detektor aufweisen, zu reduzieren. Die Störstrahlen mit den kürzeren Weglängen wirken bezogen auf die Gasdetektivität des Sensors als Rauschsignal. Zur Sicherung der Systemstabilität im Zeitbereich und bei sich ändernden Umgebungsbedingungen verfügen die klassischen Gassensoren über einen zusätzlichen Referenzstrahlweg. Da in den zunächst betrachteten Applikationen eine regelmäßige Absolutkalibrierung der Gassensoren vorgeschrieben ist, werden die hierfür vorgesehenen Gassensoren zur Vermeidung der mit dem Referenzkanal verbundenen Kosten als Einstrahlssystem

ausgeführt. Eine spezielle zeitliche Taktung des Sensors sichert dennoch eine ausreichende Systemstabilität, indem die ermittelten "Dunkelgrößen" als Korrekturwerte für die nachfolgenden Meßwerte herangezogen werden. Somit wird die elektronische Signalaufbereitung bis einschließlich Controller-Interface zum unmittelbaren Bestandteil der Sensorsysteme (siehe Abb. 2). Die Gebäudeüberwachung, speziell die Raumluftüberwachung in zentral klimatisierten Einrichtungen des öffentlichen Lebens und die Umweltmeßtechnik werden vorerst als bevorzugte Einsatzgebiete angesehen. Daher ist die Applikation als CO<sub>2</sub>-Sensor erster Untersuchungsgegenstand, wobei eine maximale Übertragbarkeit der Ergebnisse auf andere Gase (Kohlenwasserstoffe, CO ...) angestrebt wird.



**Abb. 2: Realisierter Musteraufbau des CO<sub>2</sub>-Sensors als kompakte Einheit von IR-optischen Komponenten (Kammer, Quelle, Detektor) und Sensorelektronik.**

## 3 Das optische Modell

Der Gegenstand der optischen Modellbildung ist die Beschreibung der IR-Quelle und der Absorptionskammer mittels eines parametrisierbaren optischen Modells. Diese Arbeiten wurden im Forschungszentrum Karlsruhe durchgeführt.

### 3.1 Modellbildung der IR Quelle

Die IR-Quelle besteht aus einer Breitband IR-Lampe und einem Lampenreflektor. Der Reflektor dient dazu, die Strahldivergenz zu begrenzen um einen definierten Laufwegbereich der Strahlen um eine mittlere Entfernung zu gewährleisten. Die Abstrahlcharakteristik des Quellmodells wurde anhand von Messungen mit der realen IR-Quelle abgeglichen [2].

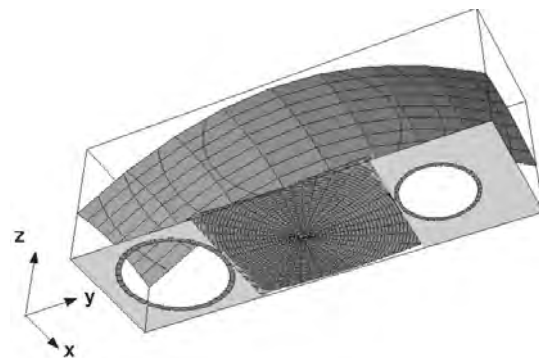
### 3.2 Modellbildung der Absorptionskammer

Die charakteristische Absorption durch das Gas findet in der Absorptionskammer statt. Die Kammer ist auf die spezifischen Absorptionseigenschaften eines Gases ausgelegt, d.h. sie ist auf die Erfassung eines einzelnen Gases abgestimmt. Um nun auch schwach absorbierende Gase zu detektieren, muss der Strahlweg durch das Gas verlängert werden. Da dies ohne eine erhebliche Vergrößerung der Dimensionen der Absorptionskammer vorzustattengehen soll, müssen die Begrenzungsflächen der Kammer reflektiv beschichtet sein und fokussierende Aufgaben übernehmen. In Abbildung 3 ist auf der linken Seite eine Fotografie der Absorptionszelle dargestellt. Der planare



In der Darstellung rechts ist das Modell der Absorptionskammer im Vergleich zu sehen. Auch hier sind die reflektierenden und fokussierenden Flächen gut zu erkennen. Die seitlichen Begrenzungsflächen der Kammer sind reflektiv beschichtet und dienen zur Strahlfaltung. Diese Flächen sind in Abbildung 3 nicht dargestellt, um einen Einblick in die Kammer zu ermöglichen.

Das optische Modell von IR-Quelle und Absorptionszelle erlaubt die Berechnung der Strahlausbreitung von der Quelle zum Detektor in einer nicht-sequentiellen Beschreibung [3]. Auf Basis dieses Modells wurden Untersuchungen der Strahlausbreitung und der Strahllaufwege innerhalb der Analyse- und Absorptionskammer durchgeführt sowie anhand von Parametersimulationen gezeigt, wie empfindlich der aktuelle Kammerentwurf hinsichtlich der Quellposition ist [4]. Die im Gassensor vorgesehenen Lichtquellen weisen eine Positionsschwankung der Glühwendel von 1mm auf. Damit der IR-Gassensor trotz dieser großen Positionsschwankungen funktionstüchtig bleibt, muss die Lichtquelle aktiv unter Zuhilfenahme einer Leistungsmessung montiert werden [2]. Dies ist zeit- und personalaufwändig und damit kostenintensiv. Für eine kostengünstige



**Abb. 3: Fotografie der Absorptionszelle (links) und parametrisierbares Modell (rechts).**

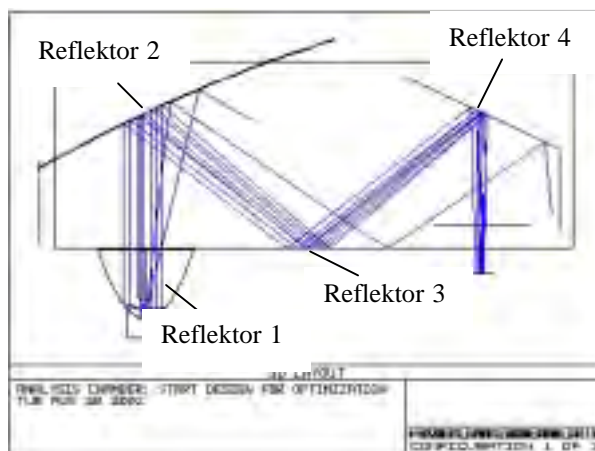
Deckelspiegel wurde für die Abbildung abgenommen, um Sicht auf die fokussierenden und reflektierenden Strukturen zu erhalten. Die Kammer ist mit einem Sensorchip (einem Thermopile-Element) bestückt, das rechts unten auf dem Foto zu sehen ist. Außerdem ist auch die kreisrunde Einlassöffnung für die Quelle zu sehen (links unten an der Kammer).

Herstellung ist eine automatisierbare Bestückung der Absorptionskammer mit Quell- und Detektormodul erstrebenswert. Ein Ansatz für die Systemoptimierung ist demgemäß einen Entwurf der Analyse- und Absorptionskammer zu finden, der hinsichtlich der Systemtoleranzen optimal robust ist.

## 4 Optimierung der Geometrie der Absorptionszelle

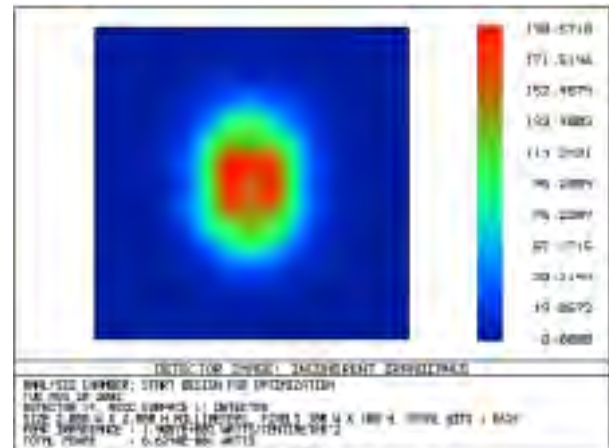
### 4.1 Startmodell für die Optimierung

Als Startmodell für die Optimierung wurde das in Abbildung 4 dargestellte Design gewählt. Es besteht aus vier Reflektoren, von denen der erste die Aufgabe hat, die Quellabstrahlung zu kollimieren. Der zweite Reflektor fokussiert die kollimierte Strahlung auf das Detektormodul. Die Spiegel 3 und 4 sind eben und dienen in diesem Startmodell nur der Strahlfaltung.

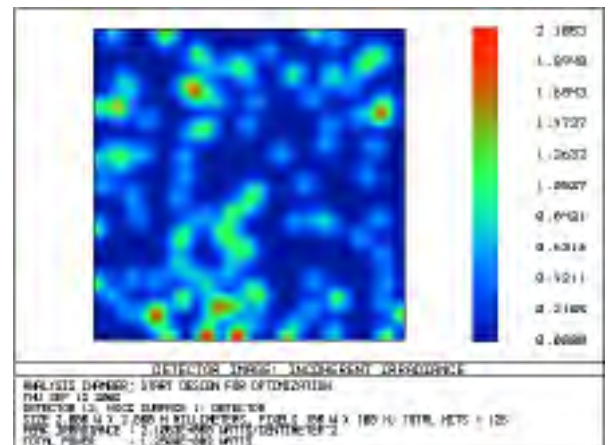


**Abb. 4: Startmodell der Absorptionskammer.**

Für dieses Modell der Absorptionskammer treffen unter Verwendung einer ideal positionierten Quelle ca. 86% der berechneten Strahlen auf den Detektor. Die Bestrahlungsstärke-Verteilung für diese Anordnung ist in Abb. 5 dargestellt. Bei einer Fehlpositionierung der Quelle um den halben Toleranzbereich treffen nur noch 1% der Strahlen auf den Detektor. Die Bestrahlungsstärke-Verteilung für diese Konfiguration ist in Abb. 6 dargestellt. Auf Basis dieses Startmodells wurde eine Optimierung der Geometrie der Absorptionskammer durchgeführt. Ziel dieser Optimierung war, ein Design zu finden, das im gesamten Toleranzbereich der Quellposition ein zufriedenstellendes Ergebnis und eine gleichmäßige Performance liefert.



**Abb. 5: Bestrahlungsstärke-Verteilung für das Startmodell bei idealer Quellposition.**



**Abb. 6: Bestrahlungsstärke-Verteilung für das Startmodell bei einer Fehlpositionierung der Quelle um den halben Toleranzbereich.**

Für die Optimierung wurden die folgenden Parameter definiert:

- die Lage der Reflektoren zueinander
- die Krümmung der Reflektoren
- die Position der Punktquelle innerhalb des Reflektor 1

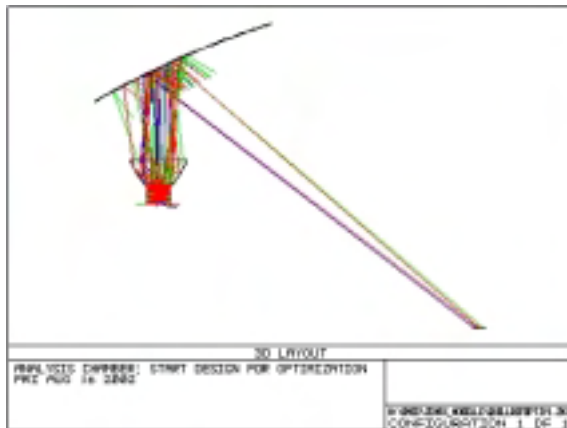
Randbedingungen der Optimierung waren die äußeren Abmessungen der Kammer, der Abstand zwischen Quelle und Detektor und der Strahlaufweg in der Kammer.

### 4.2 Optimierung

In einem ersten Schritt wurde die Form des Quellreflektors optimiert, sodass er über den gesamten Toleranzbereich ungefähr die



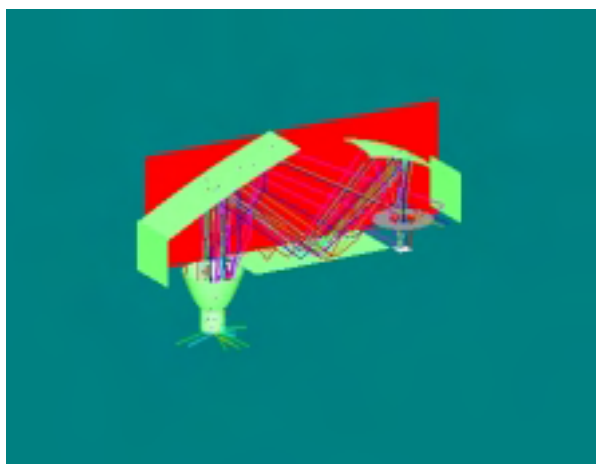
gleiche Performance besitzt (siehe Abb.7).



**Abb. 7: Optimierung des Quellreflektors hinsichtlich der Positionstoleranzen der Quelle.**

Dieser optimierte Quellreflektor wurde für die Optimierung der Kammergeometrie in das Startmodell eingefügt.

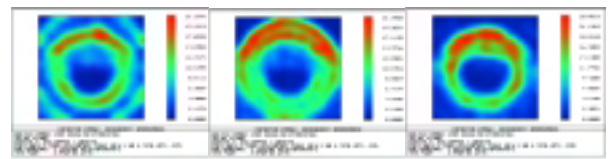
Im weiteren Verlauf der Optimierung wurden zunächst die beiden ebenen Reflektoren 3 und 4 in ihrer Form freigegeben und anschließend alle vier Reflektoren gleichzeitig. Das optimierte Modell ist in Abb. 8 dargestellt. Auffallend dabei ist, dass der Reflektor 3 seine ebene Form behält und nach wie vor nur der Strahlfaltung dient. Die stärkste Veränderung ist am Reflektor 4 auszumachen, der sowohl in x- als auch in y-Richtung stark gekrümmt ist.



**Abb. 8: Optimiertes Design der Absorptionskammer.**

In der Abb. 9 sind die Bestrahlungsstärke-Verteilungen für drei Positionen innerhalb des Toleranzbereichs dargestellt. Dabei wurde

jeweils eine Verteilung für die Quellpositionen an den Grenzen des Toleranzbereiches berechnet und eine, die mittig im Toleranzbereich lag.



**Abb. 9: Bestrahlungsstärke-Verteilungen des optimierten Designs für Quellpositionen im Toleranzbereich.**

Das optimierte Design der Absorptionskammer erreicht eine Strahlausbeute von 30% auf dem Detektor über den gesamten Toleranzbereich der Quellposition. Damit garantiert das optimierte Design über den gesamten Toleranzbereich eine definierte Systemperformance und ermöglicht eine automatisierte Bestückung der Analyse kammer mit der IR-Quelle.

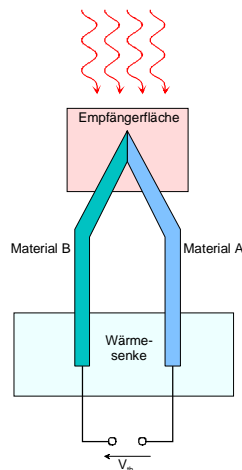
## 5 Der Thermopile Sensor

Grundlage des Modells ist der von der Firma Micro-Hybrid entwickelte Thermopile-Sensor. Ergebnisse zur Beschreibung dieser Systemkomponente wurden bereits im Verbundprojekt MIMOSYS [5] erbracht.

Wie bei vielen Mikrosystemen existieren auch bei diesem Sensor Wechselwirkungen zwischen Größen aus unterschiedlichen Domänen, in diesem Fall aus den Gebieten der Elektrotechnik und der Thermodynamik. Das thermoelektrische Prinzip des Sensors ist für eine Systemintegration zusammen mit der Auswerteelektronik besonders geeignet, da die auftretenden thermischen Spannungen nicht von den geometrischen Abmessungen des Bauteils abhängig sind [6].

### 5.1 Funktionsprinzip

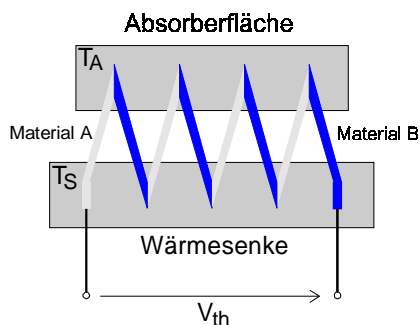
Als thermoelektrisch sensitives Bauelement fungiert ein Thermopilesensor. Die Funktion der Thermoelemente beruht auf dem Seebeck-Effekt. Bei einem Thermoelement sind je zwei Kontaktstellen vorhanden. Besteht zwischen diesen nun eine Temperaturdifferenz, so fließt aufgrund der Thermospannung ein messbarer Thermostrom (siehe Abb.10).



**Abb.10: Aufbau eines Thermoelements**

Für eine messtechnische Auswertung muss die sehr niedrige Thermospannung erhöht werden. Zu diesem Zweck werden mehrere Thermo-elemente zu einer Thermopile zusammengesaltet (siehe Abb.11), was zu einer Addition der Einzelspannungen führt.

Die Referenz-Kontaktstellen sind mit der Wärmesenke und die Meß-Kontaktstellen mit der Absorberfläche thermisch gekoppelt.

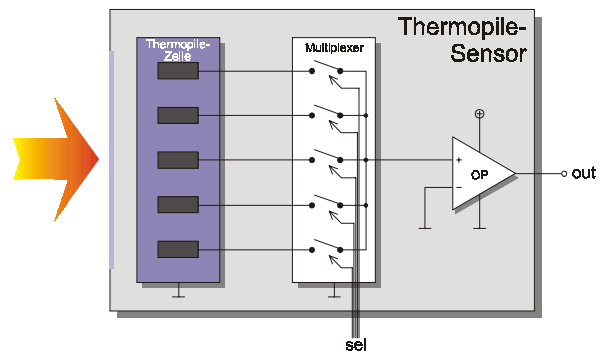


**Abb.11: Prinzipdarstellung eines Thermopiles aus vier Elementen**

Die Thermospannung ist nicht von den Abmessungen der Thermo-elemente abhängig. So ist es möglich, eine Miniaturisierung und somit eine Integration mehrerer Thermopiles zu einer Thermopile-Zeile vorzunehmen [7].

## 5.2 Aufbau des Sensors

Die in den einzelnen Thermo-elementen erzeugte Pixelspannung ist auch nach der Reihenschaltung zu Thermopiles so gering, daß sie zur weiteren Verwendung verstärkt werden muß. Dazu wird die Thermopile-Zeile mit einem Multiplexer und einem Operationsverstärker in einem Gehäuse zu einem Thermopile-Sensor integriert.



**Abb.12: Blockschaltbild des Sensors**

Der Multiplexer dient zur Adressierung des auszulesenden Thermopiles und schaltet das jeweilige Signal auf den Verstärker. In Abb.12 ist das Blockschaltbild des Thermopile-Sensors dargestellt.

## 6 Physikalisches Systemmodell

Am Institut für Theoretische Elektrotechnik und Mikroelektronik (ITEM) der Universität Bremen erfolgt die Modellierung des Sensors auf höherer Abstraktionsebene. Ziel ist es, den CO<sub>2</sub>-Gassensor der Firma MicroHybrid auf Verhaltensbasis zu beschreiben, um eine Gesamtsystemsimulation durchführen zu können.

Die bisherigen Arbeiten resultieren in der Entwicklung eines Java-Programmes, welches die physikalischen Eigenschaften des Sensors beschreibt. Das im Programm nachgebildete Verhalten wird später in eine VHDL-AMS-Beschreibung umgesetzt und anschließend einer Optimierung hinsichtlich der Systemparameter zugeführt.

Für die Charakterisierung des eingesetzten IR-Sensors wurden bereits im Verbundprojekt MIMOSYS [5] Arbeiten zur Untersuchung des thermischen Verhaltens durchgeführt. Diese beziehen sich auf die Reaktion des IR-Sensors bei Veränderungen der Umgebungstemperatur.

Ein wesentliches Argument für die Notwendigkeit einer Betrachtung des Verhaltens auf Systemebene liegt in dem mit der Simulation einhergehenden Rechenaufwand begründet. In den wenigsten Fällen ist es möglich, ein gesamtes Mikrosystem in einem numerischen Simulator auf der Basis der im System geltenden physikalischen Grundgleichungen zu betrachten. Die sich darüber hinaus bei der Einbeziehung spezieller

Simulatoren ergebende Simulatorkopplung kann zu weiteren Problemen führen.

Bei der Verhaltensmodellierung erfolgt eine Abstraktion der physikalischen Zusammenhänge. Die Vorgänge innerhalb des zu beschreibenden Elementes werden auf die Abhängigkeiten zwischen definierten Größen reduziert, wobei theoretisch bekannte, aber sehr komplexe Beziehungen auch durch heuristische Ansätze mit hoher Genauigkeit nachempfunden werden können. Die Portierung in die Verhaltensbeschreibungssprache VHDL-AMS dient zusätzlich der Unabhängigkeit von der verwendeten Simulationsplattform.

Die sich durch den Einsatz eines geeigneten Verhaltensmodells ergebenden Vorteile bezüglich des Simulationsaufwandes an Rechenzeit und Rechenleistung sind oft direkt wirtschaftlich messbar. Entwurfsbegleitende Modellierung, Simulation und Optimierung führen zu kürzeren Innovationszyklen und geringerem Ressourceneinsatz.

Aus den Berechnungen einzelner Strahlenverläufe ergeben sich deren jeweilige Länge sowie die Anzahl der Reflexionen an den Wänden der Analysekammer. Auf der Grundlage dieser Daten kann im weiteren ein physikalisches Modell des Gassensors aufgebaut werden. Gemäß dem Gesetz von Lambert-Beer wird ein Teil der Energie jedes einzelnen Strahls in einem gasförmigen Medium absorbiert. Diese Dämpfung ist von der Wellenlänge, der Länge des Weges und der Art des Gases bzw. der Zusammensetzung der beteiligten Gase abhängig. Sind die Wellenlänge oder der Wellenlängenbereich sowie die Länge des Weges bekannt, kann von einer gemessenen Intensität der Strahlung am IR-Sensor auf das Gas in der Analysekammer geschlossen werden.

## 6.1 Modell-Schnittstelle

Die wesentliche Eingangsgröße für das Systemmodell ist die Verteilung der Laufwege innerhalb der Analysekammer.

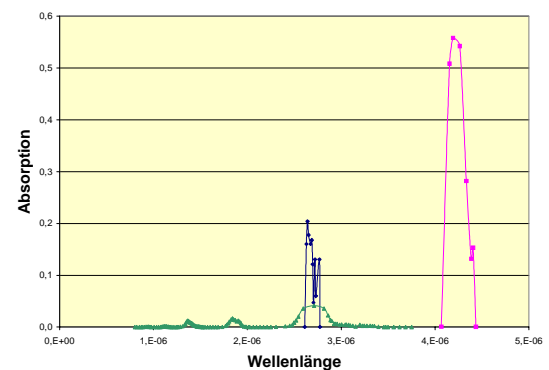
Zur Gewinnung dieser Information werden am Forschungszentrum Karlsruhe die bereits beschriebenen optischen Simulationen unter Berücksichtigung der Kammergeometrie durchgeführt.

Die Intensität der am Sensor auftreffenden Strahlung ist abhängig von der Länge des im absorbierenden Mediums zurückgelegten Weges. Stark unterschiedliche Laufwege der Strahlen durch das absorbierende Medium führen zu einer verfälschte Aussage über die Absorptionseigenschaften des involvierten Füllgases.

## 6.2 Systemmodell

Aus den physikalischen Gesetzen läßt sich ein Systemmodell formulieren. Es berechnet die Dämpfung der Intensität durch das Probegas bei einem vorgegebenen Partialdruck für jede ermittelte Strahlänge in dem gesamten Wellenlängenbereich der Lampe. In die Berechnungen wird auch das Filter mit dem entsprechenden charakteristischen Spektrum einbezogen. Da nur Strahlen berücksichtigt werden, die auf der aktiven Fläche des IR-Sensors auftreffen, kann durch Summation der Einzelintensitäten die auf dem Sensor eintreffende Leistung ermittelt werden. Daraus läßt sich die Ausgangsspannung des IR-Sensors berechnen.

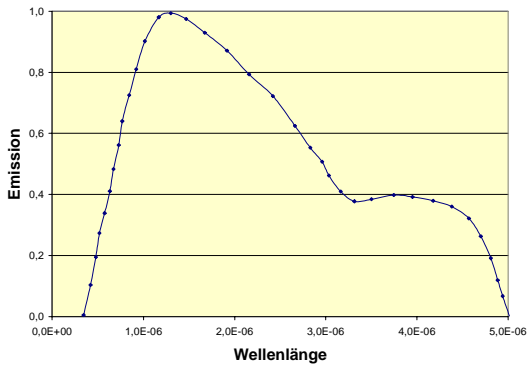
Das Systemmodell wurde zunächst in Java realisiert. Es ist mit diesem Programm sehr leicht möglich, die interessierenden Kennlinien auszutauschen oder den Partialdruck des Gases zu verändern, um erste Abschätzungen des Systems im Hinblick auf Modifikationen der beteiligten Komponenten vorzunehmen (siehe Abb.18). Dieses modulare Konzept des Einladens von Kennlinienspezifikationen erlaubt eine hohe Flexibilität bei der Anwendung des Modells.



**Abb.13: Absorptionsspektrum von CO<sub>2</sub>**

In Abb.13 ist das Spektrum von CO<sub>2</sub> dargestellt. Es ist zu erkennen, daß CO<sub>2</sub> wie

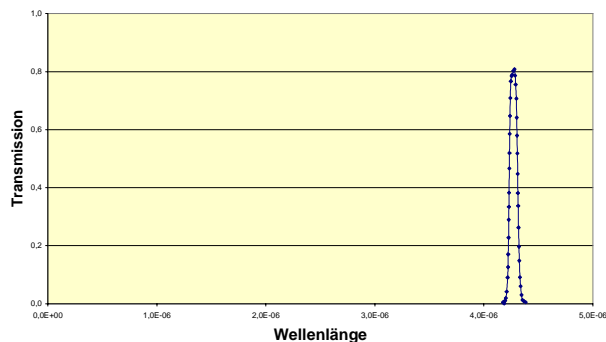
die meisten Gase größtenteils diatherm, d.h. wärmedurchlässig ist. Die Wärmestrahlung wird nur in bestimmten Bereichen absorbiert, den sogenannten Banden. Die betreffenden Wellenlängen sind für jedes Gas charakteristisch und erlauben Rückschlüsse auf die Identifikation.



**Abb.14: Spektrum der Lampe**

Die im System verwendete Lampe erzeugt Infrarotstrahlung im interessierenden Spektrum von 0.5 µm bis 5 µm. Der wichtigste Anteil der Strahlung liegt im Absorptionsband von CO<sub>2</sub> (4,0 µm bis 4,5 µm), da die Bedämpfung durch das Gas hier am stärksten ist.

Das Filter im Strahlungssensor (Thermopile-Sensor) besitzt das in Abb.16 dargestellte Spektrum. Es filtert einen Teil des Absorptionsbandes von CO<sub>2</sub> (4,0 µm bis 4,5 µm) heraus, so daß sich nur in diesem Spektralbereich die Bedämpfung der Strahlung auf das Sensorsignal auswirkt.

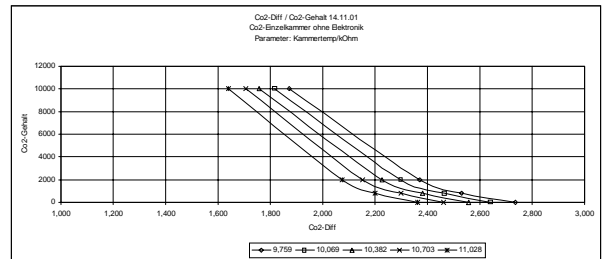


**Abb.15: Spektrum des Filters**

Durch das Filter sind im Normalfall, d.h. ohne dass ein Prüfmedium in der Analyse kammer vorhanden ist, nur Strahlen aus dem Wellenlängenbereich der charakteristischen Banden für das Spannungssignal am Sensorausgang verantwortlich. Eine Befüllung der

Anordnung durch ein Gas, auf welches das Filter ausgerichtet ist, ermöglicht die Detektion einer Auslöschung des Sensorsignals.

### 6.3 Temperaturkompensation



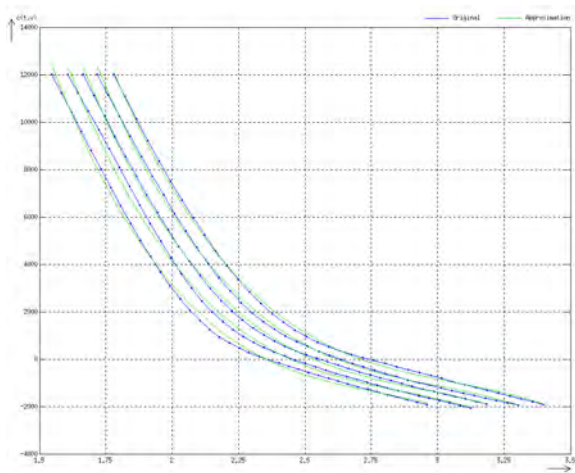
**Abb.16: Temperaturverhalten des Sensors**

Der Sensor weist eine nichtlineare Temperaturcharakteristik auf, gemäß Abb.16. Aufgetragen ist die gemessene Variation der Ausgangsspannungsdifferenz zwischen ein- und ausgeschalteter Lampe in Abhängigkeit der Umgebungstemperatur. Diese liegt in einer Drift der Sensorausgangsspannung begründet. Um in jeder Temperaturumgebung verlässliche Ausgangsdaten zu erhalten, musste daher eine Kompensation der Temperaturabhängigkeit erreicht werden. Es wurde eine einfache analytische Beschreibung für die Reproduktion durch einen Mikro-Controller in der Auswertelektronik gesucht. Ein Problem hierbei ist die wegen der sehr aufwendigen Datenerhebung geringe Anzahl von Messwerten. Die Ausgangsspannung liegt als Funktion der Gaskonzentration und der Umgebungstemperatur vor, bei insgesamt 5 Temperatur- und 4 Gaskichtewerten. Es wurde ein Polynom-Ansatz gewählt, der zwar viele, nicht physikalisch zu interpretierende Parameter enthält, dafür aber in Grundrechenarten darstellbar ist, einfach strukturiert, linear bezüglich der Fitparameter und daher in einem Schritt nicht-iterativ von Quasi-Newton-Verfahren zu lösen. Die Fitfunktion lautet:

$$c(v,t)=x_0+x_1 \cdot v+x_2 \cdot v^2+x_3 \cdot v^3+x_4 \cdot t+x_5 \cdot t \cdot v+x_6 \cdot t \cdot v^2+x_7 \cdot t \cdot v^3+x_8 \cdot t^2+x_9 \cdot t^2 \cdot v+x_{10} \cdot t^2 \cdot v^2+x_{11} \cdot t^2 \cdot v^3+x_{12} \cdot t^3+x_{13} \cdot t^3 \cdot v+x_{14} \cdot t^3 \cdot v^2+x_{15} \cdot t^3 \cdot v^3+x_{16} \cdot t^4+x_{17} \cdot t^4 \cdot v+x_{18} \cdot t^4 \cdot v^2+x_{19} \cdot t^4 \cdot v^3$$

Nach Einfügen zusätzlicher äquidistanter Stützstellen, Extrapolation des Datenbereiches, sowie Anwendung quadratischer

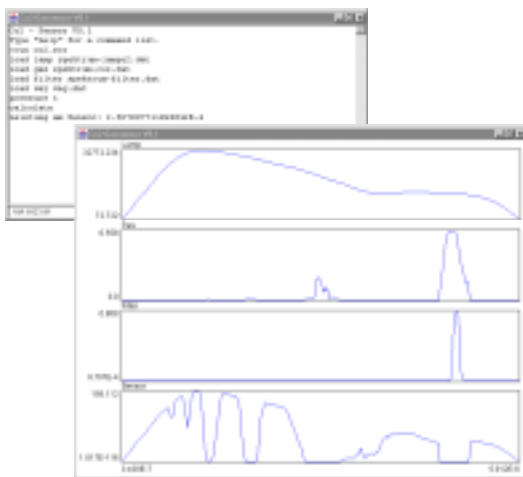
Spline-Interpolation ergibt sich eine gute Übereinstimmung von Messdaten und Approximation.



**Abb.17: Approximationsergebnis**

Die erhaltenen Ergebnisse wurden in das reale System integriert, indem die Fitfunktion in den Mikro-Controller der Auswertelektronik programmiert wurde, welche die Kompensation vornimmt.

## 6.4 Anwendung des Modells



**Abb.18: Oberfläche des Systemmodells**

Abb.18 zeigt eine Simulation unter Einsatz der im vorherigen Abschnitt vorgestellten Spektren. Die Eingabekonzole erlaubt dem Anwender, über vordefinierte Kommandos die Kennlinien der Lampe, des Gases und des Filters einzulesen. Darüberhinaus wird der Umgebungsdruck angegeben, sowie die Verteilung der Laufwege, die ebenfalls in einer Datei abgelegt sind.

Zur Erleichterung umfangreicher Untersuchungen existiert ebenfalls die Möglichkeit, die Simulation über vorgefertigte Skripte ablaufen zu lassen.

Das Ergebnis ist die am Sensor ankommende Intensität der gedämpften Lampenstrahlung, welche, über den Wellenlängenbereich integriert, die für das Ausgangssignal verantwortliche Leistung ergibt [4].

In Abb.18 ist die Strahlung am Sensor als unterste Kurve dargestellt. Die Auslöschung im relevanten Bereich der charakteristischen Banden von  $\text{CO}_2$  ist deutlich erkennbar, die angezeigte Reststrahlung resultiert aus einer Überhöhung der numerischen Randeffekte gegenüber der Auslöschung. Das Eingangssignal ist ausserhalb des Filterbereichs auf wenige Prozent reduziert.

## 6.5 Verwertung der Ergebnisse

Zunächst besteht die primäre Aufgabe des Gesamtmodells darin, die Absorption von Strahlung in einem definierten Wellenlängenbereich anzuzeigen, gemäß dem realen Verhalten des Gassensors, und somit die Identifikation ausgewählter Medien im Inneren der Strahlungskammer nachzuvollziehen.

Darüberhinaus kann das Modell auch weitere Dienste im Sinne einer entwurfsbegleitenden Anwendung leisten, indem mit dem bisher beschriebenen Funktionsprinzip eine Inversionsrechnung bezüglich der Modellparameter durchgeführt wird. Die dabei zu untersuchenden Fragestellungen beziehen sich beispielsweise auf die erreichbare Systemauflösung für ein spezielles Design und auf den Zusammenhang zu noch detektierbaren Stoffkonzentrationen, die über die Güte des Endprodukts mitbestimmen. Ein Ergebnis derartiger Untersuchungen kann eine Aussage darüber treffen, welche Laufwege benötigt werden, um über die Absorption der Strahlung das enthaltene Medium zu detektieren. Aus diesen Informationen lassen sich über weitere optische Simulationen direkte Rückschlüsse auf die benötigte Geometrie des Sensorsystems ziehen.

Das endgültige Ziel der Betrachtungen liegt in der Optimierung hinsichtlich der Modellparameter, d.h. in der Suche nach einem

bestmöglichen Designentwurf.

## 7 Zusammenfassung

In diesem Artikel wurde ein grundlegender Ansatz eines Systemmodells eines IR-Gassensors präsentiert. Im ersten Teil wurde die Optimierung des parametrisierbaren optischen Modells der Absorptionszelle und der IR-Quelle vorgestellt. Das Optimierungsziel war dabei das Auffinden eines hinsichtlich der Positionstoleranzen der Quelle robusten Designs. Beim ursprünglichen Entwurf der Gasanalysekammer war es notwendig, das Einfügen der Quelle in die Kammer manuell und mit einer Leistungsmessung durchzuführen. Das optimierte Design garantiert nun über den gesamten Toleranzbereich der Quellposition eine definierte Systemperformance. Somit wird eine automatische Bestückung der Analysechamber mit der IR-Quelle ermöglicht.

Weiterhin dienen die mit Hilfe des optischen Modells berechneten Strahlwege als Eingabedaten für die Gesamtmodellsimulation. Im zweiten Teil wurde auf der Grundlage der geltenden physikalischen Gesetze ein erstes Systemmodell vorgestellt, welches die Leistung am IR-Sensor aus den optischen Eingangsdaten, den Spektren der Lampe, des Gases und des Filters berechnet. Der modulare Aufbau der Modellkomponenten ermöglicht einen flexiblen Einsatz hinsichtlich einer entwurfsbegleitenden Simulation. Das angestrebte Ziel ist die Optimierung des Systemdesigns. Weiterhin erlaubt der vorgestellte Ansatz eine maximale Übertragbarkeit der Ergebnisse, so dass eine Entwurfsadaptation hinsichtlich der Detektion von Gasen mit unterschiedlichem Absorptionsverhalten ermöglicht wird.

## Literatur

- [1] Günzler, H., Heise, H. MIR-Spektroskopie. Ort: VCH, 1996
- [2] Sieber, I., Nüssen, O.: Optical Modellierung eines IR-Gassensors im Hinblick auf eine Entwurfsoptimierung.

Statusseminar zum Verbundprojekt OMID, 9.-10. Oktober 2001 (Bremen), S. 103-112

[3] Sieber, I., Eggert, H., Suphan, K.-H., Bechtold, S.: Simulation and Modeling of a Chamber of Analysis of an IR Gas Sensor. MICRO SYSTEM Technologies 2001, (2001), S. 71-76

[4] Sieber, I., Eggert, H., Suphan, K.-H., Nüssen, O.: Optical Modeling of the Analytical Chamber of an IR Gas Sensor. Design, Test, Integration, and Packaging of MEMS/MOEMS 2001. Proc. of SPIE, Vol. 4408 (2001), S. 272-282

[5] W. John, H. Luft, W. Groß (Hrsg.): Methoden und Werkzeuge zum Entwurf von Mikrosystemen, Tagungsband, 8. GMM-Workshop, 2.12. und 3.12.1999, Berlin

[6] Körtvélyessy, László von: Thermoelentpraxis. Vulkan-Verlag, Essen 1987

[7] J. Kirchhoff, St. Bechtold, K.-H. Suphan: Entwicklung und Simulation des Mikrosystems Thermopile-Sensor, in: W. John, H. Luft, W. Groß (Hrsg.): 8. Workshop, Methoden und Werkzeuge zum Entwurf von Mikrosystemen, Paderborn 1999.

# EIN HOCHAUFLÖSENDES RADIOMETRISCHES MEßSYSTEM ZUR QUALITÄTSKONTROLLE SCHNELL LAUFENDER MATERIALBAHNEN

D. Richter, G. Gärtner, VacuTec Meßtechnik GmbH Dresden  
Email: Dietmar.Richter@vacutec-gmbh.de

## Kurzfassung

Radiometrische Meßsysteme zur Kontrolle der Dicke oder Flächenmasse werden in der Industrie oftmals bei der Produktion von schnell laufenden Materialbahnen eingesetzt. Die Firma VacuTec Meßtechnik GmbH, Dresden, hat sich auf die Entwicklung und Produktion von strahlungssensitiven Detektoren spezialisiert, die in solchen Meßsystemen zum Einsatz kommen. Zur vollständigen Qualitätskontrolle wurde eine Mehrzellen-Ionisationskammer entwickelt, die in einem Meßvorgang die gesamte Breite der Materialbahn erfassen kann und dabei eine hohe Ortsauflösung erreicht. Sie ersetzt herkömmliche Verfahren mit traversierenden Meßköpfen.

Im Rahmen des Verbundprojektes OMID wurden stationäre und dynamische Prozesse in einer Ionisationskammer bei der Bestrahlung mit  $\gamma$ - und Röntgenstrahlung modelliert und simuliert. Dabei wurde die signalverarbeitende Elektronik (Vorverstärker) mit einbezogen.

Die entstandene Software versetzt die Firma VacuTec Meßtechnik GmbH in die Lage, vor dem Aufbau von Funktionsmustern Varianten zu berechnen und das Meßsystem entsprechend den vorgegebenen Randbedingungen gezielt zu optimieren. Dadurch läßt sich der Entwicklungsaufwand für die neue Detektorfamilie deutlich senken.

## 1 Einleitung

Zur Qualitätskontrolle und zur Steigerung der Produktivität bei der Herstellung schnell laufender Materialbahnen sind vielfältige Informationen über Materialparameter (Zusammensetzung, Dicke, Flächenmasse u.a.) und Oberflächenparameter (z.B. Glanz, Rauheit, Leitfähigkeit, Verschleiß, Korrosionsschutz)

erforderlich. Verbreitete Verfahren zur Bestimmung von Materialdicken und Flächenmassen nutzen radiometrische Meßsysteme. Sie kommen in der Papier- und Kunststoffindustrie, aber auch in Walz- und metallverarbeitenden Werken zum Einsatz.

Das mittelständische Unternehmen VacuTec Meßtechnik GmbH in Dresden hat sich mit seinen ca. 35 Beschäftigten unter anderem auf die Entwicklung und Produktion von Sensoren (Ionisationskammern, Proportionalzählrohre, Halbleiterdetektoren und Geiger-Müller-Zählrohre) spezialisiert, die in solchen radiometrischen Meßsystemen zum Einsatz kommen.

Speziell zur Qualitätskontrolle schnell laufender Materialbahnen wurde eine neue Detektorfamilie entwickelt, die eine vollständige Messung mit hoher räumlicher Auflösung über die gesamte Materialbahnbreite ermöglicht.

In den folgenden Abschnitten wird nach einem kurzen Überblick über die Wirkungsweise radiometrischer Meßsysteme und herkömmliche Verfahren, die neue Detektorfamilie vorgestellt und dargelegt, wie sich deren Modellierung und Optimierung im Rahmen des Verbundprojektes OMID einordnet.

## 2 Meßprinzip

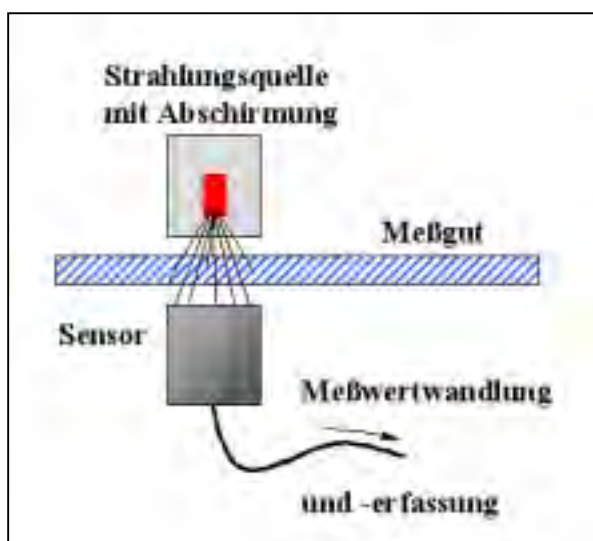
Unter dem Begriff "radiometrische oder radiographische Verfahren" werden alle Applikationen zusammengefaßt, bei denen ionisierende Strahlung in ihrer Wechselwirkung mit dem Meßgut registriert und ausgewertet wird. Diese Verfahren bestehen generell aus **drei Komponenten**. Das ist einerseits die **Quelle** der ionisierenden Strahlung. Verwendet werden Radionuklide (vorwiegend  $\beta$ - und  $\gamma$ -Quellen) aber auch Röntgengeneratoren. Andererseits muß ein sensitiver **Detektor** für die ge-

nutzte Strahlung vorhanden sein. In ihm werden elektrische Ladungen bzw. Ströme erzeugt. Diese werden anschließend in der dritten Komponente, der **Auswertelektronik**, verarbeitet und so die gewünschte Information über das Meßgut gewonnen.

Je nach den genutzten geometrischen Bedingungen werden die radiometrischen Meßverfahren in das Durchstrahlungsverfahren bzw. die Messung in Rückstreugeometrie eingeteilt.

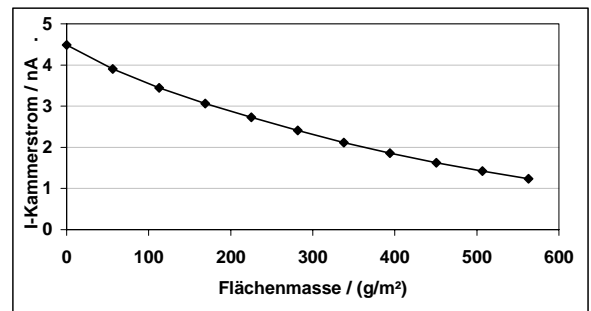
Bei der **Messung in Rückstreugeometrie** befinden sich Strahlungsquelle und Detektor auf der gleichen Seite des Meßgutes. Die Dicken- oder Schichtdickeninformation wird aus dem rückgestreuten Strahlungsanteil bestimmt. Das kann einerseits die rückgestreute Primärstrahlung sein. Es ist jedoch auch möglich, mit Röntgenquanten die charakteristische Strahlung des Meßgutes anzuregen und diese zu messen. Neben der Dickeninformation kann man hierbei auch die Zusammensetzung des Meßgutes ermitteln.

Beim **Durchstrahlungsverfahren** befinden sich Quelle und Detektor auf gegenüberliegenden Seiten des Meßgutes. Der Effekt der Absorption und Streuung der ionisierenden Strahlung im Meßgut wird ausgenutzt, um eine Information über die Dicke des Meßgutes abzuleiten. Das Meßprinzip wird in der Abbildung 1 verdeutlicht.



**Abb. 1: Prinzip einer radiometrischen Messung in Durchstrahlungsgeometrie**

Die das Meßgut durchdringende Strahlung erzeugt in einer Ionisationskammer einen Strom, der als Maß für die Dicke bzw. Flächenmasse des Meßgutes ausgewertet wird. Als Beispiel dafür wird in der Abbildung 2 die Kennlinie einer Kalibrierung mit Aluminiumfolien dargestellt. Mit wachsender Dicke (Flächenmasse) der Folie sinkt der Strom in der Ionisationskammer.



**Abb. 2: Kennlinie einer Kalibrierung mit Aluminiumfolien (Abhängigkeit des Kammerstromes von der Meßgutdicke)**

Die Auswahl einer geeigneten Strahlungsquelle und Ionisationskammer, sowie der entsprechenden Auswertelektronik muß an die Spezifika des Meßgutes (Flächenmasse, Zusammensetzung, Geschwindigkeit, geometrische Form, Genauigkeitsanforderungen etc.) und die geometrischen Bedingungen an der Produktionsanlage angepaßt werden.

Im Rahmen des Verbundprojektes OMID beschränkten wir uns auf Untersuchungen in Durchstrahlungsgeometrie und Ionisationskammern als Detektor.

Eine Übersicht über industriell häufig verwendete Radionuklide wird in der Tabelle 1 gegeben. Im Bereich der Papier- und Kunststoffindustrie mit geringeren Flächenmassen kommen vorwiegend  $\beta$ -Strahler zum Einsatz. In der Metallindustrie mit größeren Flächenmassen werden zumeist  $\gamma$ -Strahler verwendet.

Zum Schutz der Anlagenbetreiber vor der ionisierenden Strahlung sind entsprechend den Vorgaben der Strahlenschutzverordnung umfangreiche Abschirmungsmaßnahmen erforderlich. Diese müssen sowohl den Produktionsprozeß, als auch Wartungsphasen oder Ruhezeiten der Anlage berücksichtigen.



**Tabelle 1: Häufig verwendete Radionuklide für radiometrische Meßaufgaben (Zahlenwerte aus [1]).**

Strahlerart	Nuklid	max. Energie / MeV	Halbwertszeit	Flächenmassebereich / (g/m <sup>2</sup> )
β-Strahler	<sup>147</sup> Pm	0,23	2,62 a	6...150
	<sup>85</sup> Kr	0,67	10,7 a	20...1000
	<sup>90</sup> Sr	2,27	28,6 a	50...5000
		Energie / MeV	Halbwertszeit	Flächenmassebereich / (kg/m <sup>2</sup> )
γ-Strahler	<sup>241</sup> Am	0,06	432 a	10...100
	<sup>137</sup> Cs	0,66	30,1 a	50...200
	<sup>60</sup> Co	1,17; 1,33	5,27 a	80...400

**Tabelle 2: Verwendung von Röntgenstrahlung bei radiometrischer Schichtdickenmessung: Abhängigkeit der zu wählenden Hochspannung der Röntgenröhre von charakteristischen Flächenmassen nach [1]**

Röhrenspannung / kV	10	30	60	100	150
typische Flächenmasse / (kg/m <sup>2</sup> )	0,1	1	10	50	100

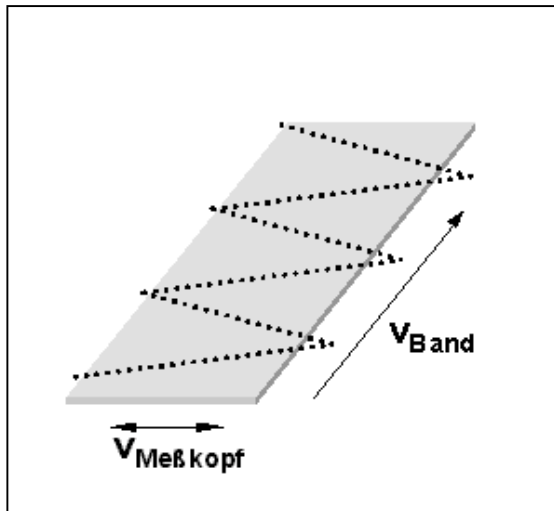
Mehrere Vorteile bietet der Einsatz von Röntgengeneratoren. Einerseits ist die Abschaltbarkeit gewährleistet. Weiterhin kann man über die Wahl der Hochspannung der Röntgenröhre die Energie (Härte) der Strahlung einstellen und für den gewünschten Flächenmassebereich optimieren. In der Tabelle 2 werden in Abhängigkeit von der gewählten Hochspannung der Röntgenröhre typische Flächenmassen aufgelistet.

### 3 Herkömmliche Verfahren

Herkömmliche radiometrische Verfahren unter Verwendung von Ionisationskammern sind dadurch gekennzeichnet, daß als Strahler vorwiegend punktförmige Quellen zum Einsatz kommen. Die edelgasgefüllten Ionisationskammern sind der Geometrie angepaßt und bestehen in der Regel aus einem zylinderförmigen Grundkörper mit einem speziell ausgewählten Eintrittsfenster für die Strahlung. Einige von der Firma VacuTec Meßtechnik GmbH produzierte Detektoren werden in der Abbildung 3 vorgestellt. Die Größe der Kammer und die Auswahl des Füllgases, des Fülldruckes und des Fenstermaterials ist von der Meßaufgabe abhängig.

**Abb. 3: Zylinderförmige Ionisationskammern der VacuTec Meßtechnik GmbH**

Um bandförmige Materialien ausmessen zu können, sind die Systemkomponenten Strahler und Ionisationskammer bei herkömmlichen Verfahren beweglich angeordnet und oszillieren quer zur Transportrichtung der Materialbahn (traversierende Meßsysteme). Durch die Überlagerung der beiden Bewegungen erhält man Informationen vom Meßgut entlang eines zickzackförmigen Weges. In der Abbildung 4 wird dieses Meßprinzip verdeutlicht. Man gewinnt so stichprobenartige Kenntnisse über das Meßgut, da sich nur wenige Prozent der Materialfläche erfassen lassen.



**Abb. 4: Meßprinzip des traversierenden radiometrischen Meßverfahrens mit Einzelionisationskammern (..... effektive Meßspur)**

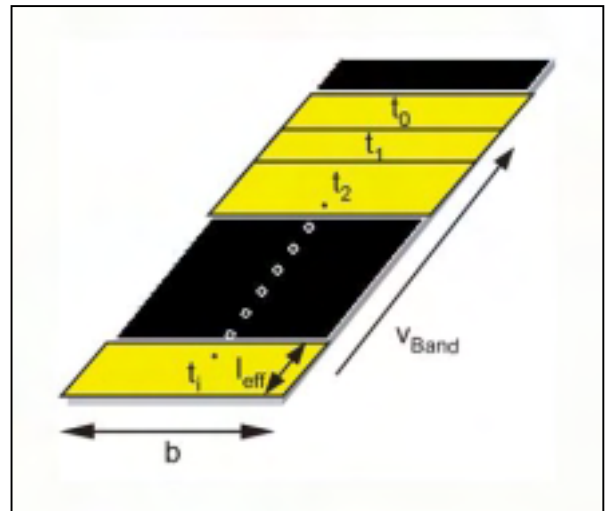
Dies ist aus Sicht der Qualitätskontrolle oftmals unbefriedigend. Als Ziel wird zunehmend die vollständige Information über die relevanten Meßparameter angestrebt.

Mit der von der VacuTec Meßtechnik GmbH neu entwickelten Ionisationskammer, die im folgenden Abschnitt vorgestellt wird, ist es möglich, bei hoher Ortsauflösung die gesamte Breite einer Materialbahn auszumessen und den hohen Anforderungen der Qualitätskontrolle bei der Produktion von Materialbahnen gerecht zu werden.

#### **4 Mehrzellen-Ionisationskammer zur Qualitätskontrolle mit hoher Auflösung**

Zur vollständigen Messung des Profils einer Materialbahn wurde eine Ionisationskammer entwickelt, die nicht mehr bewegt wird. Ihr Gasvolumen ist über die Breite in mehrere voneinander unabhängige Meßzellen unterteilt (zur Zeit maximal 64), die einzeln ausgelesen werden. So wird eine hohe Ortsauflösung erreicht. Für größere Materialbahnbreiten ist eine kaskadenartige Anordnung mehrerer solcher Ionisationskammern realisierbar.

Das Meßprinzip wird in der Abbildung 5 veranschaulicht.

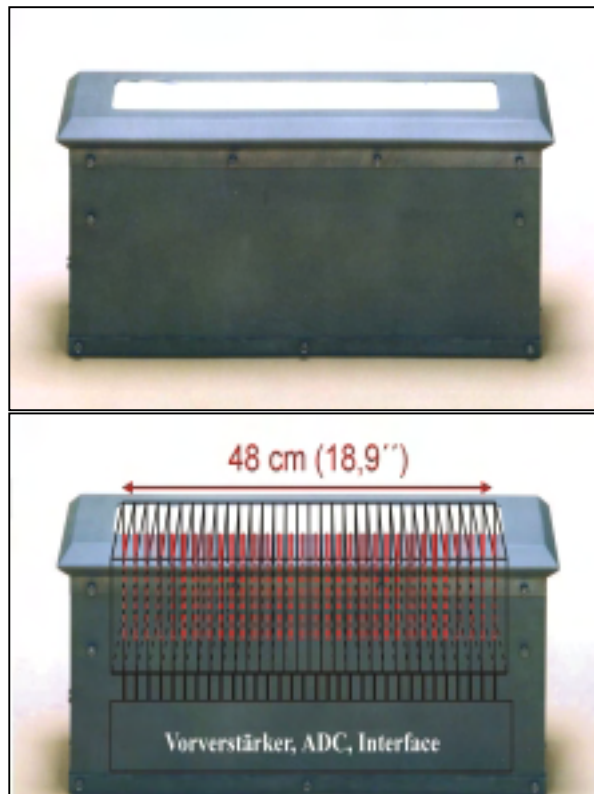


**Abb. 5: Meßprinzip des radiometrischen Meßverfahrens mit einer Mehrzellen-Ionisationskammer hoher Ortsauflösung (Erläuterung im Text)**

In Abhängigkeit von der Bandgeschwindigkeit  $v_{\text{Band}}$  und der Zeit eines Meßzyklus'  $t_i$  läßt sich die Materialbahn vollständig ausmessen. Hell dargestellt ist die auf das Material projizierte effektive Meßfläche als Produkt der Detektor- bzw. Materialbahnbreite  $b$  und der effektiven Meßlänge  $l_{\text{eff}}$ . Im zeitlichen Verlauf entstehen parallele Meßstreifen. Diese können aneinander anschließen oder sich überlappen (Redundanz, Kontrollfunktion).

Die Fotografie einer solchen, von der VacuTec Meßtechnik GmbH produzierten und patentierten Ionisationskammer, wird im oberen Teil der Abbildung 6 gezeigt. Im unteren Teil der Abbildung ist die innere Aufteilung der Ionisationskammer in die einzelnen Zellen angedeutet. Vorverstärker, ADC und Interface sind im Gehäuse mit untergebracht.

Der Einsatz der neuen Ionisationskammer zieht Veränderungen im gesamten Meßsystem nach sich. Eine linienförmige Strahlungsquelle ist besser an die Geometrie angepaßt als die bisher verwendeten Punktquellen. Für die  $\beta$ -Strahler  $^{85}\text{Kr}$  und  $^{147}\text{Pm}$  sind linienförmige Quellen verfügbar. Für andere Strahler läßt sich ein linienförmiger Quellverlauf durch die Kombination mehrerer Punktquellen realisieren.

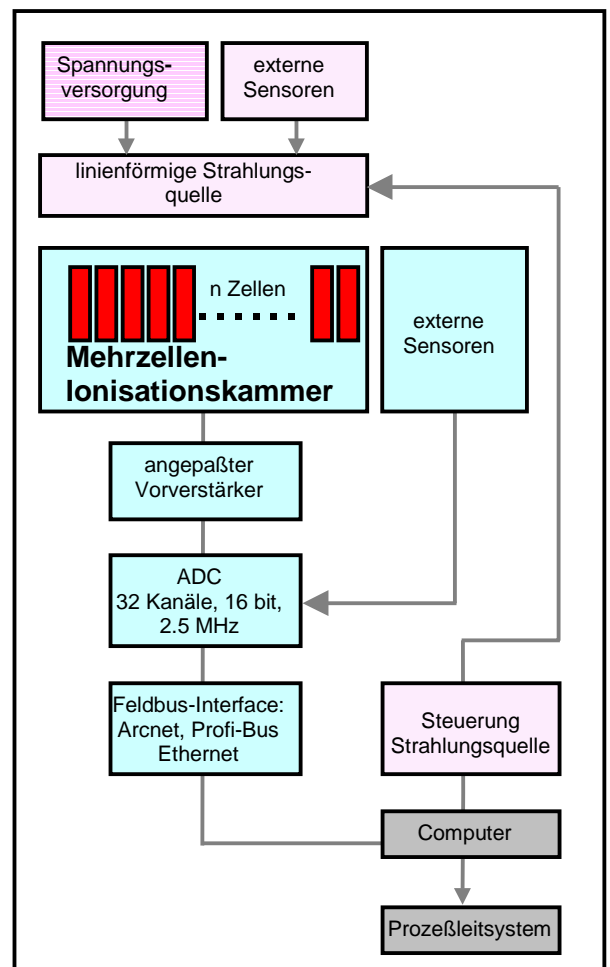


**Abb. 6: Fotografie einer Mehrzellen-Ionisationskammer (oberer Teil) und skizzierte Aufteilung in die einzelnen Meßzellen (unterer Teil)**

Weiterhin ergeben sich völlig neue Anforderungen an die Elektronik, da statt einer Kammer die Signale aller Meßzellen (maximal 64 pro Kammer) auszulesen und weiterzuverarbeiten sind.

Ein Blockschaltbild, in dem die einzelnen Komponenten des Meßprozesses dargestellt sind, wird in der Abbildung 7 gezeigt.

Jede Meßzelle des Mehrzellen-Detektors wird separat über einen Vorverstärker ausgelesen und die Strominformation einem Analog-Digital-Wandler (ADC) zugeführt. In den ADC können zusätzlich Signale externer Sensoren eingespeist werden (z.B. Sensoren für Feuchtigkeit oder Druck). Er kann durch externe Sensoren (z.B. für die Geschwindigkeit der Materialbahn) auch getriggert werden. Die digitalen Signale des ADC werden über ein Feldbus-Interface (Arcnet, Ethernet, Profi-Bus) dem Computer zugeführt. Dieser steuert auch die Strahlenquelle. Das kann einerseits



**Abb. 7: Blockschaltbild eines Meßsystems mit Mehrzellen-Ionisationskammer**

ein Röntgenerators sein, aber auch ein Radionuklidstrahler, der aus Strahlenschutzgründen hinter eine Abschirmung gefahren werden kann. Externe Sensoren, die in die Steuerung eingreifen können, sind z.B. Temperaturüberwachungssysteme. Die gesamte Einheit Strahler, Ionisationskammer, Elektronik ist in das Prozeßleitsystem eingebunden.

Die Vorteile der beschriebenen Meßmethode mit einer Mehrzellen-Ionisationskammer gegenüber dem herkömmlichen Verfahren liegen in der vollständigen Qualitätskontrolle mit hoher Ortsauflösung und im Verzicht auf bewegte Teile. Ein Nachteil ist das erhöhte Aktivitätsinventar in einer linienförmigen Quelle im Vergleich zu einer Punktquelle. Die benötigte Aktivität hängt dabei neben den physikalisch begründeten Anforderungen, auch von den Zeitkonstanten der Messung

und der benötigten Genauigkeit der Einzelmessung ab. Stärkere Strahlungsquellen erfordern in jedem Fall stärkere Abschirmungsmaßnahmen. Hierbei kollidiert man schnell mit den Platzverhältnissen an industriellen Anlagen, die die Größe des Meßsystems begrenzen.

Vorteile bringt der Einsatz eines Röntgengenerators. Allerdings muß die nötige Stabilität und Regelbarkeit der Röntgenquelle gewährleistet sein.

## 5 Ziele und Ergebnisse im Rahmen des Verbundprojektes OMID

Mit der Kompliziertheit der Detektoren wächst der zeitliche und finanzielle Aufwand zu ihrer Entwicklung, Konstruktion und zum Bau. Einen Standarddetektor, der für alle Meßaufgaben gleich gut geeignet ist, gibt es nicht. Das System aus Strahler, Detektor und Elektronik muß vielmehr an die jeweilige Meßaufgabe unter Berücksichtigung von Kundenwünschen angepaßt werden, um optimale Ergebnisse zu erzielen.

Im Rahmen des Verbundprojektes OMID sollte gezeigt werden, daß sich aufbauend auf der Modellierung und Simulation der Strahlungsverhältnisse und der elektrischen Stromverteilung Varianten- und Optimierungsrechnungen durchführen lassen. Ziel war die Reduzierung der experimentellen Arbeiten (kosten- und zeitintensive feinmechanische und vakuumtechnische Arbeiten).

Folgende Randbedingungen wurden für die Bearbeitung im Projekt festgelegt (Problemdefinition):

- Beschränkung auf Durchstrahlgeometrie,
- Verwendung von punkt- und linienförmigen  $\gamma$ -Strahlungsquellen, sowie Erweiterung auf Röntgengeneratoren,
- Ionisationskammer mit 64 Zellen (Änderung der Zellenzahl und der geometrischen Verhältnisse wie Länge, Breite und Tiefe der Zellen leicht möglich),
- Detektion einer Materialkante, sowie (als

Erweiterung) Erfassung von Dicke und lokalen Flächenmasseschwankungen.

Der erste Schritt der Bearbeitung umfaßte die Modellierung der Prozesse, die durch die Wechselwirkung von Strahlung und Meßgut bzw. Detektor hervorgerufen werden bis hin zur Berechnung des Kammerstromes für den stationären Fall. Diese Arbeiten erfolgten durch unsere Projektpartner im Institut für Theoretische Elektrotechnik und Mikroelektronik der Universität Bremen ([2]).

Anschließend erfolgte auf dieser Basis (Generationsrate der Ladungsträger) die Berechnung und Simulation der Kammerelektronik und der sich anschließenden Elektronik (Meßverstärker). Diese Teilaufgabe wurde durch unseren Projektpartner aus der Fraunhofer Gesellschaft (Institut für Integrierte Schaltungen, Außenstelle Entwurfsautomatisierung Dresden) bearbeitet ([3]). Hierbei werden auch die dynamischen Eigenschaften des Systems mit erfaßt, deren Bedeutung darin begründet ist, daß die Meßaufgaben nicht statisch sind, sondern in kurzen Zyklen bei bewegten Materialbahnen gemessen werden muß.

Der innovative Beitrag dieser Herangehensweise liegt in der Verknüpfung der beiden Teilaufgaben zu einem Gesamtsystemmodell. So ist auf der Basis der Modellierung die Optimierung des Gesamtsystems (aber natürlich auch von Teilsystemen) möglich. Hinsichtlich des Optimierungspotentials sind eine Reihe von Veränderungen an der Kammer und der Elektronik simulierbar. Das umfaßt die Größe der Zellen der Ionisationskammer, die Stärke und Materialart des Eintrittsfensters, die Art und den Druck des Edelgases in der Kammer, aber auch Elektronikbausteine im Vorverstärker, die das zeitliche Verhalten bestimmen.

Die Stärke der entwickelten Modelle liegt wesentlich in ihrer großen Flexibilität und leichten Anpaßbarkeit an veränderte Randbedingungen.

Das Ziel der Optimierung besteht in einer möglichst großen Steilheit der Kennlinie, die man erhält, wenn man den gemessenen Strom

in der Ionisationskammer in Abhängigkeit von der Dicke des Meßgutes als relevantem Meßparameter aufträgt (siehe auch Abb. 2).

## 6 Zusammenfassung

Im Rahmen des Verbundprojektes OMID wurden Prozesse in einer Ionisationskammer einschließlich der signalverarbeitenden Elektronik (Vorverstärker) bei der Bestrahlung mit  $\gamma$ - und Röntgenstrahlung modelliert und simuliert. Diese Ionisationskammer soll in radiometrischen Meßsystemen zur Qualitätskontrolle an schnell laufenden Materialbahnen eingesetzt werden.

### An dieser Aufgabe arbeiteten gemeinsam:

- die Firma VacuTec Meßtechnik GmbH, Dresden,
- das Institut für Theoretische Elektrotechnik und Mikroelektronik der Universität Bremen (ITEM) und
- das Institut für Integrierte Schaltungen, Außenstelle Entwurfsautomatisierung Dresden (EAS) der Fraunhofer Gesellschaft.

### Als Ergebnis entstanden:

- Software zur Berechnung der Ströme in der Ionisationskammer für den stationären Fall bei Bestrahlung mit  $\gamma$ - und Röntgenquanten (lauffähig auf den PC der Firma VacuTec),
- Software zur Simulation der Kammerelektronik einschließlich dynamischer Effekte und
- ein Meßplatz zur Verifikation der Rechenergebnisse.

Der Firma VacuTec Meßtechnik GmbH wird dadurch ermöglicht, den Entwicklungsaufwand für die neue Detektorfamilie zu senken. Vor dem Aufbau von Funktionsmustern können Varianten gerechnet werden und optimierte Modelle entsprechend den vorgegebenen Randbedingungen entwickelt werden. Besonders positiv ist der Synergieeffekt in der Arbeit der Projektpartner zu erwähnen.

Eine weitere Zusammenarbeit ist aus Sicht der VacuTec Meßtechnik GmbH wünschenswert, da eine Erweiterung der vorhandenen Modelle z.B. auf den Einsatz von  $\beta$ -Strahlung oder die Realisierung von Rückstreugeometrie ganz neue Ansätze in der physikalischen Modellierung erfordert.

## Literatur

- [1] K. Nitzsche, Schichtmeßtechnik, Vogel Buchverlag, Würzburg, 1996.
- [2] H. Bolte, R. Laur, Entwurfsunterstützung für ein radiometrisches Meßsystem – Ein physikalisches Modell der Ladungsträgergeneration, OMID-Statusseminar 2002, Karlsruhe, diese Tagung.
- [3] K.-H. Diener, E. Huck, P. Schneider, P. Schwarz, Entwurfsunterstützung für ein radiometrisches Meßsystem - Kammerelektronik und Gesamtsystemmodell, OMID-Statusseminar 2002, Karlsruhe, diese Tagung.

## Dank

Die Firma VacuTec Meßtechnik GmbH möchte sich bei allen Projektbeteiligten und der Projektleitung für die sehr gute Zusammenarbeit bedanken. Insbesondere den Herren Bolte und Prof. Laur vom Institut für Theoretische Elektrotechnik und Mikroelektronik der Universität Bremen, sowie Dr. Huck, Prof. Diener, Dr. Schneider und Dr. Schwarz vom Institut für Integrierte Schaltungen, Außenstelle Entwurfsautomatisierung Dresden der Fraunhofer Gesellschaft gebührt Dank für ihr stetes Engagement.

*Das diesem Beitrag zugrunde liegende Vorhaben „OMID-Optimierung von Mikrosystemen für Diagnose- und Überwachungseinrichtungen“ wird mit Mitteln des Bundesministeriums für Bildung, Wissenschaft, Forschung und Technologie (BMBF) unter dem Kennzeichen 16 SV 988/2 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt allein bei den Autoren.*

# ENTWURFSUNTERSTÜTZUNG FÜR EIN RADIOMETRISCHES MEßSYSTEM – EIN PHYSIKALISCHES MODELL DER LADUNGSTRÄGERGENERATION

H. Bolte, R. Laur, ITEM, Universität Bremen  
Email: bolte@item.uni-bremen.de

## Kurzfassung

Das ursprünglich geplante Einsatzgebiet von Mehrzellen-Ionisationskammern (MZIK) war die Detektion der Kantenlage schnell laufender Materialbahnen [1], [2]. Neuere Überlegungen erweitern die Einsatzmöglichkeiten der MZIK. Hier ist vor allem die lückenlose Überwachung des Querprofils der Materialbahn ohne bewegliche Teile hervorzuheben [3].

Damit auch diese Einsatzbereiche vor der Fertigung der Detektoren simuliert und optimiert werden können, ist eine Erweiterung des im Vorgängerprojekt DEMIS erstellten Grundmodells erforderlich.

Ziel ist es, das Meßsystem für jeden Einsatzbereich hinsichtlich Geschwindigkeit und Genauigkeit an die technologischen Forderungen, die seitens des Kunden oftmals durch eine Fehlertoleranz pro Zeit dargestellt werden, anzupassen. Eine sinnvolle Aussage über das zu erwartende Signal läßt sich nur dann treffen, wenn neben einer korrekten Berechnung der Ladungsträgergeneration dynamische Effekte und der Vorverstärker berücksichtigt werden.

Aus diesem Grund wurde ein modulares Gesamtsystemmodell konzipiert, dessen einzelne Module und Schnittstellen arbeitsteilig an der FhG IIS/EAS und der Universität Bremen entwickelt wurden.

Ziel dieses Gesamtsystemmodells ist es, den hohen Entwurfsaufwand durch Hinterlegung der Entwicklung mit Simulationen und die Verwendung von Optimierungsverfahren zu reduzieren.

Der vorliegende Beitrag stellt das Gesamtsystemmodell und seine einzelnen Module vor. Das Modul zur Berechnung der Ladungsträgergeneration wird ausführlich erläutert, für eine nähere Betrachtung der beiden anderen Module sowie die Gesamtsystemoptimierung sei auf [4] verwiesen.

## 1. Einleitung

An der Generation des Meßsignals einer MZIK wirken vielfältige physikalische und elektronische Prozesse mit. Eine umfassende und geschlossene Betrachtung aller Effekte in einem Modell ist daher nicht möglich.

Das modular konzipierte Systemmodell ermöglicht eine Bearbeitung der Teilaspekte in verschiedenen Arbeitsgruppen. Dabei wird sichergestellt, daß eine durchgängige und einheitliche Beschreibung der Problematik bzgl. der verwendeten Parameter und Schnittstellen vorliegt. Das Gesamtsystemmodell und seine Module werden in Abschnitt 2 vorgestellt.

Abschnitt 3 stellt das Modul zur Berechnung der Generationsrate dar. Hierzu wurden im Vorläuferprojekt DEMIS Fragen untersucht, bei denen es um die Detektion der Kantenlage einer Materialbahn ging. Ein neues Einsatzgebiet für MZIK-Systeme ist die lückenlose Überwachung des gesamten Querprofils von Materialbahnen während der Produktion. Dies erfordert eine Erweiterung und Anpassung der in DEMIS erstellten Modelle an die neue Meßaufgabe. Die dazu durchgeführten Erweiterungen für das Modul Generationsrate werden in Abschnitt 3 beschrieben.

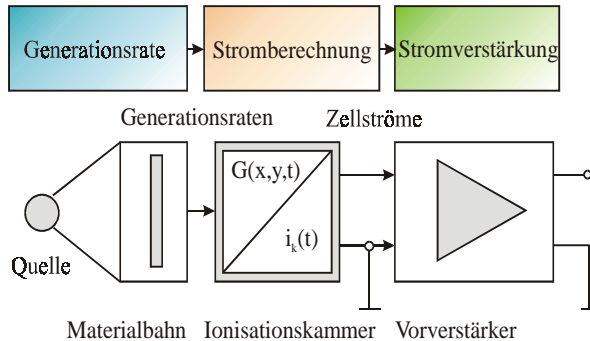
Im Abschnitt 4 werden Meßwerte im Vergleich mit Simulationsergebnissen gezeigt und Teilsystemoptimierungen beschrieben. Zum Abschluß wird in Abschnitt 5 eine Zusammenfassung und ein Ausblick gegeben.

## 2. Gesamtsystemmodell

Am Prozeß der Signalgewinnung der Mehrzellenionisationskammer sind physikalische Prozesse aus vielen verschiedenen physikalischen Domänen beteiligt. Eine exakte Abbildung in einem geschlossenen Systemmodell ist daher nicht möglich. Deshalb wurde ein modular strukturiertes, offenes Gesamt-

systemmodell mit definierten Schnittstellen als Modellstruktur gewählt (Abb. 1).

## 2.1 Struktur und Schnittstellen



**Abb. 1: Struktur des Systemmodells**

Das Gesamtsystemmodell besteht aus drei Modulen und gleicht damit im wesentlichen der technischen Struktur des Systems. Die drei Module können auch einzeln parametrisiert und für Simulations- und Optimierungsläufe eingesetzt werden (siehe auch Abschnitt 4.2). Im folgenden sollen die Module und ihre Schnittstellen näher erläutert werden.

## 2.2 Modul Generationsrate

Das erste Modul berechnet die orts- und zeitabhängige Generation von Ladungsträgern in der Mehrzellenionisationskammer. Dabei berücksichtigt es sowohl die Generation durch Absorption der Strahlung (Gamma- oder Röntgenquelle) als auch die sekundäre Generation durch die ausgelösten Photoelektronen und die Fluoreszenzphotonen. Das Modul ist variabel in den Geometrieparametern, welche die Kammer und ihr Umfeld beschreiben, den Materialkoeffizienten und hinsichtlich der Quelle.

Als Modellsprache wurde C/C++ gewählt, um Simulationszeiten zu erzielen, die noch eine Optimierung ermöglichen. Notwendig ist dies, da die Berechnung mittels Diskretisierung und Strahlverfolgung erfolgt [1], [2].

## 2.3 Modul Stromberechnung

Im zweiten Modul erfolgt die Berechnung der Kammerströme auf Basis eines dynamischen Zweiträgertransportmodells.

Ausgangspunkt hierfür ist die orts- und zeit-

abhängige Ladungsträgergenerationsrate, die im ersten Modul ermittelt wurde. Zur Sicherung der Durchgängigkeit werden für die Bestimmung der Generationsrate im Modul 1 und für die Modellierung der Kammerelektronik der gleiche Parametersatz verwendet. Dies betrifft vor allem die Strukturparameter der Kammerzellen. Bei der Kammerstromberechnung sind neben den geometrischen Abmessungen auch die Kammerspannung und Materialkoeffizienten der verschiedenen Füllgase wählbare Parameter.

Bei der Berechnung der Zellenströme wird die insbesondere bei starker Generation merkbare Feldrückwirkung der generierten Ladungsträgerpaare berücksichtigt.

Das 1D-Modell der Kammerelektronik ist als Saber-Modell implementiert, so daß eine problemlose Kopplung mit dem C/C++-Programm des Moduls „Generationsrate“ möglich wird.

## 2.4 Modul Stromverstärkung

Das dritte Modul modelliert einen wesentlichen Teil der Auswertelektronik, den Vorverstärker, der zur Verstärkung der sehr niedrigen Strompegel der MZIK-Zellen dient. Das Modul ist als Saber-Netzliste formuliert, so daß hier z.B. durch Variation von Bauelementparametern Einfluß auf die elektrischen Kenngrößen des Verstärkers genommen werden kann. An der Schnittstelle zwischen den beiden Modulen wird dabei der Strom übergeben, der in einer Zelle erzeugt wird. Eine aufwendige Kopplung entfällt, da es sich wie beim vorangegangenen Modul um ein Saber-Modell handelt.

Ausgang des Moduls ist, identisch zum technischen System, die Spannung, die sich an den Klemmen als Meßsignal abgreifen läßt.

Jedes dieser Module kann einzeln simuliert und optimiert werden. Entsprechende Ergebnisse von Teilsystemoptimierungen finden sich in Abschnitt 4.2 und für die Module Stromberechnung und Vorverstärkung in [4].

## 3. Modellerweiterungen zum Modul Generationsrate

Die Ladungsträgergeneration im Gas der MZIK beruht primär auf der Absorption von

Gamma-, Röntgen- oder Betastrahlung, je nach Anwendungsfall entsprechend des Absorptionsgesetzes:

$$I(s) = I_0 \cdot e^{-\mu s} \quad (1),$$

mit:  $I_0$  Anfangsintensität [1/s],  $s$ : zurückgelegter Weg [mm],  $\mu$ : linearer Absorptionskoeffizient [1/mm] und  $I(s)$  Intensität nach  $s$  [1/s].

Durch die Absorption wird ein Photoelektron aus dem betroffenen Gasatom gelöst, welches aufgrund seiner kinetischen Energie seinerseits weitere Gasatome ionisieren kann.

Für den interessierenden Anwendungsbereich wird die Kammer derart betrieben, daß folgende vereinfachende Annahmen mit hinreichend guter Näherung erfüllt sind:

- Als Wechselwirkungseffekt zwischen Energiestrahlung und Füllgas der Kammer kommt nur der Photoeffekt zum Tragen.
- Die Kammer wird im sog. „Ionisationskammerbereich“ betrieben, d.h. sämtliche generierten Ladungsträger werden auch abgesaugt. Ladungsträgermultiplikation durch das anliegende elektrische Feld findet nicht statt.
- Sämtliche Energie, welche durch Absorption im Kammervolumen deponiert wird, ausschließlich der Escape-Verluste durch Fluoreszenz, wird zur Generation von Ladungsträgern verwendet.

Mit Hilfe dieser Annahmen läßt sich die Generationsrate durch Multiplikation der absorbierten Photonen mit einem konstanten Faktor als Funktion des Ortes und der Zeit bestimmen. Grundlegende Arbeiten hierzu wurden im Vorgängerprojekt „DEMIS“ vorgenommen, so daß an dieser Stelle lediglich die Erweiterungen und Verallgemeinerungen aufgeführt sind. Weitere Modellgrundlagen können in [1], [2] und [5] nachgelesen werden. Die Gültigkeit der zugrundeliegenden Modellstufe ist durch Messungen verifiziert worden.

### 3.1 Berechnung des Stromes für den stationären Fall

Wird das Modul zur Berechnung der Ladungsträgergeneration als Stand-alone-Programm zur Simulation oder Optimierung verwendet, so läßt sich durch Integration der Generationsrate über die einzelnen Zellvolumina der Strom berechnen, der an der jeweiligen Zellelektrode abgegriffen werden kann. Dies gilt allerdings nur für den stationären Fall unter Vernachlässigung der Rekombination, dynamische Effekte des Ladungsträgertransportes finden erst im darauffolgenden Modul zur Stromberechnung Berücksichtigung.

### 3.2 Fluoreszenz

Im Vergleich zu den bisherigen Veröffentlichungen wird im Modell nun auch der Fluoreszenzeffekt berücksichtigt. Dieser beeinflusst vor allem die sekundäre Generation, die aus der Anzahl der absorbierten Teilchen durch Multiplikation mit einem Faktor berechnet werden kann. In der bisherigen Modellversion ohne Berücksichtigung der Fluoreszenz betrug dieser Faktor:

$$f_{gen} = \frac{(E_\gamma - E_B)}{E_{ion}} \quad (2),$$

mit  $E_\gamma$ : Quantenenergie,  $E_B$ : Bindungsenergie des herausgelösten Elektrons,  $E_{ion}$ : mittlerer Energieaufwand zur Bildung eines Ladungsträgerpaares des jeweiligen Füllgases, und beruhte auf der Stoßionisation durch Photoelektronen.

Dabei wurde die Energie, die zur Überwindung der Bindungsenergie benötigt wird, für die Generation von Ladungsträgern nicht berücksichtigt. Dies führte zu einem fehlerhaft berechneten Strom, was ein Vergleich mit Meßwerten aufzeigte. Die Einbeziehung des Fluoreszenzeffektes [6] ergab eine deutliche Verbesserung der Modellqualität.

Die Wahrscheinlichkeit  $p_{fl}$ , daß nach Absorption eines Photons vom angeregten Atom ein Fluoreszenzphoton ausgesendet wird, ist von der verwendeten Gasart und der Schale, in welcher die Absorption stattgefunden hat, abhängig. Die mittlere im Gasraum zurückgelegte Weglänge berechnet sich dabei aus



der Kammergeometrie und dem Absorptionssort. Aus der mittleren Weglänge und dem Absorptionskoeffizienten läßt sich die Absorptionswahrscheinlichkeit des Fluoreszenzphotons im Gasvolumen,  $p_{abs}$ , berechnen:

$$p_{abs} = 1 - e^{-\mu_{fl} \cdot s_{fl}} \quad (3),$$

mit:  $s_{fl}$  im Mittel zurückzulegender Weg eines Fluoreszenzphotons im Gasraum,  $\mu_{fl}$  Absorptionskoeffizient für das betreffende Gas und die Energie des Fluoreszenzphotons.

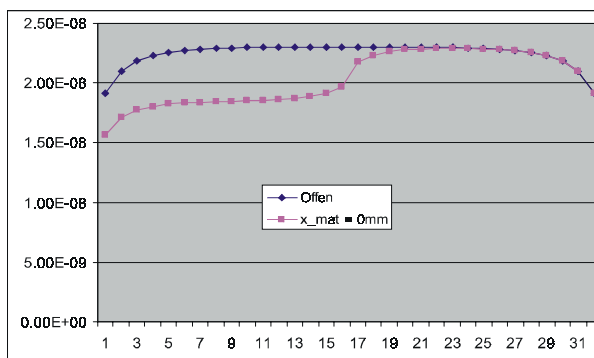
Unter Berücksichtigung von Fluoreszenz beträgt der Faktor  $f_{gen}$  also nun:

$$f_{gen} = \frac{E_{\gamma} - E_B}{E_{ion}} + \frac{E_{fl} \cdot p_{fl} \cdot p_{abs}}{E_{ion}} + \frac{(E_B - E_{fl}) \cdot p_{fl}}{E_{ion}} + \frac{(1 - p_{fl}) \cdot E_B}{E_{ion}} \quad (4),$$

mit:  $E_{\gamma}$ : Energie der Strahlung,  $E_B$ : Bindungsenergie der Schale, in welcher die Absorption stattfindet,  $E_{ion}$ : mittlerer Energieaufwand zur Bildung eines Ladungsträgerpaares innerhalb des jeweiligen Füllgases,  $E_{fl}$ : Energie des Fluoreszenzphotons,  $p_{fl}$  Wahrscheinlichkeit für das Auftreten eines Fluoreszenzphotons.

### 3.3 Linienquelle

Damit nicht nur eine Materialbahnkante, sondern das ganze Materialprofil vollständig und ohne bewegliche Teile erfaßt werden kann, ist es sinnvoll, sog. Linienquellen zu verwenden, also Strahlungsquellen, die in Richtung der Materialbahnbreite eindimensional ausgedehnt sind und diese im bestmöglichen Fall überdecken.



**Abb. 2: Linienquelle – Simulationsergebnis für eine MZIK mit 32 Zellen**

Das Modell bietet die Möglichkeit, die laterale Ausdehnung der Quelle und die Verteilung der Intensität so zu wählen, daß sich nahezu beliebige Quellengeometrien und auch Quellenverteilungen angeben lassen. Abb. 2 zeigt die normierten Zellenströme, die sich unter Verwendung einer solchen Linienquelle ergeben, wobei die Kammer in  $x$ -Richtung durch eine Materialbahn einmal mittig überdeckt und einmal unüberdeckt ist.

### 3.4 Röntgenquelle

Gasgefüllte Mehrzellen-Ionisationskammern sind in weiten Flächenmassenbereichen einsetzbar, d.h. es ist eine Anwendung für verschiedenste Materialien möglich.

Hierbei müssen allerdings gewisse Anpassungen vorgenommen werden, z.B. bezüglich der verwendeten Quellen. So werden für niedrige Flächenmassen Beta-Strahlungsquellen eingesetzt, höhere Flächenmassen erfordern dagegen Energiestahlung, wie sie Röntgen- bzw. Gammaquellen bieten. Erstere haben den Vorteil, daß der Strahlenschutz bei Ihnen leichter zu realisieren ist, allerdings emittieren sie ein kontinuierliches Spektrum an Strahlung im Gegensatz zu den zumeist monoenergetischen Gammaquellen.

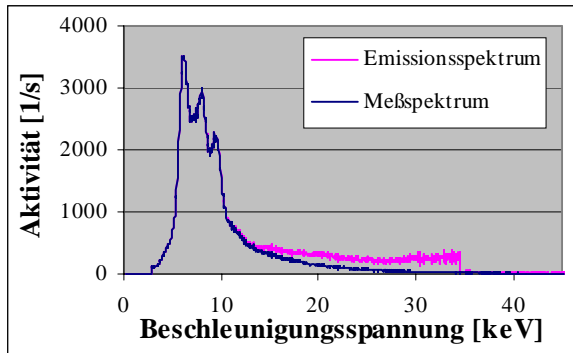
Die Einführung der Röntgenquelle erforderte eine Erweiterung des Modells. Zu lösen war vor allem das Problem der Bestimmung des Absorptionskoeffizienten als Funktion der Energie und der Kernladungszahl des bestrahlten Materials.

Zunächst galt es, das Emissionsspektrum der Quelle zu bestimmen. Dazu wurde die Quelle bei den interessierenden Beschleunigungsspannungen mit einem Standarddetektor vermessen. Auf diese Art und Weise erhält man Meßspektren, welche in die Emissionsspektren zurückgerechnet werden müssen, da die Nachweiswahrscheinlichkeit des Detektors, die ja ebenfalls eine Funktion der Detektorgeometrie, der Energie und des Füllgases ist, berücksichtigt werden muß:

$$I_{Emi}(E) = \frac{I_{Me\beta}(E)}{W_{Det}(E)} \quad (5),$$

mit:  $I_{Emi}$ : emittiertes Spektrum,  $I_{Meß}$ : gemessenes Spektrum,  $W_{Det}$ : Nachweiswahrscheinlichkeit des Detektors.

Für eine Röntgenquelle mit 49kV Beschleunigungsspannung sind gemessenes und emittiertes Spektrum in Abb. 3 gegenübergestellt.



**Abb. 3: Meß- vs. Emissionsspektrum**

Das Emissionsspektrum wird nun bezüglich der Energie diskretisiert und im weiteren so behandelt, als würde es sich um einzelne, monoenergetische Quellen handeln.

Damit ergibt sich die Wirkung einer Röntgenquelle als Summe der Wirkungen dieser monoenergetischen Quellen zu:

$$G(x, y) = \sum_{i=1}^n G_i(x, y) \quad (6),$$

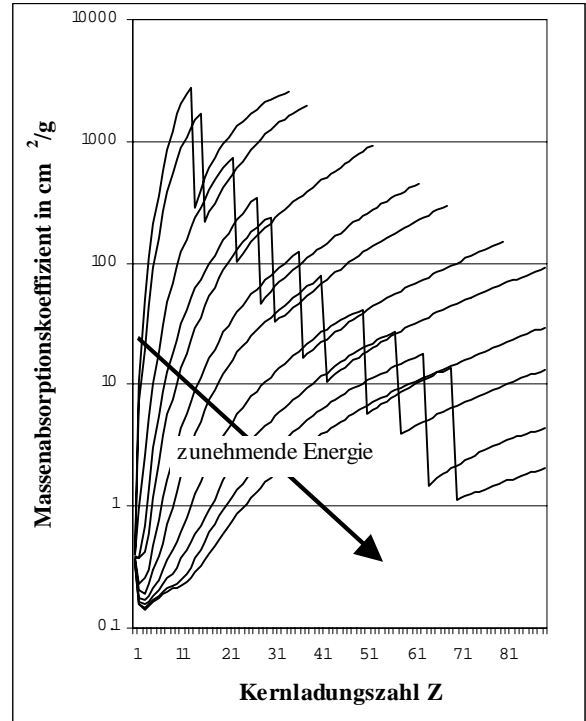
mit  $G_i(x, y)$  als Generationsrate des  $i$ -ten diskreten Energiebandes an der Stelle  $(x, y)$  in der Kammer.

Die Durchführung dieser Simulation erfordert eine automatische Berechnung der Absorptionskoeffizienten in Abhängigkeit von den Materialien und der Strahlungsenergie seitens des Modells. Dazu wurden Literaturwerte (z.B. aus [7]) für Energien  $E$  von 1,8keV bis 60keV und Kernladungszahlen  $Z$  von 1 bis 90 gefittet. Die Absorptionskanten der K- respektive der L-Schale wurden ebenfalls berücksichtigt. Das Ergebnis der automatischen Berechnung von Absorptionskoeffizienten in Abhängigkeit von der Kernladungszahl  $Z$  mit der Strahlungsenergie als Parameter ist in Abb. 4 dargestellt.

Mittels dieser Erweiterungen des Modells lassen sich Röntgenquellen als Summe von monoenergetischen Energiequellen simulieren.

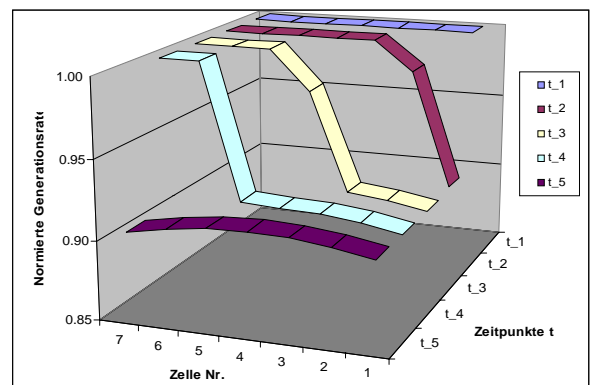
### 3.5 Transiente Simulation

Da die Materialbahnen Laufgeschwindigkeiten bis zu mehreren 10 m/s erreichen, ist die Dynamik der Kammer und der Auswertelektronik nicht vernachlässigbar.



**Abb. 4: Automatisch berechnete Massenabsorptionskoeffizienten**

Das Modul zur Berechnung der Generationsrate geht von stationären Zuständen aus. Die dynamischen Effekte des Ladungsträgertransports in der Kammer und des Verstärkungsvorgangs werden im Modul Stromberechnung bzw. im Modul Vorverstärker berücksichtigt.



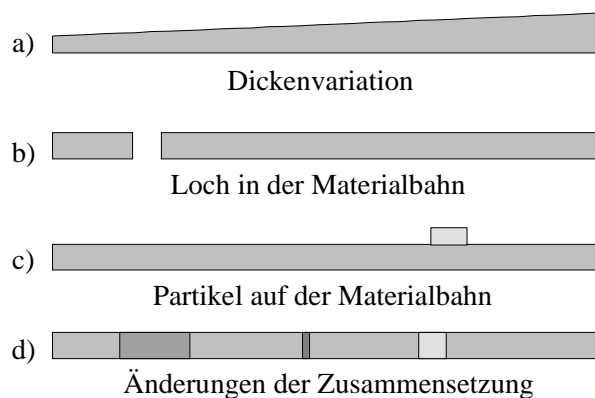
**Abb. 5: Zellgenerationsraten als  $f(t)$**

Damit das Gesamtsystem dennoch durchgängig ist, wurde das Modul zur Berechnung der

Generationsrate dahingehend erweitert, daß zeitliche Änderungen der Materialbahn oder der Materialbahnkante simulierbar sind. Entsprechende Simulationsergebnisse für eine gleichmäßig fortschreitende Überdeckung der MZIK in  $x$ -Richtung (Detektion einer Materialbahnkante) sind in Abb. 5 dargestellt.

### 3.6 Überwachung des Materialbahnprofils

Neben der Detektion der Kantenlage lassen sich MZIK ebenfalls zur lückenlosen Überwachung der Materialbahn auf ihrer gesamten Breite verwenden. Dabei sollen Unregelmäßigkeiten auf bzw. in der Materialbahn detektiert werden. Mögliche Fehlerursachen sind in Abb. 6 dargestellt.



**Abb. 6: Fehlervarianten der Flächenmasse**

Um diese neue Meßaufgabe auch im Modell berücksichtigen zu können, wird die Materialbahn in  $x$ -Richtung abschnittsweise als Mischung aus beliebig vielen Materialien frei wählbarer Dicke angelegt. Der resultierende Absorptionskoeffizient des jeweiligen Abschnitts ergibt sich dann zu:

$$\mu_{res} = \sum_{i=1}^n \mu_i \cdot p_i \quad (7),$$

mit:  $\mu_i$  als Absorptionskoeffizient des Materials  $i$  und  $p_i$  als prozentualen Anteil des Materials  $i$  an der Mischung im beschriebenen Abschnitt. Mittels dieses Ansatzes lassen sich sämtliche Fehlerarten a) - d) beschreiben [8]. Makrofunktionen im Modell sorgen für eine automatische Generation der Materialbahnbeschreibungen für gewählte Fehler. Es ist aber auch eine manuelle Eingabe der Materialbahn

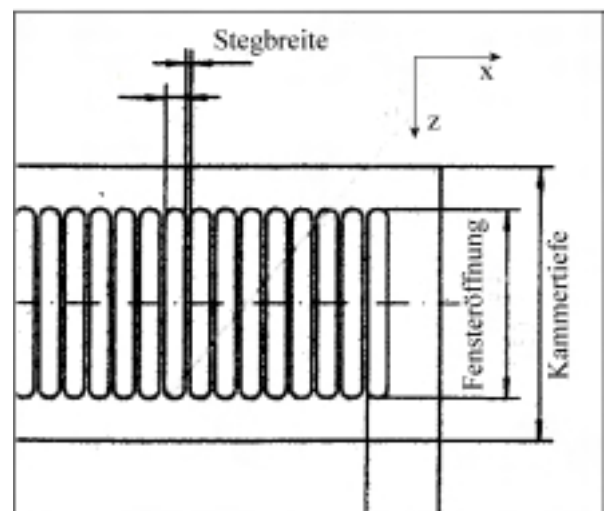
über eine Textdatei möglich.

Aufgrund des hohen Kammerinnendrucks und den verwendeten dünnen Fensterfolien wird bei den großen MZIK-Arrays ein Bersten derselben durch den Abschluß des Kammergehäuses mit einer Fenstermaske verhindert (s. Abb. 7).

Diese Maske besteht aus Stahl und stützt die Folie über jeder einzelnen Zelltrennwand mittels eines Steges ab. Da diese Maske relativ dick ist, beeinflusst sie die Strahlung erheblich und muß daher ebenfalls in dem Modell Berücksichtigung finden. Dies geschieht, indem die Maske auf gleiche Art und Weise wie eine Materialbahn angelegt wird. Beschreiben läßt sie sich durch die Angabe vierer Parameter:

1. Die Fensteröffnung in  $z$ -Richtung
2. Die Stegbreite
3. Die Dicke in  $y$ -Richtung
4. Das Maskenmaterial

Die anderen Parameter werden aus den weiteren Angaben zur Kammergeometrie abgeleitet.

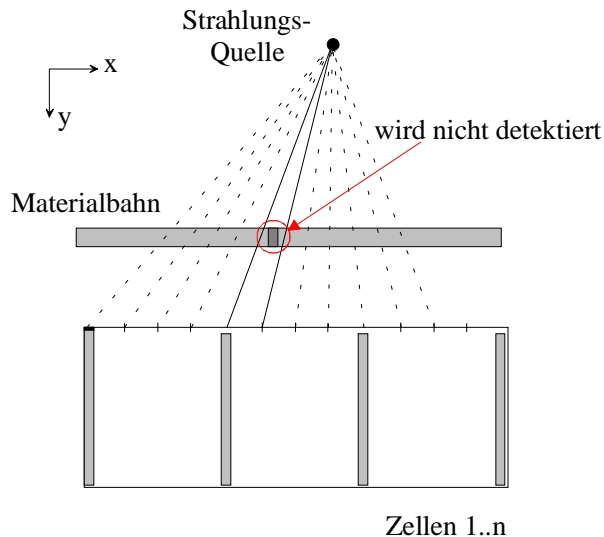


**Abb. 7: Skizze der Fenstermaske (Draufsicht,  $x$ - $z$ -Ebene)**

Die Neukonzeption der Materialbahn und die Erweiterung um die Fenstermaske erfordert auch die Berechnung einer minimalen Strahlenzahl, welche gewährleistet, daß das kleinste Materialbahnsegment noch hinreichend genau aufgelöst wird. Wählt man die Diskretisierung zu grob, werden kleine Stö-

rungen nicht detektiert (siehe Abb. 8).

Anhand des kleinsten Segments, ob in der Materialbahn oder durch die Stegbreite der Fenstermaske definiert, wird also die notwendige Auflösung bestimmt.

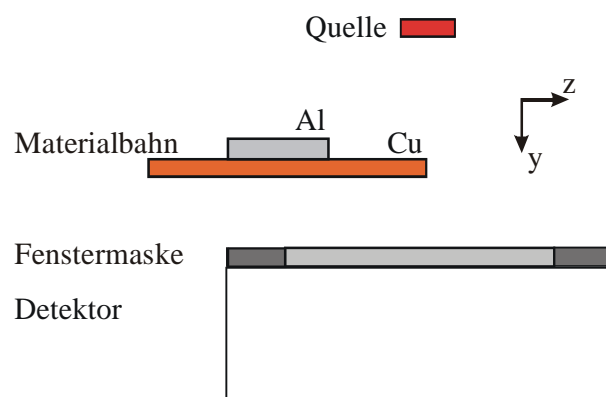


**Abb. 8: Detektionsproblem bei zu grober Diskretisierung**

Die Diskretisierung erfolgt äquidistant. Da das Modul zur Berechnung der Generationsrate sehr geringe Simulationszeiten hat, wirkt sich diese simple Auflösungsbestimmung kaum nachteilig auf die Simulations- bzw. Optimierungsdauer aus.

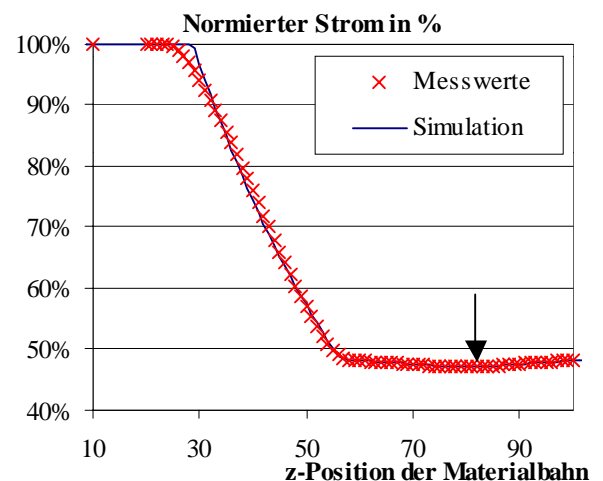
#### 4.1 Modellverifikation

Für die Flächenmasseüberwachung wurden von der Firma VacuTec verschiedene Messungen mit einem MZIK-Array mit 64 Zellen durchgeführt.



**Abb. 9: Skizze des Meßvorganges**

Die Messung, welche hier näher dargestellt werden soll (Abb. 9), untersucht die Entwicklung des Stromes einer Zelle bei sukzessiver Überdeckung der Kammer aus  $z$ -Richtung (im Gegensatz zur Meßaufgabe Materialbahnkante, bei welcher die Überdeckung aus  $x$ -Richtung erfolgt) mit einer Materialbahn, hier ein 0,5mm starkes Kupferblech, auf dem sich 25mm nach Beginn der Kante ein ebenfalls 0,5mm starkes Aluminiumblech befindet. Für die Messung wird die Materialbahn mittels eines Schlittens millimeterweise vorgeschoben und dann der resultierende Strom gemessen. Normiert man diesen auf den maximalen Strom im unüberdeckten Fall, so erhält man Werte zwischen 0 und 1 (Meßwerte: Kreuze in Abb. 10). Am Anfang befindet sich das Kupfer noch nicht über der Kammer, der Strom ist also genauso groß wie im unüberdeckten Zustand (100%). Mit zunehmender Überdeckung sinkt der Strom in der gemessenen Zelle. Das Aluminium führt noch einmal zu einer weiteren Verringerung des Stromes (Pfeil in Abb. 10). Die Wirkung des Aluminiums ist wesentlich geringer als die des Kupfers, da sein Absorptionskoeffizient geringer ist. Die durchgezogene Linie stellt die entsprechenden Simulationsergebnisse dar. Es läßt sich erkennen, daß Simulation und Messung sehr gut übereinstimmen.

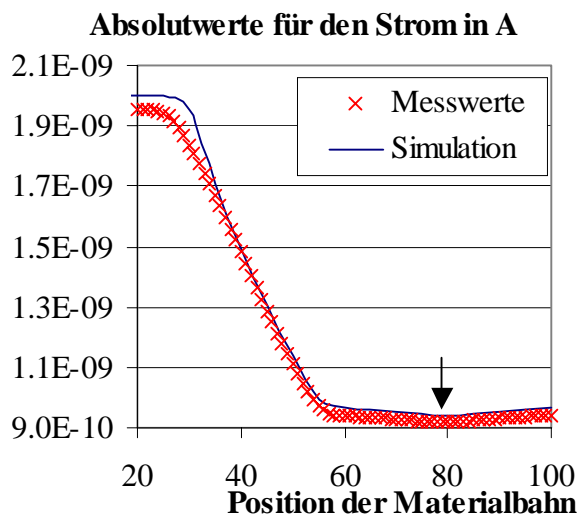


**Abb. 10: Vergleich Messung – Simulation für die normierten Werte**

In Abb. 11 ist der entsprechende Vergleich für die *absoluten* Werte des Zellstromes dargestellt. Hier ergibt sich ein sichtbarer

Unterschied zwischen Messung und Simulation. Die simulierten Ströme sind durchweg höher als die gemessenen. Der Grund hierfür ist darin zu suchen, daß die Energieeffizienz des Detektors niedriger liegt, als es die vereinfachenden Annahmen voraussetzen. So sind z.B. Rekombinations- oder Rückstreuverluste nicht im Modul zur Berechnung der Generationsrate enthalten. Erstere sind allerdings Bestandteil des Moduls Stromberechnung, das in [4] näher beschrieben wird.

Eine Kombination dieser Module dürfte also eine weitere Annäherung der Simulations- an die Meßergebnisse bewirken. Alles in allem liegt der Fehler aber unter 5%, was auf eine gute Berücksichtigung der wesentlichen Effekte im Modell schließen läßt.



**Abb. 11: Vergleich Messung – Simulation für die absoluten Stromwerte**

## 4.2 Teilsystemoptimierung

Der modulare Aufbau des Systemmodells erlaubt nicht nur eine separate Simulation der einzelnen Module, sondern auch eine separate Optimierung. Im folgenden sollen hierzu einige Untersuchungen, die im Verlauf des Verbundprojektes durchgeführt wurden, vorgestellt werden.

Bei der Teilsystemoptimierung des Moduls Generationsrate wurden für die Meßaufgabe Materialbahnkante drei zum Teil gegenläufige Ziele verfolgt:

1. die Maximierung der absoluten Zellströme für einen guten Signal-

Rauschabstand,

2. die Maximierung der Steigung der Stromflanke, welche die Kantenposition repräsentiert,
3. die Minimierung der Durchstrahlung durch die Abschirmungen zwischen den einzelnen Zellen

Vor allem Ziel 3. ist kaum mit den anderen zu vereinbaren. Da die Optimierungsaufgabe multikriteriell ist, also mehrere Bewertungskriterien gleichzeitig zu optimieren sucht, ist es notwendig, eine Transformation zu finden, die dieses Problem auf ein Problem mit nur einem Ziel zurückführt. Hierfür gibt es verschiedene Vorgehensweisen. Verfahren und Ergebnisse hierzu können in [9] nachgelesen werden.

## 5. Zusammenfassung und Ausblick

Der vorliegende Bericht hat gezeigt, daß es mit Hilfe des Moduls Generationsrate möglich ist, die Ladungsträgergeneration in einer MZIK für viele Anwendungsfälle zu berechnen. Die Gültigkeit des Modells und seiner Erweiterungen wurde durch Messungen überprüft. Des weiteren wurde demonstriert, daß sich das Modul zur Berechnung der Generationsrate auch als eigenständiges Programm zur Simulation stationärer Zellströme (ohne Berücksichtigung der Vorverstärkung) eignet. Im Verbund mit den anderen Modulen Stromberechnung und Vorverstärkung bildet es ein Gesamtsystemmodell, welches sich sowohl zur Gesamtsystemsimulation als auch zur Gesamtsystemoptimierung eignet und in nahezu sämtlichen geometrischen und materialspezifischen Parametern variabel ist.

Die Gültigkeit des Röntgenquellenmodells bedarf noch der Verifikation. Auch wäre eine Erweiterung des Modells auf den Strahlungsmechanismus von Beta-Quellen sinnvoll, wobei hier allerdings der erheblich komplexere Einstrahlvorgang einer Teilchenstrahlung neu zu modellieren ist.

*Die Autoren danken den Herren Dr. Richter und Gärtner von der Firma VacuTec sowie*

*Dr. Huck und Prof. Diener von der FhG IIS/EAS Dresden für die gute Zusammenarbeit während des Projekts.*

## Literatur

- [1] G. Keil, G. Keller, E. Huck, P. Schwarz, H. Bolte, "Optimierung einer Mehrfachionisationskammer zur Kantendetektion", 7. GMM Workshop Methoden und Werkzeuge zum Entwurf von Mikrosystemen, Paderborn '99
- [2] G. Keil, G. Keller, E. Huck, S. Parodat, P. Schwarz, H. Bolte, "Optimierung einer Mehrfachionisationskammer zur Kantendetektion", 8. GMM Workshop Methoden und Werkzeuge zum Entwurf von Mikrosystemen, Berlin '99
- [3] D. Richter, G. Gärtner: „Ein hochauflösendes radiometrisches Meßsystem zur Qualitätskontrolle schnell laufender Materialbahnen“, OMID-Statusseminar 2002, Karlsruhe, diese Tagung
- [4] E. Huck, K.-H. Diener, P. Schneider, P. Schwarz: „Entwurfsunterstützung für ein radiometrisches Meßsystem – Kammerelektronik und Gesamtsystemmodell“, OMID-Statusseminar 2002, Karlsruhe, diese Tagung
- [5] H. Bolte, R. Laur, E. Huck, K.-H. Diener, P. Schneider, P. Schwarz: „Entwurfsunterstützung für radiometrische Meßsysteme - physikalische Modellierung, Simulation und Optimierung“, OMID-Statusseminar 2001, Bremen, S. 119-127
- [6] R.-H. Menk: „Eine Vielkanalionskammer für Röntgenbildaufnahmen“, Verlag Köster, 1994
- [7] J. J. Hubbell: "Photon Cross Sections, Attenuation Coefficients and Energy Absorption Coefficients", NSRDS-NSB 29, Washington '69
- [8] K. Fiedler: „Entwicklung und Implementierung der Flächenmassenmessung in den Mehrfachionisationskammersimulator MFIK-Sim“, Diplomarbeit, Universität Bremen 2002

- [9] H. Bolte, D. Peters, O. Nüssen, R. Laur: „Behavioral Modeling and Multicriteria Optimization of a Radiometric Sensor“, DTIP 2002, S. 267-275, Cannes, 2002

*Das diesem Beitrag zugrunde liegende Vorhaben „OMID-Optimierung von Mikrosystemen für Diagnose- und Überwachungseinrichtungen“ wird mit Mitteln des Bundesministeriums für Bildung, Wissenschaft, Forschung und Technologie (BMBF) unter dem Kennzeichen 16 SV 988/2 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt allein bei den Autoren.*

# ENTWURFSUNTERSTÜTZUNG FÜR EIN RADIOMETRISCHES MEßSYSTEM – KAMMERELEKTRONIK UND GESAMTSYSTEMMODELL

K.-H. Diener, E. Huck, P. Schneider, P. Schwarz, FhG IIS/ EAS Dresden, Email: huck@eas.iis.fhg.de

## Kurzfassung

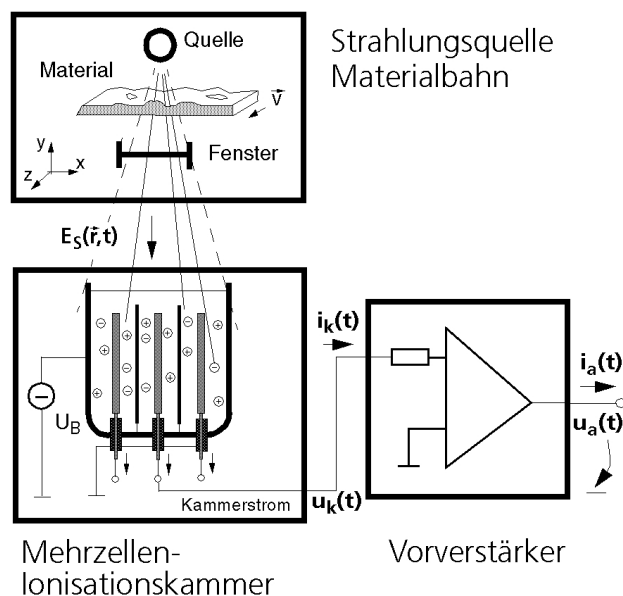
Bei der industriellen Fertigung schnell laufender Materialbahnen werden radiometrische Messverfahren mit Mehrzellen-Ionisationskammern (MZIK) zur Messung der Materialdicke und zur Detektion kurzzeitiger Flächenmasse-Änderungen (Materialfehler) eingesetzt. Dabei ist es erforderlich, jedes Meßsystem hinsichtlich Meßgeschwindigkeit und Meßgenauigkeit an die technologischen Gegebenheiten des Fertigungsprozesses anzupassen. Aus Kostengründen werden standardisierte Baugruppen verwendet und andere Teilsysteme so ausgelegt, daß die Meßanforderungen insgesamt erfüllt werden können.

Um den hohen Entwurfsaufwand bei den kundenspezifischen Entwicklungen zu reduzieren, sollen Simulationsmodelle in Verbindung mit Optimierungsverfahren angewendet werden. Die Vorgehensweise wird am Beispiel eines Meßsystems der Firma VacuTec Meßtechnik GmbH demonstriert. Vorgestellt wird die Modellierung der Teilsysteme Ionisationskammer und Meßvorverstärker sowie deren Verkopplung mit einem Programm zur Berechnung der Einstrahlungsverhältnisse zu einem Gesamtsystemmodell. Mit dem Systemmodell sind Simulationen und Optimierungsrechnungen möglich, in deren Verlauf die Komponenten des Meßsystems optimal an die Kundenwünsche angepaßt werden können.

Die Arbeiten wurden gemeinsam mit der Firma VacuTec Meßtechnik GmbH Dresden und dem Institut für Theoretische Elektrotechnik (ITEM) der Universität Bremen im BMBF-Verbundprojekt OMID durchgeführt.

## 1 Einleitung

Die Design-Optimierung eines radiometrischen Meßsystems mit den Hauptkomponenten Strahlungsquelle/Material, Mehrzellen-Ionisationskammer und Meßvorverstärker (Abb. 1) setzt die Modellierung der genannten Teilsysteme und des Gesamtsystems voraus.



**Abb. 1: Komponenten eines radiometrischen Meßsystems**

Das vom Projektpartner ITEM entwickelte Modell zur Berechnung der Einstrahlungsverhältnisse [2] liefert die Generationsrate, die neben anderen Parametern die entscheidende Eingangsgröße für den Auf- und Umbau von Trägerdichteprofilen in der Kammer ist. Im Hinblick auf die Optimierung ist eine ingenieurgerechte Modellierung der Kammerelektronik zwingend. Diese kann mit problemspezifischen Anpassungen aus dem Modellierungskonzept der Halbleiterelektronik gewonnen werden. Die im weiteren verfolgte Anwendung der angepaßten Halbleitergleichungen [3],[4] auf die Modellierung der inneren Elektronik der MZIK ist ein neuartiger Ansatz, der bisher in der Literatur nicht dokumentiert ist. Das entwickelte Funktionsmodell der Kammerelektronik ist als SABER-Modell in der Modellierungssprache MAST implementiert und ermöglicht, daß für einen großen Variationsbereich der Struktur-, Prozeß- und Betriebsparameter wesentliche Design- und Optimierungsparameter der Kammerzelle tendenziell korrekt und in der richtigen Größenordnung ermittelt werden können. Diese neuartige Modellierung der Ionen- und

Elektronentransportprozesse im Kammerinneren gestattet die numerische Berechnung dynamischer Kammerströme, die Eingangssignale für den Meßvorverstärker sind.

Der Meßvorverstärker ist eine wesentliche Komponente der Auswerteelektronik, die zur Verstärkung der sehr niedrigen Signalpegel des Kammerstroms erforderlich ist.

Die gewählten Schnittstellen ermöglichen eine funktionelle Kopplung der drei Modellkomponenten zu einem Gesamtsystemmodell, das für eine Optimierung des Gesamtsystems benötigt wird.

Die Optimierung erfolgt mit dem bei FhG EAS Dresden entwickelten System MOSCITO [20]. Neben einer funktionellen Beschreibung ist ein parametrisierter Zugang zum ausnutzbaren Optimierungspotential ein wesentliches Ziel der Modellierung der Teilsysteme. Die Zielfunktion zur Optimierung des Gesamtsystems ist anwendungsspezifisch und wurde in Abstimmung mit dem Projektpartner VacuTec entwickelt.

Im Beitrag werden Optimierungsergebnisse vorgestellt und Folgerungen für ein verbessertes Design diskutiert.

## 2 Stromberechnung

### 2.1 Physikalisches Grundmodell

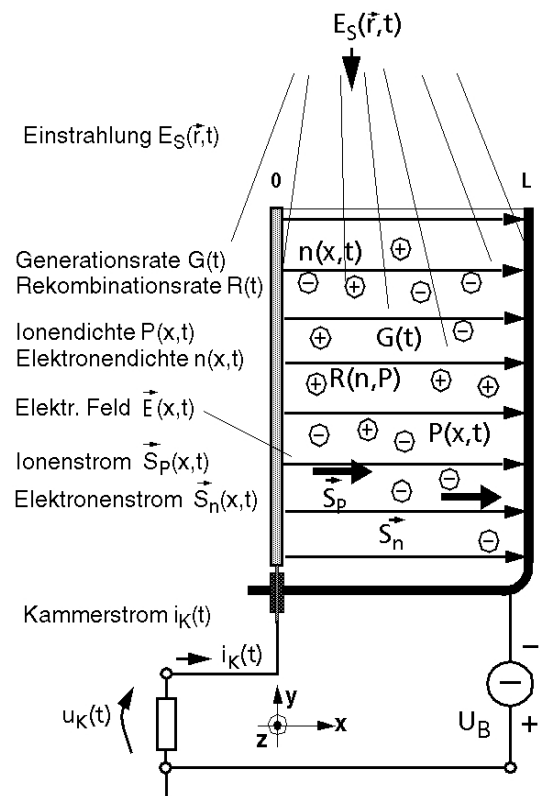
Jede Meßzelle der MZIK ist ein Sensorelement, in dem Ladungsträgerpaare durch ionisierende Strahlung generiert werden. Dieser Prozeß wird summarisch mit einer Generationsrate  $G$  beschrieben. Die erzeugten Ionen (Konzentration  $P$ ) und Elektronen (Dichte  $n$ ) werden durch ein elektrisches Feld (Feldstärke  $\vec{E}$ ) in der Kammer getrennt und bilden dabei den Ionen- und Elektronenstrom mit den Stromdichten  $\vec{S}_p$  bzw.  $\vec{S}_n$ .

Wie in Abb. 1 veranschaulicht, verursachen die zu detektierenden Unregelmäßigkeiten von bewegten Materialbahnen eine zeitabhängige Generationsrate  $G(t)$  in der Kammer und führen folglich zu einem entsprechend modulierten Kammerstrom  $i_K(t)$ .

Die Berechnung von  $i_K(t)$  erfordert neben der phänomenologischen Beschreibung der Generation und Rekombination (Rekombinationsrate  $R$ ) von Ladungsträgerpaaren insbeson-

dere die Modellierung des Auf- und Umbaus von Trägerdichteprofilen  $P(\vec{r},t)$  und  $n(\vec{r},t)$  infolge der Trägergenetik sowie ihre Rückwirkung auf die Feldgestaltung im Kammerinneren. Wegen dieser Wechselwirkungen muß darauf geachtet werden, unzulässige Idealisierungen der funktionsbestimmenden Phänomene zu vermeiden. Deshalb wird beispielsweise bei der Kammerstromberechnung von vornherein ein Zweiträgertransportmodell zugrunde gelegt.

Dabei kann die Berücksichtigung von konstruktionsbedingten zwei- bzw. dreidimensionalen Randfeldern zunächst zurückgestellt werden, wenn diese 2D- bzw. 3D-Effekte ausreichend genau über geeignete Korrekturfaktoren nachträglich in einem 1D-Funktionsmodellansatz berücksichtigt werden können.



**Abb. 2: Veranschaulichung physikalischer Vorgänge in der Kammer**

Die konstruktive Auslegung der voneinander unabhängigen Zellen der MZIK gestattet eine solche 1D-Analyse (Ortskoordinate  $x$ ), weil die Kammerabmessungen in  $y$ - und  $z$ -Richtung vergleichsweise sehr groß sind (Abb. 2). Unter Vernachlässigung der Randeffekte an den Elektroden kann die Generationsrate in  $x$ -



Richtung als konstant angesehen werden.

Das Modellierungskonzept fußt i.w. auf vier Säulen. Die zugehörigen Grundgleichungen werden nachfolgend für eine 1D-Analyse angegeben.

### 1. Poissongleichung

$$\frac{\partial^2}{\partial x^2} \varphi(x, t) = -\frac{e}{\epsilon} \cdot (P(x, t) - n(x, t)) \quad (1)$$

$\varphi(x, t)$ : Potentialverteilung in der Kammer

$e$ : Betrag der Elementarladung

$\epsilon$ : Dielektrizitätskonstante des Füllgases

### 2. Transportgleichungen

Das hier zugrunde gelegte Zweiträgertransportmodell stützt sich auf eine Ionen- und Elektronenteilchenstromdichte

$$S_p = e \cdot (\mu_p \cdot E \cdot P - D_p \cdot \frac{\partial}{\partial x} P) \quad (2)$$

bzw.

$$S_n = e \cdot (\mu_n \cdot E \cdot n + D_n \cdot \frac{\partial}{\partial x} n). \quad (3)$$

$\mu_p, \mu_n$ : Ionen- bzw. Elektronenbeweglichkeit

$D_p, D_n$ : Diffusionskonstanten für Ionen bzw. Elektronen.

Die zur numerischen Auswertung der Teilchenstromkomponenten benötigten Trägerbeweglichkeiten  $\mu_p$  und  $\mu_n$  bzw. Diffusionskoeffizienten  $D_p$  und  $D_n$  sind über die Temperaturspannung  $U_T$  miteinander verknüpft. Es ist

$$\frac{D_p}{\mu_p} = \frac{D_n}{\mu_n} = U_T \approx 26 \text{ mV}, \quad (4)$$

wobei der Zahlenwert für eine Temperatur von 25 °C (Zimmertemperatur) gilt.

Die Trägerbeweglichkeit ist ein phänomenologischer Parameter, der die „Nettobewegung“ der Träger infolge eines elektrischen Driftfelds gegen den „Reibungswiderstand“ des Gases beschreibt [5]. Die Beweglichkeit ist in erster Näherung der mittleren freien Weglänge der Träger direkt proportional, während die Masse der Träger sowie ihre mittlere Geschwindigkeit zwischen den Stößen reziprok mit ihr zusammenhängen. Wenn sich beispielsweise durch Druckerhöhung in der Kammer die mittlere Freiflugstrecke der Träger bis zur ihrer nächsten Kollision mit anderen Trägern verringert, muß sich das in einer Abnahme ihrer Beweglichkeit äußern.

Wenn der Betriebsdruck  $p$  des Füllgases ein Designparameter für Neukonstruktionen und damit ein Optimierungsparameter ist, muß sein Einfluß auf die Trägerbeweglichkeiten und Diffusionskonstanten modelliert werden. Dafür wird die Druckabhängigkeit der Trägerbeweglichkeiten in erster Näherung mit dem Ansatz

$$\mu_p = \mu_p' \cdot \frac{P_0}{p} \quad \text{und} \quad \mu_n = \mu_n' \cdot \frac{P_0}{p} \quad (5)$$

erfaßt ( $\mu_p'$  bzw.  $\mu_n'$  ist die Ionen- bzw. Elektronenbeweglichkeit beim Bezugsdruck  $p_0$ ) [5], [7].

Da die mittlere Geschwindigkeit der stoßenden Teilchen vom Driftfeld  $E$  abhängt, kann die Beweglichkeit bei vergleichsweise hohen Kammerfeldstärken nicht mehr als feldunabhängige Konstante angesehen werden. Dieser Hochfeldbereich wird jedoch im vorgesehenen Betriebsbereich der Ionisationskammer i.a. nicht erreicht und kann bei den weiteren Betrachtungen ausgeklammert werden.

Beim Umbau von Trägerdichteprofilen kommt es zu zeitlichen Änderungen des elektrischen Feldes mit Verschiebungsströmen in der Kammer. Die Gesamtstromdichte ist

$$S = S_p + S_n + \epsilon \cdot \frac{\partial}{\partial t} E. \quad (6)$$

### 3. Kontinuitätsgleichungen

Wegen der strahlungsinduzierten Trägerpaarzeugung und ihrer Vernichtung durch Rekombinationsprozesse können Teilchenströme in der Kammer neu entstehen und auch verenden, wobei nach der Maxwell'schen Theorie die Quellenfreiheit des Gesamtstroms gewahrt bleiben muß. Dieser Sachverhalt wird durch die Kontinuitätsgleichungen

$$\frac{\partial}{\partial x} S_p = e \cdot (G(t) - R(P, n)) - e \cdot \frac{\partial}{\partial t} P \quad (7)$$

und

$$\frac{\partial}{\partial x} S_n = -e \cdot (G(t) - R(P, n)) + e \cdot \frac{\partial}{\partial t} n \quad (8)$$

beschrieben.

### 4. Trägergenetik

Die Modellierung der Trägergenetik umfaßt sowohl die Trägerpaarbildung als auch ihre paarweise Vernichtung durch Rekombination. Dabei kann ein Teil der generierten Trägerpaare durch spontante Rekombination un-

mittelbar nach dem Generationsprozeß vernichtet werden, während andere Träger erst nach Ablauf ihrer mittleren „Lebensdauer“ und Zurücklegung ihrer mittleren „Diffusionslänge“ rekombinieren. Die an Transportprozessen in der Kammer gebundene (nicht spontane) Rekombination findet sowohl im Kammervolumen als auch auf den Kammelektroden statt [5]. Da die Kammerensoren mit speziellen Gasfüllungen unter Überdruck ( $p = 1 \dots 5$  bar) arbeiten [1], ist in der Kammerzelle die Oberflächenrekombination gegenüber der Volumenrekombination zu vernachlässigen [5].

Zur ihrer Berechnung wird ein nichtlineares Volumenrekombinationsmodell

$$R(t) = \alpha \cdot P(t) \cdot n(t) \quad (9)$$

angesetzt. Der Rekombinationskoeffizient  $\alpha$  [9] erfaßt summarisch den Einfluß von Prozeß- und Betriebsparametern (z.B. die Art des Füllgases, den Betriebsdruck  $p$  und die Feldstärkeverteilung in der Kammer) auf die Rekombination.

Bei sonst vergleichbaren Zustandsgrößen des Füllgases (z.B. Temperatur  $\vartheta$ ) nimmt mit steigendem Druck  $p$  seine Dichte zu und gleichzeitig die mittlere freie Weglänge der stoßenden Teilchen ab. Das wirkt sich näherungsweise in einer druckproportionalen Zunahme des Rekombinationskoeffizienten aus, die mit

$$\alpha = \alpha' \cdot \frac{p}{p_0} \quad (10)$$

beschrieben werden kann, wobei  $\alpha'$  und  $p_0$  Anpassparameter sind ( $\alpha' = 10^{-3}$  mm/s,  $p_0 = 2.5$  bar [5], [6], [11]). Darüber hinaus ist  $\alpha$  noch von einigen z.T. auch meßtechnisch schwer erfaßbaren Größen (z.B. Gasverunreinigungen) sensibel abhängig. Diese Vielfalt von Einflußgrößen mit z.T. gegenläufigen Wirkungen auf die Rekombination ist ein Schwachpunkt bei der Quantifizierung des Rekombinationsmodells. Eine genauere Beschreibung ist jedoch sehr aufwendig und aus den in der Literatur verfügbaren Quellen sind z.Z. keine numerisch zuverlässigen Daten zur Qualifizierung eines detaillierteren Modells zu entnehmen [6].

Neben der Trägergeneration infolge ionisierender Strahlung kann es bei hohen Kammerfeldstärken zur Trägermultiplikation durch

Stoßionisation kommen, wenn Impuls und kinetische Energie von im Feld beschleunigten Trägern ausreichen, neutrale Gasmoleküle zu dissoziieren. Dieser Fall wird im weiteren nicht betrachtet, da dieses Phänomen im Arbeitsbereich der MZIK vernachlässigbar bleibt [8], [9].

Das mit den Gln. (1) bis (10) formulierte System nichtlinearer partieller Differentialgleichungen (DGL) repräsentiert i.w. das physikalische Grundmodell der Kammerelektronik.

Zur Lösung des Systems sind noch Anfangswerte und Randbedingungen erforderlich, die physikalisch sinnvoll vorzugeben sind.

Beispielsweise sind zur Lösung der Poissongleichung die Randpotentiale

$$\varphi(x=0, t \geq 0) = 0 \quad (11)$$

und

$$\varphi(x=L, t \geq 0) = -U_B \quad (12)$$

( $U_B$ : Kammeranspannung) vorgeschrieben.

Für die Stromübernahme an den Elektroden wird angenommen, daß an der Anode ( $x=0$ ) nur ein Elektronen- und an der Katode ( $x=L$ ) nur ein Ionenstrom übernommen wird (Abb 2). Somit ist

$$S_p(x=0, t \geq 0) = S_n(x=L, t \geq 0) = 0. \quad (13)$$

Mit

$$P(x, t=0) = n(x, t=0) = 0 \quad (14)$$

sind die Anfangsverteilungen der Trägerdichteprofile in der Kammer vorgeschrieben.

## 2.2 Netzwerkmodell der Kammer

Zur computergestützten Wahl und Optimierung von Designparametern der Kammer werden ein implementiertes Modell und ein Simulator benötigt, mit deren Hilfe die innere Elektronik des Kammerzellensensors in Abhängigkeit der funktionsbestimmenden Struktur-, Prozeß- und Betriebsparameter in ausreichender Näherung berechnet werden kann. Wegen des physikalischen Grundmodell-Ansatzes ist es naheliegend, die Berechnungen mit *FEM-basierten Bauelementesimulatoren* der Halbleiterelektronik [12], [13] durchzuführen. Das ist zwar prinzipiell möglich, erfordert aber z.B. die Neuimplementierung der kammer-spezifischen Randbedingungen (13). Hinzu kommt der Aufwand zur Kopplung des FEM-Simulators mit den Programmen zur

Berechnung der Generationsrate und des Vorverstärkers sowie die Kosten für die Bereitstellung des FEM-Simulators. Deshalb wurde dieser Weg zunächst nicht weiter verfolgt.

Alternativ dazu wurde versucht, das angegebene Grundmodell mittels geeigneter *Netzwerksimulatoren* (NW-Simulatoren) zu analysieren. Diese Simulatoren lösen große nichtlineare Differentialgleichungssysteme (DGL-Systeme), wobei die Modellierung durch Modellbeschreibungssprachen und C-Schnittstellen unterstützt wird. Die DGL-Systeme können problemlos um weitere Netzwerkgleichungen, beispielsweise zur Modellierung des elektrischen Vorverstärkers, ergänzt werden.

NW-Simulatoren lösen Modellgleichungen, die durch *gewöhnliche* DGL beschrieben werden. Es ist deshalb erforderlich, das System *partieller* DGL des physikalischen Grundmodells durch ein System gewöhnlicher DGL zu approximieren. Das bedeutet hier, die Differentialquotienten der örtlichen Ableitungen nach Diskretisierung des Transportraumes bezüglich der  $x$ -Richtung in Differenzenquotienten umzuwandeln. Die Diskretisierung kann z.B. so erfolgen, daß der Transportraum in  $m$  gleiche Abschnitte der Länge  $\Delta x = L/m$  unterteilt wird. In jedem dieser Kammerabschnitte sind alle Grundgleichungen zu approximieren. Es entstehen  $m$  gewöhnliche DGL-Systeme mit entsprechenden Schnittstellengrößen und Randbedingungen.

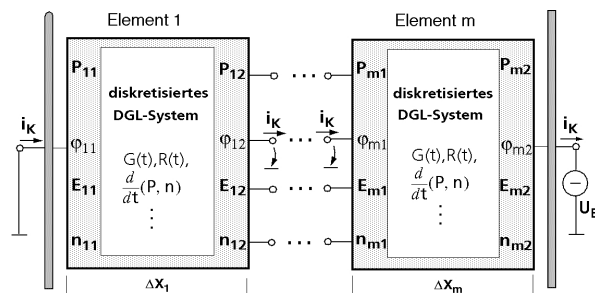


Abb. 3: Netzwerkmodell der Kammer

Mit Blick auf die numerische Lösung kann die Struktur des entstandenen Systems gewöhnlicher DGL auch als NW-Struktur im Sinne der NW-Theorie [15], [16] interpretiert werden. Dazu müssen die abschnittsweise vorliegenden gewöhnlichen DGL gekapselt und für die Schnittstellengrößen an den

Rändern *Klemmengrößen* (Fluß- und Differenzgrößen, [15]) eingeführt werden. Die sich ergebende NW-Struktur, bestehend aus  $m$  NW-Elementen und den Koppelgrößen, zeigt Abb. 3.

Zur Einbindung in den Design- und Optimierungs-Flow für das Gesamtsystem wird ein in weiten Grenzen numerisch robustes und flexibel erweiterbares Funktionsmodell benötigt. Diesen Grundforderungen wird ein Netzansatz mit mindestens zwei Elementen und linearisierten Trägerdichteprofilen gerecht. Abb. 4 zeigt das für die Approximation verwendete Netzwerk-Volumenelement mit einigen geometrischen Abmessungen und Feldgrößen.

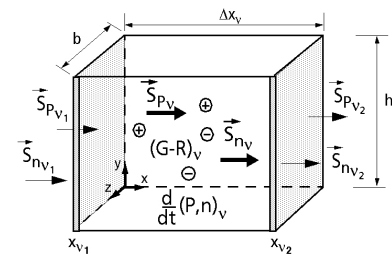


Abb. 4: v. diskretes Volumenelement

Mit der Beschreibung der Trägerdichteprofile in jedem Volumenelement lässt sich die numerisch kritische zweite örtliche Ableitung in der Poissongleichung auf die weniger kritische Berechnung des Feldverlaufes  $E(x)$  im Volumenelement zurückführen.

Wenn im  $v$ . diskreten Kammerelement linearisierte Trägerdichteprofile angenommen werden, liefert die Lösung der Poissongleichung (1) die Feldstärke am linken und rechten Rand

$$E_{v_1} = -\frac{\varphi_{v_2} - \varphi_{v_1}}{\Delta x_v} - \frac{e}{\epsilon} \cdot \frac{\Delta x_v}{6} \cdot (P_{v_2} - n_{v_2} + 2 \cdot (P_{v_1} - n_{v_1})) \quad (15)$$

bzw.

$$E_{v_2} = -\frac{\varphi_{v_2} - \varphi_{v_1}}{\Delta x_v} + \frac{e}{\epsilon} \cdot \frac{\Delta x_v}{6} \cdot (2 \cdot (P_{v_2} - n_{v_2}) + P_{v_1} - n_{v_1}) \quad (16)$$

Darin sind  $P_{v_1}$ ,  $P_{v_2}$  bzw.  $n_{v_1}$  und  $n_{v_2}$  die Konzentrationen der Ionen bzw. Elektronen am linken und rechten Rand des Elements.  $\varphi_{v_1}$  und  $\varphi_{v_2}$  sind Randpotentiale, die - wie auch die o.g. Randkonzentrationen (mit Aus-

nahme der fest vorgeschriebenen Randwerte (s. Gln. (11) und (12)) - Variable sind, die bei der Lösung der Systemgleichungen festgelegt werden.

Die Dynamik der Kammerelektronik wird i.w. von den Kontinuitätsgleichungen (7) und (8) beschrieben. Unter Beachtung der Vereinfachungen des entwickelten Kammermodells findet man für die Differenz der Teilchenstromdichten an den Rändern des betrachteten Elements

$$S_{P_{v_2}} - S_{P_{v_1}} = e \cdot \Delta x_v \cdot \left( G - \alpha \cdot P_{z_v} \cdot n_{z_v} - \frac{d}{dt} P_{z_v} \right) \quad (17)$$

und

$$S_{n_{v_2}} - S_{n_{v_1}} = -e \cdot \Delta x_v \cdot \left( G - \alpha \cdot P_{z_v} \cdot n_{z_v} - \frac{d}{dt} n_{z_v} \right) \quad (18)$$

wobei

$$P_{z_v} = \frac{P_{v_2} + P_{v_1}}{2}, \quad n_{z_v} = \frac{n_{v_2} + n_{v_1}}{2} \quad (19)$$

die zentralen Trägerkonzentrationen im Volumenelement sind, die näherungsweise auch für die Modellierung des Rekombinationsbeitrags in den beiden Kontinuitätsgleichungen verwendet werden.

Die Gleichungen (15) bis (19) ergeben mit den diskretisierten Transportgleichungen (2) und (3) sowie der Gesamtstromdichte (6) das DGL-System des Kammermodells, dessen Lösung mit den Rand- und Anfangsbedingungen (11) bis (14) die innere Elektronik des diskreten Kammerelements beschreibt. Das funktionsgerechte Zusammenwirken der Netzelemente wird durch Stetigkeitsforderungen an die Trägerdichten, das elektrische Potential und das elektrische Feld an den Schnittstellen zwischen benachbarten Elementen, also die Erfüllung der verallgemeinerten Kirchhoffschen Sätze im NW-Modell, gewährleistet.

Das beschriebene Netzwerkmodell wurde in der Beschreibungssprache MAST codiert und liegt für den Simulator Saber [17] als Simulationsmodell vor.

### 2.3 Simulationsergebnisse

Bei der qualitativen Diskussion der Kammerdynamik ist zu beachten, daß sich die Beweglichkeiten und Diffusionskoeffizienten der

beiden beteiligten Trägerarten um Größenordnungen unterscheiden können. Das bedeutet, daß Elektronen und Ionen auf eine zeitliche Modulation der Generationsrate mit stark unterschiedlichen Zeitkonstanten reagieren. Bevor sich die wesentlich langsameren Ionen mit ihrem Teilchenstrom am Umbau der Feld- und Trägerverteilungen beteiligen, können nur die schnellen Elektronen einen Teilchenstrom bilden, der mit dem zugehörigen Verschiebungsstrom im Kammerinneren die Kontinuität des Gesamtstroms sichert. Dieses für Elektronen und Ionen unterschiedliche Reaktionsverhalten läßt sich durch Simulationsrechnungen mit dem implementierten Analysemodell nachweisen. Abb. 5 zeigt z.B. das Übergangsverhalten des Kammerstromes einer 7.5 mm breiten Kammer ([1]) nach einem Sprung der Generationsrate für verschiedene Betriebsspannungen  $U_B$ .

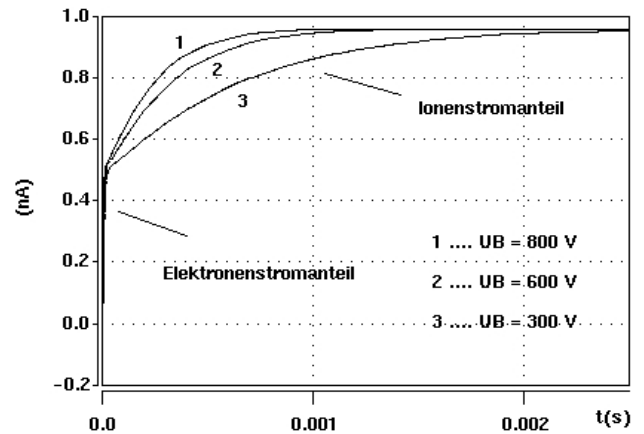
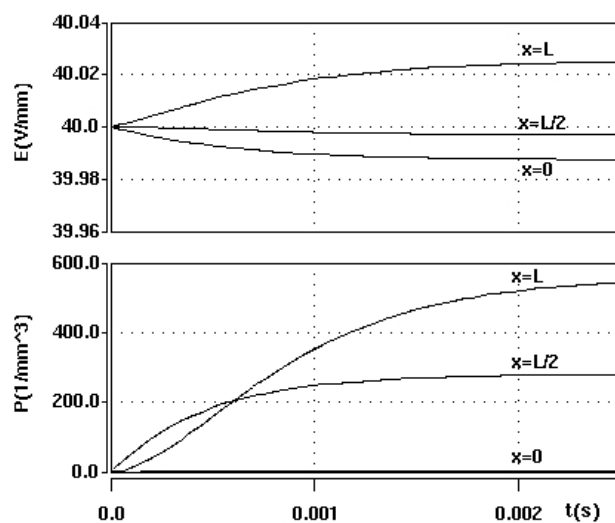


Abb. 5: Übergangsverhalten des Kammerstromes

Zu erkennen ist der sehr schnelle Anstieg des Elektronenstromes auf einen stationären Wert von ca. 0.5 nA und die dann folgende deutlich langsamere Zunahme des Stromes (Ionenstromanteil) bis zum stationären Kammerstromendwert von ca. 0.95 nA. Die Kammerzeitkonstante  $\tau_K$  setzt sich somit aus zwei Anteilen zusammen und ist, besonders im Bereich der Ionenstromkomponente, erheblich spannungsabhängig. Die Spannungsabhängigkeit trifft hingegen auf den Endwert des Kammerstromes bei den hier gewählten Parameterverhältnissen nicht zu. Für  $U_B = 600$  V läßt sich für die im Beispiel berechnete Kammer eine „Gesamtzeitkonstante“ von ca. 200  $\mu$ s (70%-Wert des Stromes) ablesen.

Abb. 6 zeigt die zeitlichen Übergangsvorgänge und die örtlichen Verteilungen des elektrischen Feldes und der Ionendichte für das verwendete 2-Elemente-NW-Modell (Koppelstelle bei  $x=L/2$ , vgl. Abb. 3). Die Ortsverteilung des elektrischen Feldes  $E(x,t)$  weist eine sehr schwache Rückwirkung der Raumladungen auf die Feldgestaltung aus - das Feld ist nahezu homogen. Die Ursache hierfür liegt in dem hier zu betrachtenden Betriebsfall mit vergleichsweise *schwacher Generation* in einer Kammer kleiner Abmessung (7.5 mm) und genügend hoher Betriebsspannung (300...600 V).



**Abb. 6: Elektrisches Feld (oben) und Ionendichte für ein 2-Elemente-Netzwerkmodell**

Die Trägerdichteverteilung  $P(x,t)$  (Abb. 6 unten) folgt nach Erreichen der stationären Endwerte einer fast linearen Verteilung im Kammerraum, ausgehend von einem Trägerdichte-Wert nahe 0 an der Stelle  $x=0$ . Dieses Verhalten ist ebenfalls auf die schwache Generation und die relativ geringe Kammerausdehnung in  $x$ -Richtung zurückzuführen. Gleiches trifft in entsprechender Weise auf die Elektronendichte zu. Betrachtet man beide Trägerarten hinsichtlich der Größenordnung, so ist festzustellen, daß sich indirekt proportional zum Verhältnis der Beweglichkeiten sehr unterschiedliche stationäre Dichteprofile von Ionen- und Elektronenkonzentrationen einstellen, d.h., es liegt eine sehr geringe Elektronendichte und damit eine vernachlässigbare Rekombination im Kammerraum vor. Dieser Sachverhalt erklärt auch, weshalb eine erhöhte Kammerausdehnung keinen höheren stationären

Stromwert bewirken kann.

Die dargestellten Simulationsergebnisse stehen in Übereinstimmung mit Berechnungsergebnissen eines analytischen Ansatzes, der von den Grundgleichungen (1) bis (10) und weiteren Vereinfachungen (rückwirkungsfreies elektrisches Feld, vernachlässigbare Rekombination, sehr geringer Diffusionsstromanteil) ausgeht, aber nicht die diskretisierten Netzwerkgleichungen des numerischen Modells verwendet.

Während der analytische Ansatz für genau die dort gewählten Vereinfachungen Gültigkeit besitzt, ist das implementierte Simulationsmodell auch für die Berechnung anderer Betriebsmodi ausgelegt, z.B. für stärkere Generation und Kammern mit größerer Ausdehnung in  $x$ -Richtung. Numerische Berechnungen solcher Betriebsfälle zeigen, daß sich die Profile beider Trägerarten annähern und in einigen örtlichen Bereichen die in der Literatur beschriebene „Quasineutralität“, d.h. nahezu wertmäßige Gleichheit von Ionen- und Elektronendichte beobachtet werden kann [5]. Außerdem ist in diesen Fällen eine merkliche Rückwirkung der generierten Ladungsträger auf das elektrische Feld und eine daraus resultierende sehr starke Inhomogenität des Feldes festzustellen.

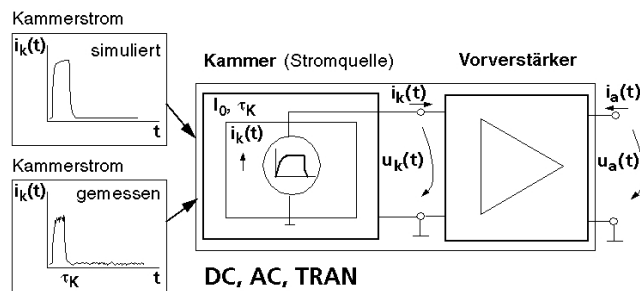
### 3 Stromverstärkung

Die Verstärkung des sehr niedrigen Kammerstromsignals wird durch eine Vorverstärkerschaltung mit Operationsverstärker realisiert. Hierbei handelt es sich um einen I-U-Wandler mit kleinem differentiellen Eingangswiderstand, ein Schaltungstyp, der für die Verstärkung sehr geringer Ströme aus hochohmigen Quellen eingesetzt wird. Problematisch bei diesen Schaltungen ist, daß der gewünschte Pegel der Ausgangsspannung nur mit genügend großen Rückkoppelwiderständen erreichbar ist, woraus sich verschiedene ungünstige Schaltungseigenschaften ergeben können. Dies betrifft vor allem das dynamische Verhalten der Verstärkerschaltung, aber auch Temperaturdrift, zeitliche Stabilität und Rauschen.

Gegenstand der simulationsgestützten Untersuchungen war das dynamische Verhalten der Schaltung. Zum Aufbau des Modells konnte

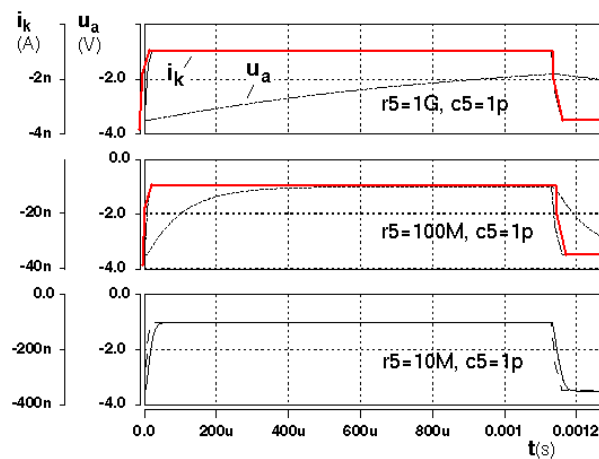
eine Saber-Implementierung des Operationsverstärkers "opa\_128" der Firma Burr-Brown genutzt werden [17], [18].

Für dieses Modell ist mit entsprechender äußerer Beschaltung und Spannungsversorgung ein komplettes Vorverstärker-Modell für den Simulator Saber implementiert worden. Als Testsignale (Stromquellen) für die Simulation wurden aus Messungen und Simulationen gewonnene Stromverläufe genutzt. Ein Schema des Vorgehens zeigt Abb. 7.



**Abb. 7: Modellschema zur Vorverstärkersimulation**

Die Simulationsergebnisse für verschiedene Eingangsstromsignale einer sehr schnellen Kammer sind in Abb. 8 dargestellt.



**Abb. 8: Transientes Verhalten des Vorverstärkers**

Es wird deutlich, daß der Vorverstärker, insbesondere bei kleinen Eingangsstromwerten und den dann erforderlichen hohen Stromverstärkungen, eine dynamisch kritische Komponente darstellt. Optimierungsparameter sind vor allem die Elemente des Rückkopplungszweiges (hier r5, c5).

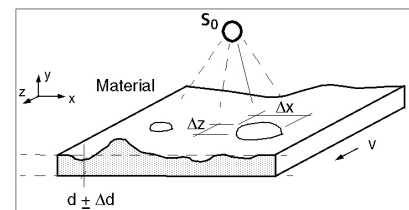
## 4 Optimierung

Zur computergestützten Wahl und Optimierung von wesentlichen Designparametern des Gesamtsystems wird ein durchgehender Design- und Optimierungsflow benötigt. Seine Implementierung muß die Physik der Strahlungsquelle nebst der Trägergenerationsrate in der Kammer, die Kammerelektronik und die angepaßte Dimensionierung des Meßvorverstärkers einschließen.

Dazu werden das oben beschriebene Modell des Gesamtsystems, die mathematische Beschreibung des Optimierungsproblems sowie ein Optimierungssystem benötigt.

### 4.1 Optimierungsproblem

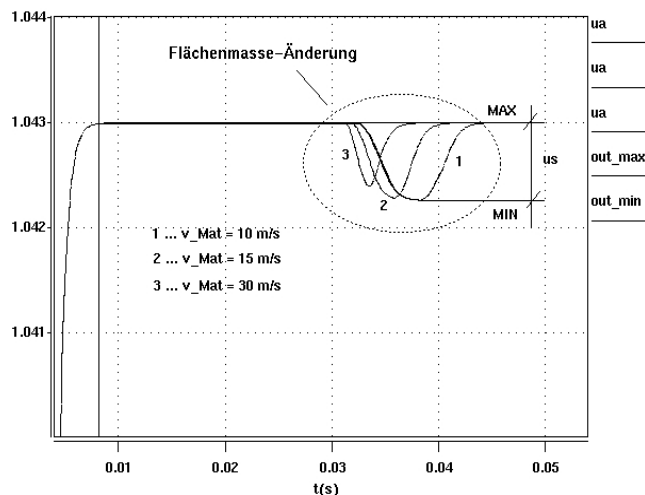
Aufgabe des untersuchten radiometrischen Meßverfahrens ist die Erkennung von Dickenschwankungen und Materialfehlern in x-, z-Richtung (im weiteren „Störungen“ genannt). Die Größenordnung dieser Fehler und die Materialgeschwindigkeiten bei industriellen Herstellungsverfahren sind in Abb. 9 aufgeführt.



$\Delta d_{\min}$	1% von d	(Dickenabweichungen)
$\Delta x, \Delta z$	5...30 mm	(Materialstörungen)
v	0.2...30 m/s	(Bandgeschwindigkeit)

**Abb. 9: Technologische Forderungen**

Die hohe Bandgeschwindigkeit erfordert hohe Abtastraten und eine geringe Zeitkonstante in der Signalverarbeitung des Meßsystems. Durch Simulationen am Gesamtmodell (s. Abschn. 4.2) läßt sich zeigen, daß die technologisch vorgegebene Bandgeschwindigkeit nicht nur die verfügbare Detektionszeit, sondern auch den Pegel des ohnehin sehr schwachen Störsignals beeinflusst. In Abb. 10 ist dieser Sachverhalt beispielhaft für das Durchlaufen einer 0.01 mm starken flächigen Materialverdickung aus Kupfer auf einem 0.5 mm starken Trägermaterial für verschiedene Bandgeschwindigkeiten dargestellt.



**Abb. 10: Verstärker-Ausgangsspannung bei verschiedenen Bandgeschwindigkeiten**

Neben den dynamischen Forderungen ist gleichzeitig eine hohe Verstärkung des von der Kammer gelieferten sehr kleinen Stromsignals erforderlich ( $\Delta i_K(t) \ll 10 \text{ nA}$ ). Geringe Zeitkonstante und hohe Verstärkung sind konkurrierende Auslegungskriterien. Sie werden von allen Teilsystemen beeinflusst und können nur sinnvoll im Zusammenspiel der einzelnen Komponenten beurteilt werden. Damit ist eine Optimierung der gesamten Systemkette erforderlich. Ziel der Optimierung ist, einen ausreichend hohen (maximalen) Störsignalpegel  $u_S$  (s. Abb. 10) auch bei großen Bandgeschwindigkeiten zu erreichen. Damit berechnet sich der Wert  $Z$  der Zielfunktion (ZF) unter Einbeziehung der während der Simulation extrahierbaren Kenngrößen  $MAX$ ,  $MIN$  (s. Abb. 10) zu

$$Z = |MAX - MIN| \Rightarrow \text{Maximum,}$$

mit  $MAX$ ,  $MIN$  ... Ausgangsspannungen am Vorverstärker vor bzw. infolge einer Flächenmasse-Änderung.

Optimierungsparameter, die den größten Einfluß auf den Zielfunktionswert erwarten lassen sind

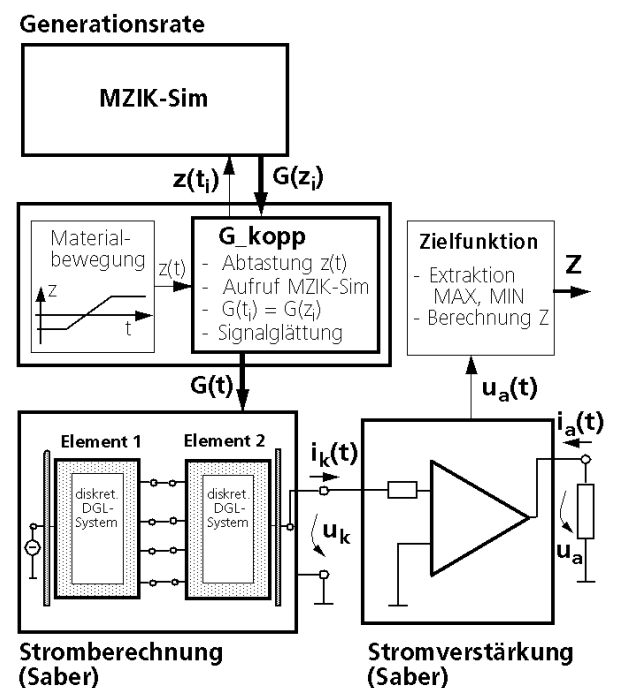
- die Geometrie der Einstrahlungsverhältnisse (Abstände  $a_1$ : Quelle-Bahn und  $a_2$ : Bahn-Kammer),
- der Kammerinnendruck  $p$ ,
- die Kammer-Betriebsspannung  $U_B$  und insbesondere
- die Verstärkerbeschaltung  $r5$ ,  $c5$ .

Die Kammergeometrie hat ebenfalls einen Einfluß auf den Zielfunktionswert. Die Opti-

mierung soll hier jedoch beispielhaft für den Fall einer vorgegebenen Ionisationskammer als Teil eines Kammer-Arrays [1] durchgeführt werden. Optimale Systemparameter bei variierbaren Kammergeometrien lassen sich jedoch sehr leicht in gleicher Weise ermitteln.

## 4.2 Gesamtsystemmodell

Das Gesamtsystemmodell besteht aus den eingangs beschriebenen Modellteilen zur Berechnung der Generationsrate (C-Programm MZIK-Sim [2]), den Saber-Modellen zur Stromberechnung und Stromverstärkung sowie einem Schnittstellen-Modell  $G\_kopp$ , das die berechnete Generationsrate in das Saber-Modell zur Stromberechnung überträgt (Abb. 11).



**Abb. 11: Gesamtsystemmodell mit Schnittstellenmodul  $G\_kopp$**

Der für die Berechnung von  $i_K(t)$  benötigte Zeitverlauf  $G(t)$  der Generationsrate wird aus diskreten Werten  $G(z_i)$  des Programmes MZIK-Sim gewonnen. Dazu wird MZIK-Sim zu jedem Zeitpunkt  $t_i$  einmal abgearbeitet, wobei sich eine Materialstörung mit der Geschwindigkeit  $v = \Delta z / \Delta t$  über der Kammer bewegt. Der Wert  $z = z_0 + v \cdot i \cdot \Delta t$ , der die Lage der Störung über der Kammer angibt (s.a. Abb. 9), ist Eingabewert für MZIK-Sim und wird zu jedem diskreten Simulations-

zeitpunkt  $t_i = t_0 + i \cdot \Delta t$  von den Programmteilen „Materialbewegung“ und „G\_kopp“ des Gesamtmodells generiert. Auf diese Weise kann das ursprünglich für statische Simulation entwickelte Programm MZIK-Sim auch für dynamische Simulationen des Gesamtsystems eingesetzt werden.

Mit dem Gesamtmodell ist somit eine vollständige dynamische Simulation des Meßvorgangs bei bewegtem Materialband einschließlich nachfolgender Signalwandlung durch Ionisationskammer und Vorverstärker möglich.

### 4.3 Optimierungssystem

Für die Gesamtsystemoptimierung wurde das bei Fraunhofer EAS entwickelte Internetbasierte Simulations- und Optimierungssystem MOSCITO verwendet [19],[20],[21]. MOSCITO stellt Basissoftware

- für die Einbindung verschiedener Simulatoren (Spice, Saber, ANSYS, Matlab, C-Funktionen, ...) und Optimierungsprogramme (OPAL, MODOS, GADO, ...) mit einem breiten Spektrum an Optimierungsverfahren (z.B. Nelder-Mead-Simplex, Powell, Quasi-Newton, ...),
- für den einheitlichen Datenaustausch zwischen diesen Programmen und
- für die Ablaufsteuerung und Ergebnisvisualisierung

bereit, so daß es sich sehr gut zur Bearbeitung der hier vorliegenden Optimierungsaufgabe eignet. Abb. 12 zeigt den prinzipiellen Optimierungsablauf in MOSCITO, Abb. 13 die grafische Oberfläche. Über diese Oberfläche startet und konfiguriert der Entwerfer den verwendeten Simulator (hier Saber), den gewählten Optimierer (OPAL, [21]) sowie die Ergebnisdarstellung.

Sehr vorteilhaft ist, daß Parameter- und Zielfunktionswertänderungen über die Visualisierungskomponenten von MOSCITO direkt verfolgt werden können. Damit wird eine sehr praktikable dialoggestützte Arbeit mit dem Optimierungssystem möglich.

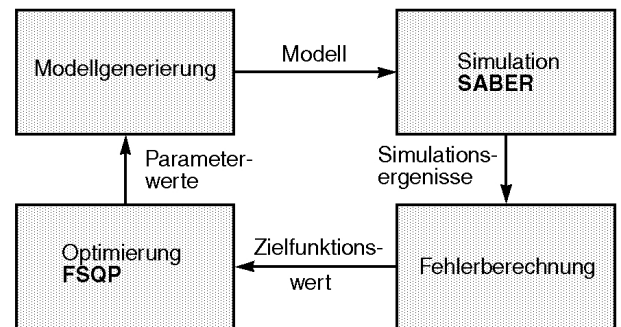


Abb. 12.: Prinzip der simulationsgestützten Optimierung in MOSCITO

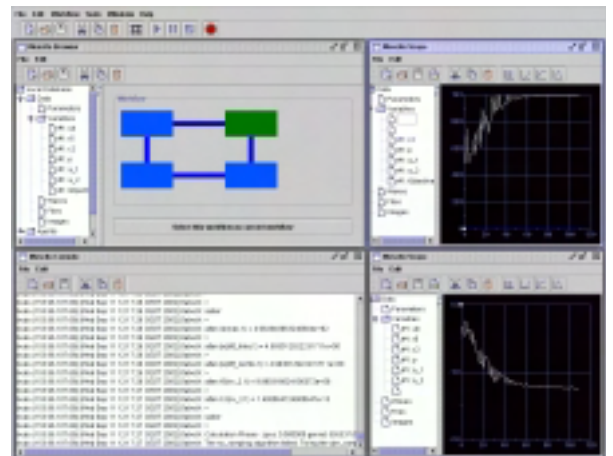


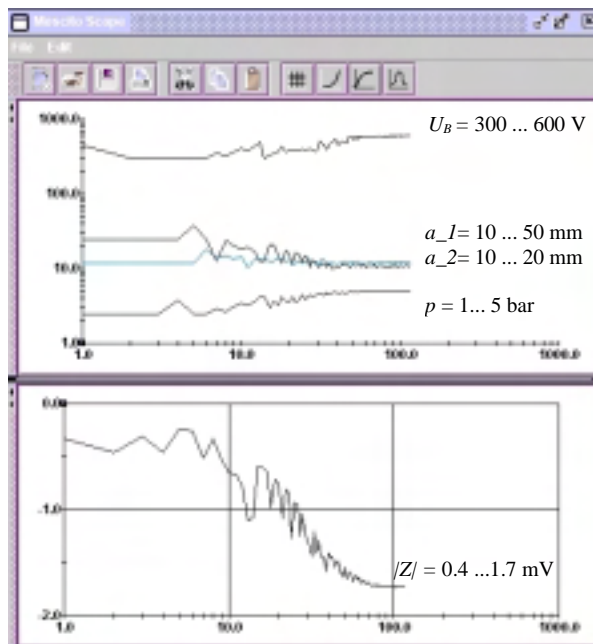
Abb. 13: User -Interface von MOSCITO

### 4.4 Optimierungsergebnisse

Die Optimierungsergebnisse sind in den Abbildungen 14-16 dargestellt. Abb. 14 zeigt den Verlauf von Optimierungsparametern (oben) und Zielfunktionswert für das gewählte Optimierungsverfahren Nelder-Mead-Simplex. Die Berücksichtigung von Beschränkungen in den Variationsbereichen der Parameter erfolgt über eine exakte Straffunktion (Betragfunktion – absolute value penalty function, [20]). Der Zielfunktionswert der optimierten Lösung liegt bei  $Z \sim 1.7$  mV. Dies ist eine deutliche Verbesserung gegenüber der

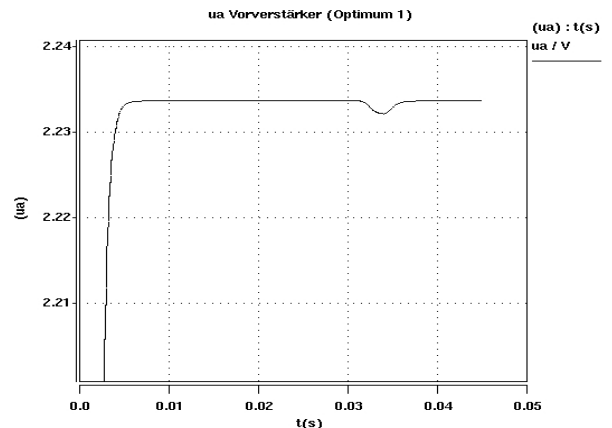


Startlösung mit  $Z < 1$  mV (s. Abb.10,  $Z=U_S$ ). Beim Optimum handelt es sich um ein Randoptimum, erkennbar daran, daß die Extremwerte der Optimierungsparameter an den Grenzen der Variationsräume liegen (Abb. 14). Dies ist a priori nicht voraussagbar, da z.B. der Kammerinnendruck in unterschiedlicher Weise auf das System wirkt (vgl. dazu Abschn. 2, Stromberechnung). Nicht dargestellt sind die Optimierungsparameter des Vorverstärkers, die ebenfalls an den Variationsbereichsgrenzen liegen ( $r5 = 1000$  M $\Omega$ ,  $c5 = 0.1$  pF).



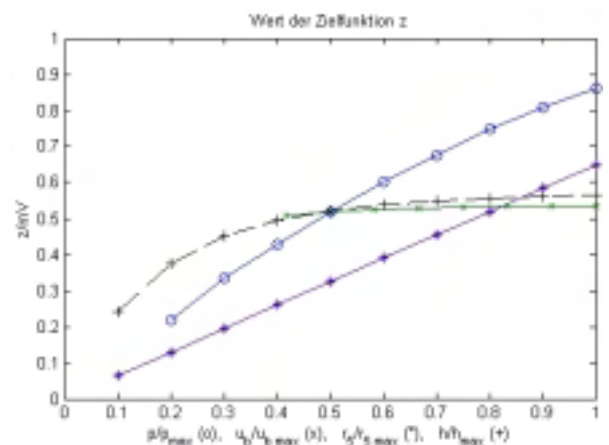
**Abb. 14: Verlauf von Optimierungsparametern und Zielfunktionswert**

Abb. 15 zeigt den Spannungsverlauf am Verstärkerausgang für ein Meßsystem mit optimalen Designparametern, berechnet für eine Materialstörung von 0.01 mm Dicke auf 0.5 mm Kupferträger (vgl. Abschnitt 3.2).



**Abb. 15: Spannung am Verstärkerausgang bei optimalen Designparametern**

Eine weitere für den Entwerfer wichtige Information ist die Kenntnis des Einflusses eines einzelnen Designparameters auf den ZF-Wert. In Abb. 16 sind beispielhaft die normierten Parameter  $U_B$ ,  $p$ ,  $r5$  und hier auch die Kammerhöhe  $h$  als Variationsparameter (bezogen auf ihre zulässigen Maximalwerte) dargestellt. Variiert wird jeweils ein Parameter, für die anderen Parameter werden jeweils die Werte der Startlösung ( $U_B = 300$  V,  $p = 2.5$  bar,  $r5 = 800$  M $\Omega$ ,  $h = 48,5$  mm) genutzt. Die Verläufe sind ebenfalls mit MOSCITO automatisiert generierbar.



**Abb. 16: Einfluß wichtiger Systemparameter auf den Zielfunktionswert**

## 5 Zusammenfassung und Ausblick

Im Beitrag wird das komplexe Problem der Modellierung, Simulation und Optimierung eines radiometrischen Meßsystems behandelt.

Es werden Modelle und Simulationsergebnisse zur Berechnung der Einstrahlungsverhältnisse, des dynamischen Kammerstromes einer Ionisationskammer sowie zur Analyse von dynamischen Eigenschaften eines Meßverstärkers vorgestellt. Die Kopplung der entwickelten Teilsystemmodelle zu einem Gesamtmodell wird beschrieben und eine Zielfunktion zur Optimierung definiert. Mit dem Gesamtsystemmodell ist eine Simulation des Systemverhaltens auch für den Fall der Detektion von Materialinhomogenitäten bei schnell laufenden Materialbahnen möglich.

Zur Optimierung wesentlicher Design-Parameter des Gesamtsystems wurden Optimierungsrechnungen mit dem Optimierungssystem MOSCITO durchgeführt. Die gefundenen Randoptima zeigen Designpotentiale und mögliche Entwicklungsansätze für die Verbesserung zukünftiger Meßsysteme auf.

Die Verfasser danken Dr. D. Richter, G. Gärtner und Dr. W. Burmeister (VacuTec Meßtechnik GmbH Dresden) sowie Herrn H. Bolte (ITEM, Universität Bremen) für die sehr gute Zusammenarbeit im Projekt.

## Literatur

- [1] D. Richter, G. Gärtner, "Ein hochauflösendes radiometrisches Meßsystem zur Qualitätskontrolle schnell laufender Materialbahnen", diese Tagung
- [2] H. Bolte, R. Laur, "Entwurfsunterstützung für ein radiometrisches Meßsystem – Ein physikalisches Modell der Ladungsträgergeneration", diese Tagung
- [3] A.S. Grove, "Physics and Technology of Semiconductor devices", Wiley, New York 1977
- [4] H. Elschner, A. Möschwitzer, A. Reibiger, "Rechnergestützte Analyse in der Elektronik", Hüthig, Heidelberg 1977
- [5] G. Mierdel, "Elektrophysik" Verlag Technik, Berlin 1970
- [6] J.P. Boeuf, L.C. Pitchford, "The SIGLO series, gas parameters and cross section data", Kinema Software, Autumn Way Monument, Colorado 1996
- [7] A. v. Engel, M. Steenbeck, "Elektrische Gasentladungen – Ihre Physik und Technik, Bd.1" Springer Verlag, Berlin 1932
- [8] G. Keil, G. Keller, E. Huck, P. Schwarz, H. Bolte, "Optimierung einer Mehrfachionisationskammer zur Kantendetektion", 7. GMM Workshop Methoden und Werkzeuge zum Entwurf von Mikrosystemen, Paderborn 1999
- [9] G. Keil, G. Keller, E. Huck, S. Parodat, P. Schwarz, H. Bolte, "Optimierung einer Mehrfachionisationskammer zur Kantendetektion", 8. GMM Workshop Methoden und Werkzeuge zum Entwurf von Mikrosystemen, Berlin 1999
- [10] J. J. Hubbell, "Photon Cross Sections, Attenuation Coefficients and Energy Absorption Coefficients", NSRDS-NSB 29, Washington 1969
- [11] K. Kleinknecht, "Detektoren für Teilchenstrahlung", Teubner, Stuttgart 1987
- [12] R. Stenzel, et al., "Simulation of AlGaIn&GaIn-HFETs including spontaneous and piezoelectric polarisation charges", Proc. 26th Int. Symp. On Compound Semiconductors, Berlin 1999
- [13] MEDICI, User's Manual, Avant! Corporation, Fremont, 1999
- [14] O.C. Zienkiewicz, R.L. Taylor, "The Finite Element Method" McGraw-Hill, London/ New York, 1994
- [15] P. Schwarz, "Physically Oriented Modeling of Heterogeneous systems" 3. IMACS Symp. Mathematical Modeling (MATHOD), Wien 2000, pp.
- [16] C. Clauß, J. Haase, P. Schwarz, "An Approach to Analog Behavioural Modelling" Proc. VHDL User Forum Europe, SIG-VHDL Spring'96 Working Conference, Dresden 1996, Shaker, Aachen 1996
- [17] Saber Template Library, Release 5.2, Analogy Inc., Beaverton, 2000
- [18] <http://www.burr-brown.com>
- [19] A. Schneider, P. Schneider, D. Peters, W. Jakob, A. Quinte, "Simulationsgestützte Optimierung von heterogenen Systemen", Tagung Multi-Nature-Systems 2001, Hamburg 2001
- [20] P. Schneider, A. Schneider, J. Bastian, S. Reitz, P. Schwarz, "MOSCITO – a pro-

gramm system for MEMS optimization",  
Proc. DTIP 2002, Cannes, France, 5.-6.  
Mai 2002, 246-257  
<http://www.eas.iis.fhg.de/publications/papers/2002/001>

- [21] A. Schneider, P. Schneider, J. Bastian,  
"MOSCITO – ein Internet-basiertes Pro-  
grammsystem für die Optimierung von  
Mikrosystemen", diese Tagung

*Das diesem Beitrag zugrunde liegende Vorhaben „OMID-Optimierung von Mikrosystemen für Diagnose- und Überwachungseinrichtungen“ wird mit Mitteln des Bundesministeriums für Bildung, Wissenschaft, Forschung und Technologie (BMBF) unter dem Kennzeichen 16 SV 988/2 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt allein bei den Autoren.*

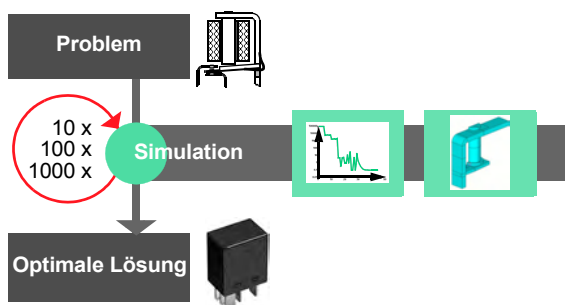
# MOSCITO – Ein modulares, Internet-basiertes Programmsystem für die Optimierung von Mikrosystemen

A. Schneider, P. Schneider, J. Bastian  
Fraunhofer-Institut für Integrierte Schaltungen, Außenstelle EAS Dresden  
Zeunerstraße 38, 01069 Dresden  
E-Mail: Andre.Schneider@eas.iis.fhg.de

**Kurzfassung:** Mit der Optimierung von Mikrosystemen wird das Ziel verfolgt, beginnend bei der Entwurfsphase eine Produktlösung zu schaffen, die hinsichtlich Funktionalität, Materialkosten sowie Fertigungsaufwand optimale Eigenschaften aufweist. Simulation und Optimierung nehmen dabei eine Schlüsselstellung ein. Der Beitrag stellt mit MOSCITO ein Programmsystem vor, das genau diese beiden Entwurfsphasen unterstützt. Der Entwickler erhält mit Hilfe von MOSCITO Zugriff zu Simulationswerkzeugen und Optimierungsverfahren und kann diese über das Internet miteinander verbinden und auf entfernten Rechnern nutzen. Zwei Anwendungsbeispiele illustrieren dabei Funktionsprinzip und Leistungsfähigkeit des Ansatzes.

## 1 Motivation

Im Entwurfsprozess von komplexen Systemen der Elektronik, der Mikrosystemtechnik, im Maschinenbau und in der Automatisierungstechnik sind Eigenschaften wie geringes Gewicht, minimaler Energieverbrauch, hohe Lebensdauer u.v.a.m. wichtige Randbedingungen für den Entwurf und entscheiden letztlich neben den eigentlichen Systemmerkmalen über die Marktakzeptanz eines Produktes.



**Bild 1:** Prinzip der simulationsgestützten Optimierung

Aufgrund der physikalischen Heterogenität gestaltet sich der Entwurf solcher Systeme oft sehr kompliziert. Versuche mit Testmustern sind in der Regel zeit- und kostenintensiv, so dass die Nutzung von Simulationswerkzeugen unter Verwendung eines an die jeweilige Untersuchung bzw. den Untersuchungsgegenstand angepassten Modells (analytische Funktion, Verhaltensbeschreibung, Netzliste o.ä.) oft die einzig sinnvolle Alternative darstellt.



**Bild 2:** Mehrzellenionisationskammer (MZIK)

Folgende Entwurfsphasen spielen dabei eine zentrale Rolle: die **Modellierung** zum Finden eines Modells, welches alle für den jeweiligen Entwurfschritt wesentlichen Eigenschaften des realen Systems nachbildet, die **Simulation** zum Prüfen der prinzipiellen Funktionsweise eines Mikrosystems und die **Optimierung** für die systematische Suche nach einer möglichst optimalen Realisierung. Verbindet man nun Simulation und Optimierung zu einem Iterationszyklus, ergeben sich interessante und vielfältige Möglichkeiten für den rechnergestützten Entwurf. Welche Probleme bei der Realisierung eines solchen kombinierten Simulations-Optimierungs-Ansatzes jedoch in der Praxis auftreten, sollen die folgenden beiden Entwurfsszenarien illustrieren.

Kernstück radiometrischer Messsysteme zur Bestimmung von Materialdicken, Flächenmassen und Materialkanten im Fertigungsprozess von Bandmaterialien (Papier, Walzstahl, Folien etc.) sind **Mehrzellenionisationskammern**. Diese de-

tektieren die Abschwächung radioaktiver Strahlung ( $\gamma$ , Röntgen), die durch die vollständige oder teilweise Abdeckung der Einzelkammern durch das zu vermessende Material hervorgerufen wird. Zur Verbesserung der Messsysteme sollten unter anderem die geometrischen Abmessungen der Kammer sowie einige Parameter der Auswerteschaltung optimiert werden. Das Ziel bestand dabei darin, kleinste Abweichungen der Materialkante von der Sollvorgabe bzw. kleine Materialstörungen oder Dicken-schwankungen zuverlässig und schnell zu detektieren. Kammervolumen und damit Materialeinsatz und Herstellungskosten mussten minimiert werden.

Die Erfassung aller für die Entwurfsaufgabe relevanten physikalischen Effekte an der Ionisationskammer gelang mit Hilfe von Verhaltensmodellen. Zur Unterstützung der Modellierung wurden FEM-Simulationen der Felder in der Kammer (ANSYS) und die Analyse spezieller Effekte aus der Halbleiterelektronik [18] durchgeführt. Für die Simulation des Messsystems wurden SABER (Kammerverhalten und Elektronik) sowie ein spezielles C-Programm (Strahlungscharakteristik) verwendet. Da der Einfluss der Geometrieparameter auf das zu optimierende Volumen und das Ausgangssignal kaum abschätzbar war, waren Experimente mit unterschiedlichen Optimierungsalgorithmen notwendig, um zu einem vernünftigen Ergebnis zu gelangen. Der Entwerfer sollte diese Versuche nach Möglichkeit vor Ort am eigenen PC durchführen können.



**Bild 3:** Kraftsensor

In einem zweiten Szenario sollte ein **Kraftsensor** optimiert werden. Das Ziel bestand darin, über Dehnmessstreifen die zu detektierenden Kräfte möglichst präzise und mit hohem Signalpegel zu erfassen und dabei gleichzeitig die Empfindlichkeit gegenüber Fertigungstoleranzen für die Position der Messstreifen zu vermindern. Mit Hilfe von z.T. sehr rechenintensiven ANSYS-Simulationen des Verformungskörpers wurde die Sensorgeometrie schritt-

weise optimiert. Auch in diesem Szenario sollte der Entwerfer an seinem PC arbeiten; für die ressourcenintensiven Simulationen war allerdings der Zugriff auf leistungsfähige Rechentechnik erforderlich.

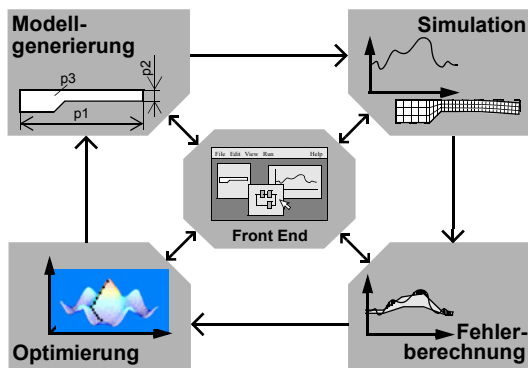
Zusammenfassend lassen sich für die genannten Szenarien folgende Probleme nennen:

- Die Heterogenität der Entwurfsaufgaben erfordert eine große Vielfalt von Modellierungsansätzen. In der Praxis ist man damit auf unterschiedliche Simulationswerkzeuge angewiesen.
- In Bezug auf die Optimierung besitzt der Entwerfer vorab kaum Informationen darüber, welches Verfahren für ein spezielles Problem geeignet und effizient ist. Hilfreich ist in diesem Fall ein breites Spektrum an verfügbaren Algorithmen für unterschiedliche Problemklassen.
- Die benötigten Werkzeuge (z.B. Simulatoren) sind vor Ort (am eigenen Rechner auf dem Schreibtisch) nicht immer verfügbar. Gründe dafür können fehlende Lizenzen, mangelnde Rechenleistung oder die Inkompatibilität des installierten Betriebssystems (UNIX oder Windows) sein. Ausweg bietet hier die Nutzung entfernter Ressourcen über LAN oder Internet.
- Der Entwerfer kann bei der Vielfalt der zum Einsatz gelangenden Werkzeuge nicht in jedem Fall über das notwendige Wissen zur Bedienung verfügen. In der Regel genügt es, unter Verzicht auf interaktives Arbeiten mit grafischen Oberflächen ein Werkzeug lediglich für die Erbringung eines Dienstes (z.B. für eine Simulation) zu nutzen.

In den folgenden Abschnitten greifen die Autoren die genannten Probleme auf und stellen ein Konzept für ein Simulations-Optimierungs-System vor. Mit MOSCITO wird anschließend eine realisierte Softwarearchitektur vorgestellt. Der Schwerpunkt liegt dabei auf der Darstellung der in das System integrierten Simulationswerkzeuge und Optimierungsalgorithmen. Abschließend werden Aspekte der Internet-basierten Arbeitsweise erläutert und Anwendungsbeispiele vorgestellt.

## 2 MOSCITO

Das unter dem Namen MOSCITO konzipierte und realisierte Simulations-Optimierungs-System gliedert sich in die vier funktionalen Module Modellgenerierung, Simulation, Fehlerberechnung und Optimierung. Hinter den Modulen verbergen sich eigenständige Bearbeitungsphasen, die nacheinander zyklisch – im Sinne einer Iteration – abgearbei-

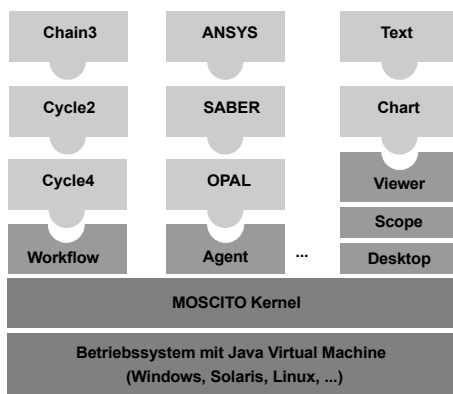


**Bild 4:** Modulares Konzept für ein Simulations-Optimierungs-System

tet werden. Bild 4 veranschaulicht das Gesamtsystem. Die Funktionsweise wird ausführlich in [16] beschrieben.

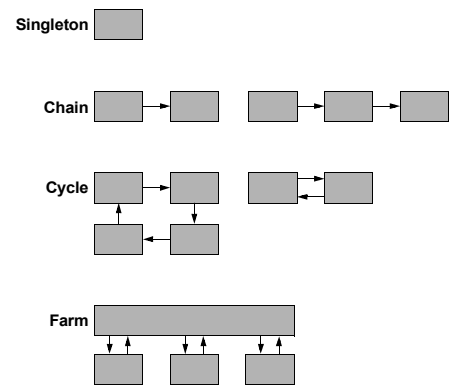
Entscheidend für die Flexibilität des Systems sind folgende Aspekte:

- Jedes Modul (im Bild 4 die äußeren vier Blöcke) entspricht in der Praxis einem eigenständig arbeitenden MOSCITO-Agenten, in den ein konkretes Werkzeug (z.B. der Schaltungssimulator SABER oder das Optimierungsprogramm OPAL) eingebettet ist.
- Die Daten- und Steuerverbindungen zwischen den Modulen bzw. MOSCITO-Agenten (im Bild 4 mit schwarzen Pfeilen dargestellt) werden über TCP/IP-Sockets realisiert. Damit können die einzelnen Agenten verteilt auf unterschiedlichen Rechnern laufen und über ein lokales Netzwerk (LAN) oder über Internet gekoppelt werden.
- Die gesamte MOSCITO-Architektur ist komplett in Java implementiert und damit auf allen Plattformen, für die eine Java Virtual Machine verfügbar ist (u.a. Windows, Solaris, Linux), einsetzbar.

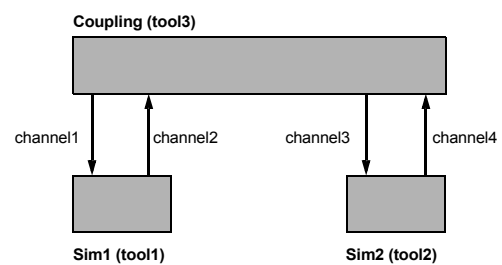


**Bild 5:** MOSCITO-Softwarearchitektur mit offenen Schnittstellen

MOSCITO ist als offenes System konzipiert. Anwender können somit die Funktionalität von MOSCITO über Schnittstellen, die jeweils über Java-Interfaces bzw. über abstrakte Java-Klassen definiert sind, erweitern. Auf diese Weise lassen sich u.a. neue Werkzeuge, Datenkonverter bzw. Viewer für die Ergebnisanzeige in MOSCITO zusätzlich integrieren. Weiterhin gibt es eine offene Schnittstelle für das Einbringen neuer Workflows. Während das ursprüngliche Konzept zunächst nur den in Bild 4 dargestellten Zyklus mit vier Modulen vorsah, kann der Anwender nun aus einer Vielzahl von verfügbaren Workflows auswählen (vgl. Bild 6) oder aber selbst neue Workflows definieren. Damit eröffnen sich neben der simulationsgestützten Optimierung auch Anwendungsbereiche unter anderem bei der Simulatorkopplung.



**Bild 6:** Auswahl an verfügbaren MOSCITO-Workflows



**Bild 7:** Workflow für eine Simulatorkopplung nach dem Farmer-Worker-Prinzip

Die Definition eines anwenderspezifischen Workflows erfolgt über eine Java-Klasse. Der dem Workflow zugrundeliegende bipartite Graph wird ähnlich einer Netzliste beschrieben.

Das folgende Beispiel illustriert die Beschreibung für einen bei Simulatorkopplungen üblichen Farm-Workflow mit einem Farmer (Simulatorkopplung) und zwei Workern (Simulatoren).

```

channel1 = new MoscitoChannel("Sim1Input");
channel2 = new MoscitoChannel("Sim1Output");
channel3 = new MoscitoChannel("Sim2Input");
channel4 = new MoscitoChannel("Sim2Output");

tool1 = new MoscitoTool("Sim1", "Sim1", 1, 1);
tool2 = new MoscitoTool("Sim2", "Sim2", 1, 1);
tool3 = new MoscitoTool("Coupling", "Cp1", 1, 1);

tool1.setInputChannel(0, channel1);
tool1.setOutputChannel(0, channel2);

tool2.setInputChannel(0, channel3);
tool2.setOutputChannel(0, channel4);

tool3.setOutputChannel(0, channel1);
tool3.setInputChannel(0, channel2);
tool3.setOutputChannel(1, channel3);
tool3.setInputChannel(1, channel4);

```

Interessanter für den Anwender ist in der Regel die Einbindung neuer Werkzeuge in das Simulations-Optimierungs-System. Das Vorgehen soll im Folgenden skizziert werden.

Voraussetzung dafür, dass ein Werkzeug in MOSCITO integriert werden kann, ist, dass es als Hintergrundprozess im Batchmodus arbeiten kann. Für einen Simulator bedeutet das, dass eine Simulation ohne Nutzer-Interaktionen und ohne die Anzeige einer grafischen Oberfläche (GUI) durchführbar sein muss. Dafür muss die gesamte Simulation über Kommandozeilenoptionen bzw. Eingabe- oder Konfigurationsdateien steuerbar sein. Sind diese Bedingungen erfüllt, werden für die Integration genau zwei Dateien erstellt:

- eine Agenten-Beschreibungsdatei, die alle Steuerparameter enthält, mit denen das Verhalten eines Werkzeugs von außen beeinflusst werden kann;
- eine Java-Quelldatei, die die Schnittstelle zwischen MOSCITO und dem eingebetteten Programm realisiert.

Für beide Dateien gibt es Templates und hinreichend viele Beispielrealisierungen, so dass eine Einbindung eines neuen Werkzeugs in MOSCITO in der Regel nur wenige Stunden erfordert.

Die Agenten-Beschreibungsdatei ist eine XML-Datei, die in Listenform alle Konfigurationsparameter mit Namen, Typ, Initialwert sowie Kurzbeschreibung enthält. Der MOSCITO-Desktop nutzt diese Beschreibung für die automatische Generierung werkzeugspezifischer Dialogfenster (Bild 8).

```

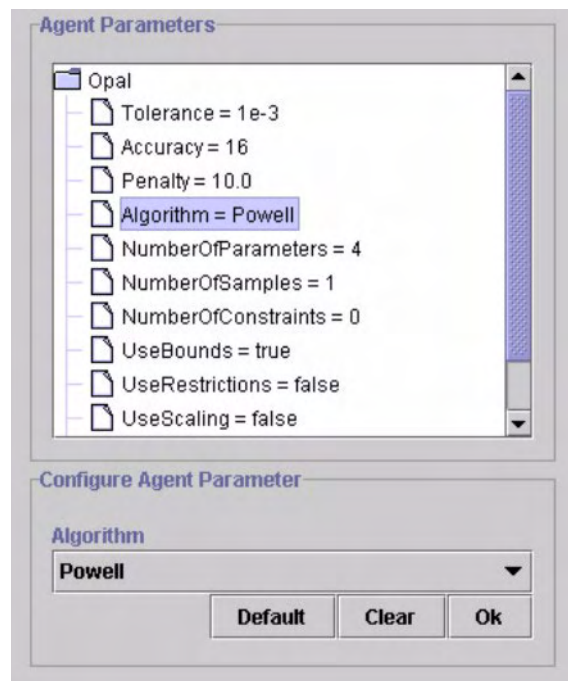
<?xml version="1.0" standalone="no"?>
<!DOCTYPE MoscitoML SYSTEM
  "http://www.eas.iis.fhg.de/moscito/lib/
  MoscitoML.dtd">
<MoscitoML version="1.0">
<Agent
  name="Opal"
  service="Optimization"
  call="java moscitox.agents.Opal"
  description="The agent provides ...">

```

```

<AgentConfiguration>
  <AgentParameter
    name="Tolerance"
    type="Double"
    default="1e-3"
    description="..." />
  <AgentParameter
    name="Accuracy"
    type="Integer"
    default="16"
    description="..." />
  <AgentParameter
    name="Restrictions"
    type="File"
    default="noname.res"
    description="..." />
  ...
</AgentConfiguration>
</Agent>
</MoscitoML>

```



**Bild 8:** Eingebundene Werkzeuge lassen sich über Dialogfenster konfigurieren.

Für die Realisierung des Agenten liegt die vorbereitete Schnittstelle analog zu den oben beschriebenen Workflows als Java-Klasse bereit. Hier müssen lediglich noch die Methoden *configure()*, *initialize()*, *service()*, *stop()* und *terminate()* implementiert werden.

```

// Agenten-Klasse zur Einbettung des
// FEM-Simulators Ansys

public class Ansys
  extends MoscitoAgent
{
  // Deklaration der werkzeugspezifischen
  // Variablen

  public Ansys(String args[])
  {
    // Konstruktor
  }
}

```

```

public void configure()
{
    // Konfiguration des eingebetteten Programms
    // (z.B. Aufbau der Kommandozeile für den
    // Programmstart)
}

public void initialize()
{
    // Initialisierung
    // (z.B. Bereitstellen von Eingabedateien)
}

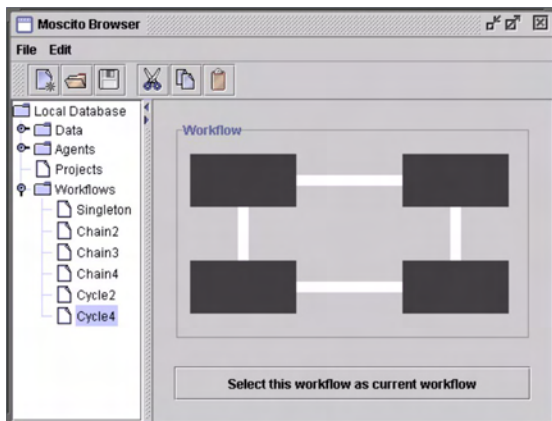
public String service(String input_data)
{
    // Kernfunktion: Bereitstellung des Dienstes
    // z.B. Aufruf des Simulators
    runtime.exec("ansys -p ANE3 -l C -j ...");
    // ...
    return(output_data);
}

public void stop()
{
    // Beenden bzw. Anhalten des Programms
}

public void terminate()
{
    // Freigabe der belegten Ressourcen
}
}

```

Die bisher gewonnenen Erfahrungen zeigen, dass sich nach diesem Schema viele verfügbare Werkzeuge einbetten lassen. Die Integration der für die Optimierung von Mikrosystemen wichtigsten Werkzeuge soll in den folgenden beiden Abschnitten ausführlich beschrieben werden.



**Bild 9:** Auswahl eines Workflows mit Hilfe des Browsers

### 3 Simulationswerkzeuge

Viele physikalische Effekte in heterogenen Mikrosystemen lassen sich in der Regel nur mit Hilfe von partiellen Differentialgleichungen (PDE) auf Komponentenebene bzw. durch Algebra-Differenti-

algleichungen (DAE) auf Teil- und Gesamtsystemebene hinreichend exakt modellieren. Für die Simulation sind folglich leistungsfähige, spezialisierte Programme erforderlich, zu denen u.a. folgende frei bzw. kommerziell verfügbare Produkte zählen:

- **SABER:** Netzlisten- und Verhaltenssimulation für analoge elektrische Schaltungen, auch mit digitalen Teilen (mixed-signal); über die Verhaltensbeschreibungssprache MAST ist die Einbeziehung nichtelektrischer Systeme möglich (vgl. [2]);
- **Spice:** Simulation elektrischer Netzwerke (frei verfügbarer Public-Domain-Simulator); gilt als De-facto-Industriestandard im Bereich der Halbleitersimulation;
- **ANSYS:** FEM-Simulation; insbesondere geeignet für genaue Simulationen mechanischer Komponenten, elektromagnetischer Felder und gekoppelter physikalischer Effekte in Mikrosystemen;
- **Matlab:** weit verbreitetes Programm, geeignet zur Simulation allgemein kontinuierlicher Systeme (insbesondere Regelungs- und Steuerungssysteme);
- **Dymola:** Simulator, der auf Basis der Modellierungssprache Modelica [9] die Simulation physikalischer und damit insbesondere heterogener Systeme gestattet;
- **ITI-SIM:** Simulator für mechatronische Systeme mit umfangreichen Bibliotheken für regelungstechnische Blöcke, mechanische, hydraulische und elektrische Komponenten.

Obwohl die einzelnen Simulationswerkzeuge auf jeweils unterschiedlichen theoretischen Ansätzen beruhen, können sie in MOSCITO nach einem einheitlichen Schema integriert werden. Jeder Simulator erhält am Eingang ein simulationsfähiges Modell in Form einer oder mehrerer Dateien (Netzliste, Verhaltensbeschreibung, ...), und am Ausgang werden nach durchgeführter Simulation die Ergebnisse in werkzeugspezifischer Syntax ausgegeben. Da sich die Simulatoren in der Regel über Aufrufoptionen bzw. Konfigurationsdateien oder -kommandos steuern lassen, kann die Simulation im Rahmen von MOSCITO als Dienst (im Hintergrund) erbracht werden. Das ist eine entscheidende Voraussetzung für die Realisierung der zyklischen, iterativen Arbeitsweise des Simulations-Optimierungssystems. Die derzeit integrierten Simulatoren gestatten, ein breites Spektrum an Problemstellungen beim Entwurf von Mikrosystemen mit MOSCITO zu bearbeiten.



## 4 Optimierungsverfahren

Die Optimierungsverfahren sind im Gegensatz zu den Simulatoren nicht einzeln in MOSCITO integriert, sondern stehen als Programm mit dem Namen OPAL (OPTimierungsALgorithmen) zusammengefasst zur Verfügung. Für OPAL wurden Verfahren aus verschiedenen Klassen ausgewählt, welche aus bewährten Software-Bibliotheken stammen. Dadurch kann ein breites Spektrum an Problemen bearbeitet werden; der Nutzer sollte stets ein gut angepasstes Verfahren vorfinden. Es wurde für die Algorithmen ein einheitliches Interface erstellt, so dass sie steuerbar sind und mit den anderen Agenten kommunizieren können. Wie in Bild 4 illustriert, wird als Ausgabe der aktuelle Parametersatz an den Generator übergeben, welcher sie für den Simulator aufbereitet. Die Eingabe besteht aus dem vom Evaluator berechneten Zielfunktionswert. Über den MOSCITO-Desktop (Front-End-Programm) wird der Optimierer konfiguriert und rekonfiguriert, der Optimierungsprozess kann sowohl unter- als auch abgebrochen werden, er kann im Schrittbetrieb laufen und über die Konsole beobachtet werden. Auf der Konsole werden Fehlermeldungen und weitere verfahrensspezifische Informationen ausgegeben.

Bei praktischen Optimierungsproblemen kommt es vor allen Dingen darauf an, schnell und mit möglichst geringem Aufwand eine Näherung für die Optimallösung zu bestimmen oder zumindest eine deutliche Verbesserung gegenüber den am Anfang vorliegenden Werten zu finden. Werden Optimierungsverfahren mit Simulatoren gekoppelt, so benötigt der Simulator den weitaus größten Teil der Rechenzeit. Man wird deshalb Optimierungsalgorithmen wählen, welche mit möglichst wenig Funktionsaufrufen zum Ziel führen.

Die bisher eingebundenen Verfahren (Tabelle 1) sind ausschließlich für die stetige nichtlineare Optimierung geeignet, der wichtigsten Problemklasse für technische Anwendungen. Dabei darf die Dimension des Problems nicht zu groß sein, für mehr als etwa 1000 Parameter müssen spezielle Verfahren verwendet werden.

Durch den Einsatz von Simulatoren ist i. a. über die Zielfunktion a priori nichts bekannt. Deshalb sind bei Problemen mit mehr als zwei Parametern als globale Optimierungsverfahren nur stochastische bzw. heuristische Verfahren sinnvoll. Weiterhin ist es von Vorteil, verschiedene Algorithmen zur Auswahl zu haben, da es für spezielle Problemklassen angepasste Verfahren gibt.

Verfahren	direkt	global	Schranken
Nelder-Mead Simplex	ja	nein	nein
Simulated Annealing	ja	ja	nein
BFGS	nein	nein	nein
L-BFGS-B	nein	nein	ja
Powell	ja	nein	nein
N2FB	nein	nein	ja
Conjugate Gradient	nein	nein	nein
DIRECT	ja	ja	ja
BTRK	nein	ja	ja/nein
FSQP	nein	nein	ja
Praxis	ja	nein	nein

**Tabelle 1:** Optimierungsverfahren in OPAL

Für das Ergebnis der Optimierung ist die Wahl der Zielfunktion entscheidend. So steht man häufig vor dem Problem, den Ist-Verlauf  $x^i$  einer Simulation (zeitliches Verhalten, Frequenzgang) an einen Soll-Verlauf  $x^s$  anzupassen, was auf eine nichtlineare Approximationsaufgabe führt. Misst man die Abweichung an den Punkten  $k$  der Diskretisierung in der euklidischen Norm, so führt dies auf eine Quadratmittelapproximation:

$$\|x^i - x^s\|_2 = \sqrt{\sum_k (x_k^i - x_k^s)^2} \rightarrow \text{Min!}$$

Als angepasstes Verfahren steht N2FB [13], eine Implementierung des Levenberg-Marquardt-Algorithmus [5], in OPAL zur Verfügung.

Besser ist eine Approximation in der Maximumnorm, auch als gleichmäßige Approximation bezeichnet:

$$\|x^i - x^s\|_\infty = \max_k |x_k^i - x_k^s| \rightarrow \text{Min!}$$

Für diese Aufgabe ist FSQP [8] mit entsprechend eingestellten Optionen das angepasste Verfahren in OPAL.

Globale Optimierungsverfahren benötigen im allgemeinen sehr viele Funktions- bzw. Simulatoreufrufe. Es ist also sinnvoll, globale und lokale Techniken zu koppeln. Dieser Ansatz wird z. B. in den Cluster-Methoden realisiert, von denen ein Verfahren (BTRK) von Boender, Rinnooy Kan, Timmer und Strougie in der Implementation von Csendes [19] integriert wurde. Cluster-Verfahren sind im Prinzip verbesserte Multistart-Verfahren. Bei letzte-

ren wird der Optimierungsprozess mit unterschiedlichen Startwerten durchgeführt, wodurch mehrere lokale Minima entdeckt werden können. Dabei werden i. a. jedoch viele unnötige Simulatoreufrufe durchgeführt, wenn der Optimierer für unterschiedliche Startwerte gegen das gleiche lokale Minimum konvergiert. Cluster-Verfahren durchmustern dagegen in einem ersten Schritt den Suchraum, entweder deterministisch mittels eines Gitters oder stochastisch. Anschließend bilden sie Cluster, welche die Umgebung lokaler Minima identifizieren. Danach wird in jedem Cluster ein lokales Verfahren benutzt.

Ein weiteres populäres Verfahren der globalen Optimierung ist Simulated Annealing [15], welches auch für diskrete Problemstellungen geeignet ist. Es zählt ebenfalls zu den stochastischen Verfahren.

Schließlich wurde noch DIRECT [4] in OPAL eingebunden, ein Verfahren zur Minimierung Lipschitz-stetiger Funktionen. Es ist jedoch nur für höchstens zweidimensionale Probleme geeignet, da sonst zu viele Simulatoreufrufe benötigt werden.

Optimierungsverfahren können in direkte und indirekte Verfahren unterteilt werden. Bei letzteren werden erste Ableitungen bzw. Gradienten der Zielfunktion benötigt, welche numerisch über Differenzenquotienten approximiert werden müssen. Obwohl zentrale Differenzenquotienten mit Ordnung 2 approximieren, werden Vorwärtsdifferenzen – am Rand des zulässigen Bereiches auch Rückwärtsdifferenzen – verwendet:

$$\frac{\partial}{\partial x_k} f(x) \approx \frac{f(x + h_k e_k) - f(x)}{h_k}$$

Diese approximieren zwar mit erster Ordnung, benötigen aber nur halb so viele Simulatoreufrufe.

Ein kritischer Punkt ist die Wahl der Schrittweite. Dafür ist es notwendig zu wissen, mit welcher Genauigkeit der Simulator seine Ergebnisse ausgibt. So gibt z. B. SABER höchstens 8 gültige Ziffern aus. Bei den oben erwähnten einseitigen Differenzenquotienten ist  $h_k = |x_k| \sqrt{10^{-m}}$  sinnvoll, wobei  $m$  die Anzahl der gültigen Ziffern ist. Eine ausführliche Diskussion zur Wahl der Schrittweite findet man in [5], dort jedoch im Zusammenhang mit der Maschinengenauigkeit. Bei sehr wenigen gültigen Ziffern (weniger als 4) oder wenn in der Zielfunktion Rauschen vorhanden ist, sollten nur direkte Suchverfahren benutzt werden.

Oft unterscheiden sich die Parameter um mehrere Größenordnungen. Ein wichtiger Punkt ist deshalb die Möglichkeit, die Parameter skalieren zu können, damit die Approximation der Differenzenquotienten, die Schrittweitenbestimmung bei der Strahlminimierung und die Konvergenzkriterien der

Verfahren funktionieren. In OPAL ist die diagonale Skalierung möglich, d. h. Division durch einen (bzw. Multiplikation mit einem) durch den Nutzer zu bestimmenden Faktor.

Zu indirekten Verfahren gibt es eine Vielzahl von Lehrbüchern, u. a. [5], [17]. Diese Algorithmen bestimmen in einem ersten Schritt eine Abstiegsrichtung, um danach durch eine Liniensuche (Strahlminimierung) entlang dieser Richtung eine neue Näherung für das Minimum zu finden.

Wegen ihrer schnellen Konvergenz und Zuverlässigkeit sind die Quasi-Newton-Verfahren sehr populär. Sie bilden im aktuellen Punkt ein quadratisches Modell der Zielfunktion. Die (inverse) Hesse-Matrix wird durch Aufdatierungsformeln (BFGS, DFP) approximiert, welche nur erste Ableitungen enthalten. Neben dem BFGS-Verfahren aus [15] wurde in OPAL auch L-BFGS-B [20] eingebunden, welches bereits Schranken berücksichtigt und effektiver mit dem Speicher umgeht (Limited-Memory Quasi-Newton-Verfahren).

Verfahren vom Typ der konjugierten Gradienten (cg) liefern im Gegensatz zu Quasi-Newton-Verfahren Suchrichtungen, ohne eine Matrix zu speichern, und sind deshalb auch für etwas größere Probleme geeignet. In OPAL ist das Verfahren aus [15] zu finden.

Als die gegenwärtig effizienteste universell einsetzbare Methode zur Lösung von nichtlinearen Optimierungsproblemen mit Restriktionen wird die sequentielle quadratische Approximation (SQP) angesehen. Das FSQP-Verfahren [8] ist als Vertreter dieser Klasse in OPAL eingebaut worden.

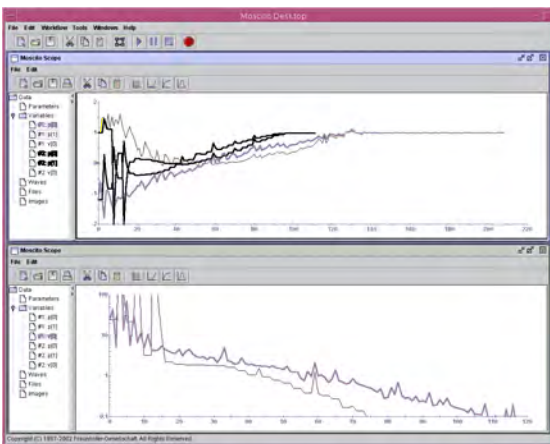
Direkte Suchverfahren arbeiten ableitungsfrei und bestimmen Suchrichtung und Schrittweite heuristisch. Das wohl am meisten verbreitete Verfahren dieser Klasse ist das Simplex-Verfahren von Nelder und Mead [15]. Es ist in der Anzahl der benötigten Simulatoreufrufe vergleichbar mit Quasi-Newton-Verfahren, kann jedoch in manchen Fällen auch versagen [14]. Hat das Problem  $n$  Parameter, so wird mit einem Simplex der Dimension  $n + 1$  gestartet und durch eine Kombination aus Reflexion, Expansion und Kontraktion dieses Simplex in einem iterativen Prozess die Optimallösung gefunden. Der Simplex passt sich dabei der lokalen Landschaft – wie z. B. engen Tälern – an.

Andere direkte Verfahren nutzen konjugierte Richtungen, so etwa das Verfahren von Powell oder der Praxis-Algorithmus von Brent [15].

Ein wichtiges Merkmal praktischer Probleme ist die Beschränkung der Parameter. Das können einfache Schranken, aber auch lineare oder nichtlineare Restriktionen sein. Sie ergeben sich aus dem Gültigkeitsbereich von Modellen oder aus prozesstechni-

schen Randbedingungen. Manche der in OPAL eingebundenen Verfahren berücksichtigen Schranken für die Parameter automatisch, FSQP kann sogar mit linearen und nichtlinearen Restriktionen umgehen. Ist das ausgewählte Verfahren für unrestringierte Probleme bestimmt, so werden Schranken und Restriktionen über Straf- (Penalty-) bzw. Barrierefunktionen berücksichtigt. Dabei ist zu beachten, dass bei praktischen Problemen der zulässige Bereich nicht verlassen werden darf. Die Verfahren sind in der Praxis durch viele Simulatoreufrufe sehr aufwendig und arbeiten wegen der zunehmend schlechten Kondition der Hessematrizen ziemlich unzuverlässig. Einen Ausweg bieten exakte Straffunktionen. Bei diesen wird der Strafparameter nicht beliebig groß, und wählt man ihn geeignet, so erhält man die Optimallösung schon nach einer einzigen unrestringierten Optimierung. Deshalb wird in OPAL auch die Betragsfunktion (absolute value penalty function) als Straffunktion eingesetzt. Um Simulationen außerhalb des zulässigen Bereiches zu vermeiden, wird dort der bisher maximale Funktionswert verwendet.

Von den zahlreichen Optimierungsverfahren in OPAL wurden in den folgenden zwei Abbildungen jeweils das Nelder-Mead-Verfahren als der im Allgemeinen am besten arbeitende Vertreter der direkten Verfahren und ein ableitungsbasiertes Verfahren gegenübergestellt.



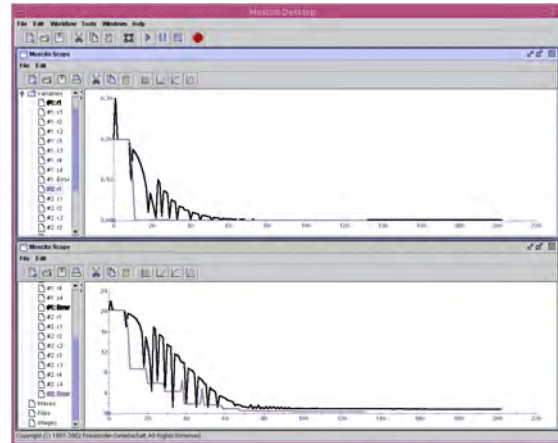
**Bild 10:** *Banana-Funktion, optimiert mit Nelder-Mead-Simplex- und FSQP-Verfahren*

Bei der Banana-Funktion

$$f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1^2) \rightarrow \text{Min!}$$

ermittelte das FSQP-Verfahren schneller das Minimum. Bei der Optimierung des Rosa-Filters sollte der Ist-Verlauf des Frequenzgangs an einen vorgegebenen Soll-Verlauf angepasst werden. Als Fehlermaß wurde die Quadratmittelnorm verwen-

det, als ableitungsbasiertes und angepasstes Verfahren N2FB. Es lieferte nicht nur etwas bessere Ergebnisse als das Nelder-Mead-Verfahren, sondern benötigte auch nur etwa halb so viel Funktionsaufrufe.



**Bild 11:** *Rosa Filter, optimiert mit Nelder-Mead-Simplex- und N2FB-Verfahren*

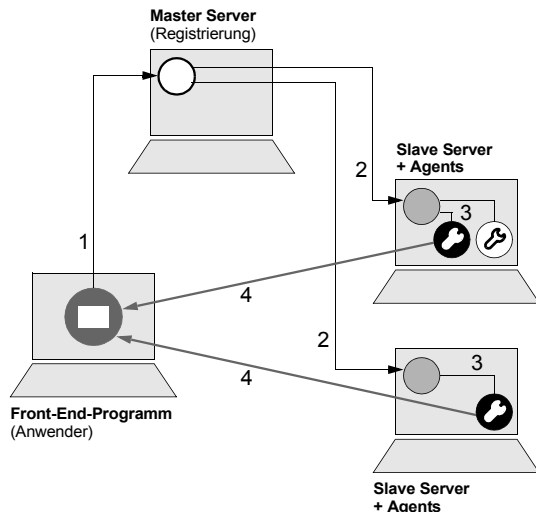
Durch den modularen Aufbau von MOSCITO können neben dem Optimierungsagenten OPAL auch andere Optimierer verwendet werden. So wurden bereits das Programm MODOS der Universität Bremen und die Optimierungs-Toolbox von Matlab erfolgreich integriert.

## 5 Internet-basierte Arbeitsweise

Neben der Möglichkeit, im Rahmen von MOSCITO Simulations-, Optimierungs- und andere Werkzeuge im Verbund nutzen zu können, besteht ein wesentlicher Vorzug in der verteilten Arbeitsweise. Das bedeutet, dass die Einzelprogramme auf verschiedenen Rechnern und ggf. an verschiedenen Orten arbeiten können. Für die Kommunikation und den Datenaustausch ist lediglich eine LAN- oder Internetverbindung erforderlich.

Zur Erläuterung der Internet-basierten Arbeitsweise von MOSCITO sollen zunächst einige Begriffe eingeführt werden.

Damit ein Anwender im Internet oder LAN die gewünschten Dienste in Form der MOSCITO-Agenten findet, ist eine zentrale Registrierung erforderlich. Diese wird über den **MOSCITO Master Server** verwaltet. Weiterhin ist es notwendig, pro Rechner, wo Dienste bzw. Agenten zur Verfügung gestellt werden, einen **MOSCITO Slave Server** zu betreiben. Während am Master-Server alle Slave-Server registriert werden, werden pro Rechner alle Agenten am (lokalen) Slave-Server angemeldet. Ein

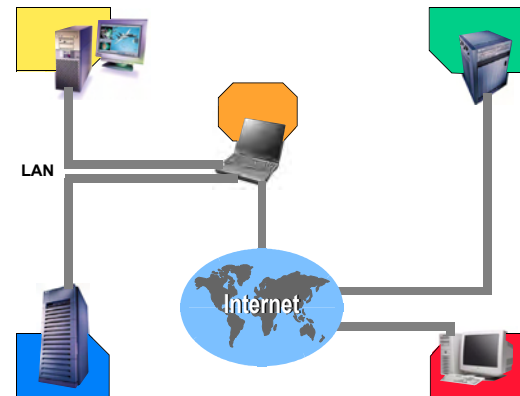


**Bild 12:** MOSCITO ermöglicht netzwerkweit das Auffinden und Starten von Agenten.

Anwender muss nun lediglich den Master-Server kontaktieren, um eine vollständige Übersicht über alle in einem Netzwerk registrierten Agenten zu erhalten. Das **MOSCITO-Front-End-Programm** unterstützt das Auffinden, Auswählen und Konfigurieren der gewünschten Agenten. Bild 12 illustriert den Ablauf. Das Front-End des Anwenders kontaktiert den Master-Server (1), dieser gibt die Suchanfrage an alle Slave-Server weiter (2). Die Slave-Server können nun Auskunft über die registrierten Agenten geben und auf Anforderung die vom Anwender gewählten Agenten starten (3). Wurde der Start erfolgreich ausgeführt, kontaktiert der Agent unmittelbar das Front-End-Programm (4), über das der Anwender dann die weitere Kontrolle aller Funktionen des Agenten übernimmt. So kann jeder Agent und damit das in ihm eingebettete Werkzeug jederzeit angehalten oder nach dem Staus befragt werden. Weiterhin ist jeder Agent in der Lage, Ergebnisdaten (z.B. nach einer Simulation) direkt an das Front-End-Programm des Anwenders zu schicken.

Durch die verteilte, Internet-basierte Arbeitsweise von MOSCITO lassen sich in erster Linie die in Abschnitt 1 genannten Ressourcen- und Verfügbarkeitsprobleme elegant lösen. So kann der Entwerfer das gesamte von MOSCITO bereitgestellte Leistungsspektrum von seinem Arbeitsplatz aus nutzen und z.B. die rechenzeitintensiven Simulationen auf entsprechend leistungsfähiger Hardware laufen lassen. Da außerdem nicht alle vom Entwerfer benötigten Programme vor Ort installiert und gewartet werden müssen, sinkt der Administrationsaufwand drastisch.

Der Preis der gewonnenen Flexibilität besteht im zusätzlichen Kommunikationsaufwand über die Netzverbindungen. Hier haben aber alle praktisch relevanten Beispielrechnungen gezeigt, dass der Zeitaufwand für die Simulation und zum Teil auch für die Optimierung deutlich höher ist als die benötigte Zeit für den Datentransfer.



**Bild 13:** MOSCITO-Agenten arbeiten im Internet verteilt auf verschiedenen Rechnern. Das Szenario entspricht dem Simulations-Optimierungs-Zyklus von Bild 4.

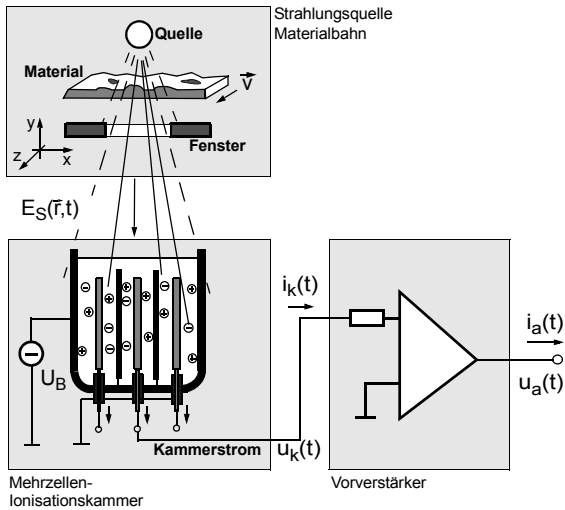
Kritisch für das korrekte Funktionieren von MOSCITO ist die Stabilität der Netzverbindungen. Spezielle Funktionen in der MOSCITO-Kommunikationsschicht sorgen bei Problemsituationen dafür, dass alle gestarteten Agenten ordnungsgemäß beendet und die Kommunikationsverbindungen korrekt abgebaut werden.

Ein weiteres Problem besteht dann, wenn aus Gründen der Netzsicherheit eine Firewall die direkte TCP/IP-Kommunikation verhindert. Da die Verbindungen zwischen den Agenten dynamisch generiert werden, genügt es nicht, einzelne Ports der Firewall für MOSCITO freizuschalten. Eine erste prototypische Lösung für dieses Problem wurde auf der Basis des SOCKS-Protokolls implementiert. In Verbindung mit verfügbaren Proxy-Servern ist es damit möglich, MOSCITO auch über Firewall-Grenzen hinweg zu betreiben.

## 6 Anwendungsbeispiele

Für das in Abschnitt 1 skizzierte Optimierungsproblem beim Entwurf der Mehrzellenionisationskammer wurde im Rahmen des Projektes ein Gesamtsystemmodell erstellt, das die Basis für die Optimierung bildet.

Aufgabe des zu optimierenden Messsystems ist es, Materialfehler – im weiteren „Störungen“ genannt – bei Bandmaterialien in x,z-Richtung sowie Dickenschwankungen zu detektieren. Die prinzipielle Struktur des Messsystems sowie die Größenordnung von Materialfehlern und die Materialgeschwindigkeiten bei industriellen Herstellungsverfahren sind in Bild 14 aufgeführt.



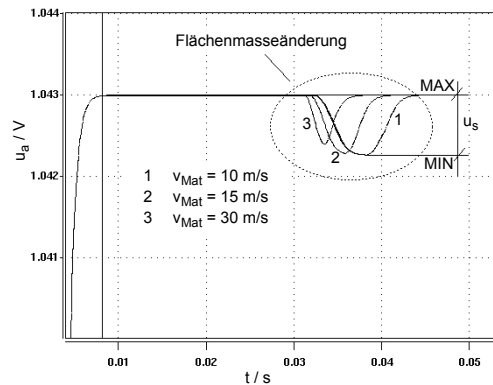
- Dickenabweichungen:  $\Delta d = 1\%$  von  $d$
- Materialstörungen:  $\Delta x, \Delta z < 5 \dots 30$  mm
- Bandgeschwindigkeit:  $v = 0,2 \dots 30$  m/s

**Bild 14:** Komponenten des radiometrischen Messsystems und technologische Anforderungen

Die hohe Bandgeschwindigkeit erfordert hohe Abtastraten und eine geringe Zeitkonstante in der Signalverarbeitung des Messsystems. Durch Simulationen am Gesamtmodell [3] lässt sich zeigen, dass die technologisch vorgegebene Bandgeschwindigkeit nicht nur die verfügbare Detektionszeit, sondern auch den Pegel des ohnehin sehr schwachen Störsignals beeinflusst. In Bild 15 ist der Signalpegel am Vorverstärkerausgang für das Durchlaufen einer 0,01 mm starken flächigen Materialverdickung aus Kupfer auf einem 0,5 mm starken Trägermaterial für verschiedene Bandgeschwindigkeiten dargestellt.

Neben den dynamischen Forderungen ist gleichzeitig eine hohe Verstärkung des von der Kammer gelieferten sehr kleinen Stromsignals erforderlich ( $\Delta i_K(t) \ll 10$  nA). Geringe Zeitkonstante und hohe Verstärkung sind konkurrierende Auslegungskriterien und können nur sinnvoll im Zusammenspiel der einzelnen Komponenten beurteilt werden. Damit ist eine Optimierung der gesamten Systemkette erforderlich. Ziel der Optimierung war

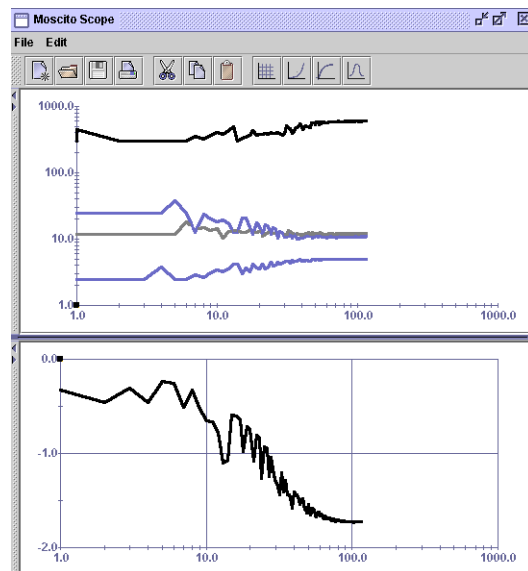
es, einen ausreichend hohen (maximalen) Störsignalpegel  $u_S$  (s. Bild 15) auch bei großen Bandgeschwindigkeiten zu erreichen.



**Bild 15:** Ausgangsspannung des Verstärkers bei verschiedenen Bandgeschwindigkeiten.

Optimierungsparameter, die den größten Einfluss auf den Zielfunktionswert erwarten lassen, sind

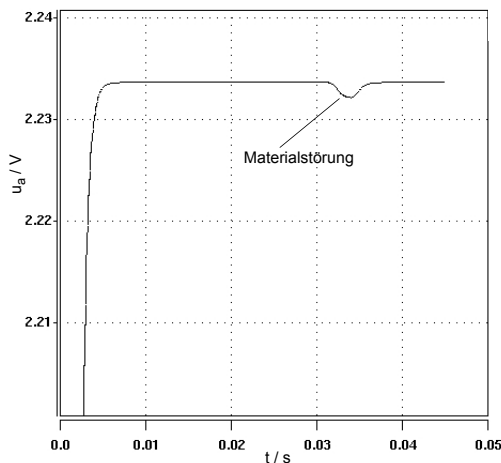
- die Geometrie der Einstrahlungsverhältnisse, d.h. die Abstände Quelle-Bahn  $a_1$  und Bahn-Kammer  $a_2$ ,
- der Kammerinnendruck  $p$ ,
- die Kammer-Betriebsspannung  $U_B$  und insbesondere
- die Verstärkerbeschaltung  $r5, c5$  sowie
- die geometrischen Parameter der Kammer.



**Bild 16:** Verlauf der Optimierung: Parameter (oben) und Zielfunktionswert (unten).

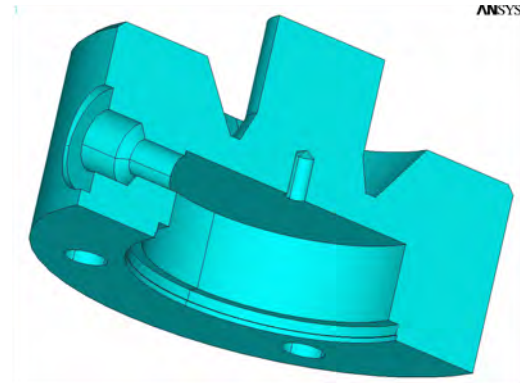
Das Gesamtsystemmodell des Messsystems wurde für den Simulator SABER erstellt [3] – es kommt somit der MOSCITO-SABER-Agent zum

Einsatz. Da eine Simulation nur wenige Sekunden dauert, kann der Agent auf einer Workstation mit durchschnittlicher Leistungsfähigkeit mit dem Betriebssystem Solaris laufen. Ergebnisse der Optimierung [3] unter Verwendung der o.a. Parameter, allerdings für eine feste Kammergeometrie, sind in Bild 16 und Bild 17 dargestellt. Bild 16 zeigt den Verlauf von Optimierungsparametern (oben) und Zielfunktionswert für das gewählte Optimierungsverfahren Nelder-Mead-Simplex. Der Zielfunktionswert der optimierten Lösung liegt bei  $Z \sim 1,7$  mV. Dies ist eine deutliche Verbesserung gegenüber der Startlösung mit  $Z = u_S < 1$  mV. Bild 17 zeigt den Spannungsverlauf am Verstärkerausgang für ein Messsystem mit optimalen Designparametern, berechnet für eine Materialstörung von 0,01 mm Dicke auf 0,5 mm Kupferträger.



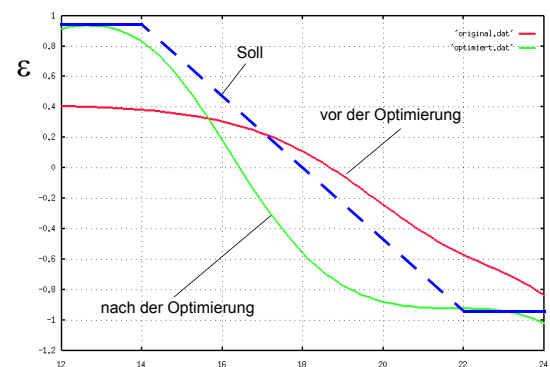
**Bild 17:** Auswirkung einer Materialstörung auf die Spannung am Verstärkerausgang bei optimalen Designparametern.

Für das zweite Beispiel – die Optimierung eines Kraftsensors – wurde zur Berechnung des Systemverhaltens der MOSCITO-ANSYS-Agent verwendet. Da eine FEM-Simulation je nach Komplexität der Struktur, verwendetem Modell und gewähltem Detaillierungsgrad zum Teil extrem lange Rechenzeiten benötigt, war es für die Optimierung, bei der pro Optimierungsschritt ein Simulationslauf notwendig ist, zwingend notwendig, schnelle Hardware zu nutzen. Mit Hilfe von MOSCITO konnte ANSYS auf einer Sun Workstation mit zwei Prozessoren und 1 GByte Hauptspeicher laufen, während der Optimierungsfortschritt einschließlich der simulierten Zwischenergebnisse am lokalen PC online verfolgt werden konnte. Die Zeit für eine 2D-Simulation dauerte etwa 40 Sekunden, für das 3D-Modell wurde sogar eine halbe Stunde benötigt. Bild 19 zeigt die mit dem in OPAL integrierten BTRK-Ver-



**Bild 18:** 3D-Modell des Kraftsensors. Die Geometrie der Kerbe an der Oberseite soll optimiert werden.

fahren erzielten Ergebnisse. Die Kraftverteilung konnte deutlich an die Sollvorgaben angenähert werden, woraus sich eine Verbesserung der Fertigungstoleranzen ergab.

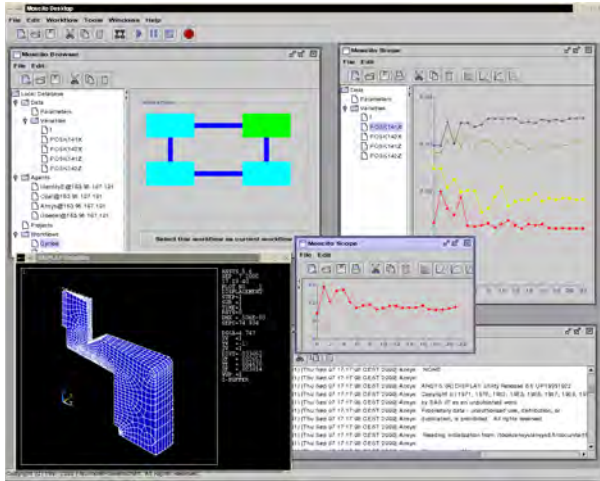


**Bild 19:** Verbesserte Kraftverteilung am Sensor durch Optimierung der Kerbengeometrie.

## 7 Zusammenfassung

MOSCITO wurde als Internet-basiertes Simulations-Optimierungs-System vorgestellt, das dem Entwerfer über ein grafisches Front-End eine breite Palette an Simulationswerkzeugen und Optimierungsverfahren zur Verfügung stellt. Die modulare, Agenten-basierte Architektur von MOSCITO gestattet, frei verfügbare wie kommerzielle Entwurfswerkzeuge zu Workflows miteinander zu verbinden und über LAN oder Internet vom Arbeitsplatz des Entwerfers aus zu nutzen. Speziell für die Optimierung von Mikrosystemen wurden 6 Simulationswerkzeuge und 11 Optimierungsverfahren in MOSCITO integriert und an Anwendungsbe-

spielen getestet. Derzeit befinden sich MOSCITO-Installationen bei Partnern in Bremen, Tallinn, Darmstadt und Bratislava, so dass vor allem bei der Internet-basierten Werkzeugnutzung wertvolle Erfahrungen gesammelt werden können.



**Bild 20:** MOSCITO-Desktop – die grafische Nutzerschnittstelle.

Das diesem Beitrag zugrunde liegende Vorhaben „OMID – Optimierung von Mikrosystemen für Diagnose- und Überwachungsanwendungen“ wird mit Mitteln des Bundesministeriums für Bildung, Wissenschaft, Forschung und Technologie (BMBF) unter dem Kennzeichen 16 SV 988/2 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt allein bei den Autoren.

## 8 Literatur

- [1] Blume, C.: GLEAM - A System for intuitive Learning. Proc. of PPSN I, LNCS 496, Springer-Verlag, S.48-54, 1991.
- [2] Bolte, H.; Laur, R.; Huck, E.; Diener, K.-H.; Schneider, P.; Schwarz, P.: Entwurfsunterstützung für radiometrische Messsysteme - physikalische Modellierung, Simulation und Optimierung. OMID-Statusseminar, Bremen, 2001
- [3] Diener, K.-H.; Huck, E.; Schneider, P.; Schwarz, P.: Entwurfsunterstützung für radiometrische Messsysteme - Kammerelektronik und Gesamtsystemmodell. diese Tagung
- [4] Gablonsky, J. M.: Modifications of the DIRECT Algorithm. PhD Thesis, North Carolina State University, 2001
- [5] Gill, P. E.; Murray, W.; Wright, M. H.: Practical Optimization. Academic Press, San Diego, 1997.
- [6] Jakob, W.; Gorges-Schleuter, M.; Blume, C.: Application of Genetic Algorithms to Task Planning and Learning. In: Männer, R.; Manderick, B. (eds.): Conf. Proc. PPSN II, North-Holland, Amsterdam, S.291-300, 1992
- [7] Jakob, W.; Gorges-Schleuter, M.; Sieber, I.; Süß, W.; Eggert, H.: Solving a Highly Multimodal Design Optimization Problem Using the Extended Genetic Algorithm GLEAM. In: S. Hernandez, A.J. Kassab, C. A. Brebbia: Computer Aided Design of Structures VI, WIT Press, Southampton, S.205-214, 1999.
- [8] Lawrence, C. T.; Zhou, J. L.; Tits, A. L.: User's Guide for CFSQP Version 2.3: A C Code for Solving (Large Scale) Constrained Nonlinear (Minimax) Optimization Problems, Generating Iterates Satisfying All Inequality Constraints. Institute for Systems Research, College Park, Maryland, 1995
- [9] Modelica:  
<http://www.Modelica.org>
- [10] MOSCITO:  
<http://www.eas.iis.fhg.de/solutions/moscito>
- [11] Peters, D.; Bolte, H.; Böhnke, R.; Bischoff, L. und Laur, R.: MODOS - Model Based Design Optimization System. IEEE Proceedings of the 43rd Midwest Symposium on Circuits and Systems, Lansing, Michigan, August 2000
- [12] Peters, D.; Laur, R.: An Improvement of Rosenbrocks Algorithm for Model Based Design Optimization of Microsystems. Sixth IEEE International Conference on Electronics, Circuits and Systems; Zypern, 1999
- [13] PORT Mathematical Subroutine Library:  
<http://www.bell-labs.com/project/PORT/>
- [14] Powell, M. J. D.: Direct search algorithms for optimization calculations. In: Acta Numerica Vol. 7. Cambridge University Press, 1998
- [15] Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T.: Numerical recipes in C. The art of scientific computing. 2. Auflage. Cambridge University Press, 1992
- [16] Schneider, P.; Huck, E.; Reitz, S.; Parodat, S.; Schneider, A.; Schwarz, P.: A modular approach for simulation-based optimization of MEMS. Proc. ISMA 2000, Singapore, November 2000.
- [17] Spellucci, P.: Numerische Verfahren der nichtlinearen Optimierung. Basel [u. a.]: Birkhäuser, 1993
- [18] Stenzel, R. et al.: Simulation of AlGaIn&GaIn-HFETs including spontaneous and piezoelectric polarisation charges. Proc. 26th International Symposium On Compound Semiconductors, Berlin 1999
- [19] Törn, A.; Zilinskas, A.: Global optimization. Berlin [u. a.]: Springer, 1989
- [20] Zhu, C.; Byrd, R. H.; Lu, P.; Nocedal, J.: L-BFGS-B: FORTRAN subroutines for large-scale bound constrained optimization. ACM Trans. Math. Software 23 (1997), S. 550-560

# ORBIT – INTERNEBASIERTE PLATTFORM FÜR REMOTE-ANWENDUNGEN

Dipl.-Inform. Andreas Reiffer  
InterConnect Software GmbH, Karlsruhe  
Email: [reiffer@interconnect.de](mailto:reiffer@interconnect.de)

Dipl.-Inform. Peter Rohnacher  
InterConnect Software GmbH Karlsruhe  
Email: [rohnacher@interconnect.de](mailto:rohnacher@interconnect.de)

## Kurzfassung

Der vorliegende Beitrag beschreibt die im Rahmen des Verbundprojekts OMID entwickelte Plattform für Remote-Anwendungen **ORBIT**. Sie gestattet es über normale WWW-Browser komplexe Simulations- und Optimierungsläufe auf entfernten Rechnern durchzuführen und zu verwalten. Die Anbindung der Benutzer erfolgt durch das Standard-http-Protokoll, die Kommunikation mit den Tool-Servern wird mit SOAP durchgeführt. Kernkomponente des ORBIT-Systems die objekt-orientierte Datenbank, in der alle Modell- und Ergebnisparameter abgespeichert und abgerufen werden.

## 1 Einleitung

Im Rahmen des Verbundprojekts OMID soll durch die beteiligten Forschungseinrichtungen ein verteiltes Dienstleistungszentrum aufgebaut und eingesetzt werden. Das in diesem Beitrag vorgestellte System ORBIT unterstützt hierbei sowohl die Forschungseinrichtungen als Dienstleister, als auch die Industrieunternehmen als Nutzer.

Die netzwerktechnische Basis für das Softwaresystem ist das Internet. Alle Projektpartner verfügen über Zugangsmöglichkeiten zum Internet und sind jederzeit online über das Internet zu erreichen. Der Nutzen für die Anwender ist klar erkennbar: Forschungseinrichtungen oder Partner der Großindustrie setzen das ORBIT-Framework als Dienstleistungsbasis für die effiziente und kostengünstige Bereitstellung ihrer Ressourcen zur Nutzung durch andere Partner ein.

## 2 Systemarchitektur

Die ORBIT-Systemarchitektur besteht aus drei Hauptkomponenten, wobei jede dieser Komponenten im Bedarfsfall auch mehrfach vertreten sein kann (Abb. 1):

- Tool-Anwender
- Tool-Server
- ORBIT-System

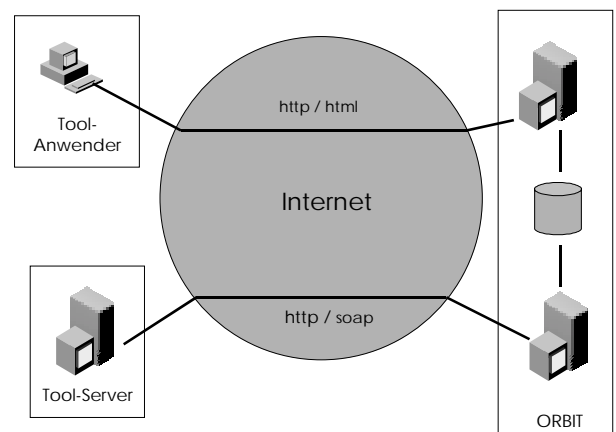


Abbildung 1: ORBIT-Systemarchitektur

### 2.1 Tool-Anwender

Der Tool-Anwender verbindet sich über das Internet mit dem ORBIT-System. Dazu ist auf Anwenderseite lediglich ein herkömmlicher http-Browser notwendig. Die gesamte Kommunikation von und zum ORBIT-System erfolgt über HTML-Seiten und -Formulare.

### 2.2 Tool-Server

Der Tool-Server ist für die Berechnung und Durchführung der eigentlichen Simulations- und Optimierungsaufträge zuständig. Er steht



ebenfalls über das Internet in Verbindung mit ORBIT-System. Die Kommunikation erfolgt dabei ausschließlich über das HTTP-basierte SOAP-Protokoll (Simple Object Access Protokoll). Der entscheidende Vorteil dieses Plattform-unabhängigen Protokolls ist offensichtlich: Da als „Transportmedium“ das weit verbreitete HTTP-Protokoll verwendet wird, kann ein Tool-Server auch in geschützten Umgebungen eingesetzt werden, ohne dass Interferenzen mit Firewalls oder anderen Sicherheitsinstanzen zu erwarten sind.

## 2.3 ORBIT-System

Das ORBIT-System besteht aus drei weiteren Einzelkomponenten sowie einem unterstützenden Tool zur Workflow-Modellierung:

- Internet Web-Server
- Data-Server
- ORBIT-Server
- Unterstützend: Workflow-Editor

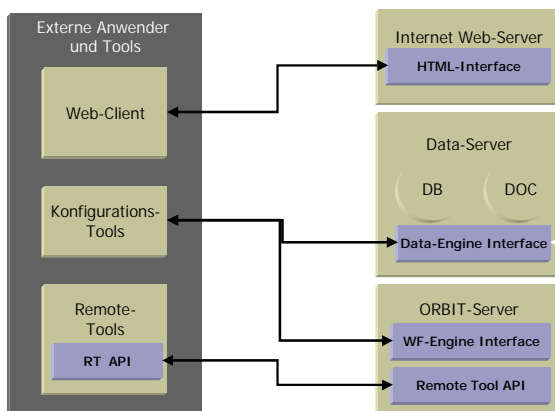


Abbildung 2: ORBIT-System

### 2.3.1 Internet Web-Server

Die Hauptaufgabe des Internet Web-Servers ist die Kommunikation mit dem Tool-Anwender. Auf Anforderung des Benutzers werden die Daten aus der Datenbank ermittelt, bzw. dort abgelegt und dem Benutzer in Form von HTML-Seiten zur Verfügung gestellt. Der http-Server ist ebenfalls für die Verwaltung der Datenbank, die Auftragsgenerierung für den Kommunikationsserver, sowie die XML-Generierung der Auftragsdaten zuständig. Über die generische HTML-Schnittstelle werden aus dem Data-Server automatisch HTML-

Formulare für die Konfiguration der Modelle sowie für die Abfrage der Berechnungsergebnisse generiert.

### 2.3.2 Datenbank-Server (Data-Server)

Der Datenbank-Server sorgt für Speicherung der Auftrags- und Ergebnisdaten in der Datenbank. Das zugrundeliegende relationale Datenmodell ist aus Gründen der Polymorphie mit einem objektorientierten Aufsatz versehen. Das objektorientierte Datenmodell ist im weiteren Verlauf dieses Beitrags beschrieben. Neben den eigentlichen Nutzdaten der jeweiligen Anwendung sind im Data – Server zusätzlich die notwendigen Workflowbeschreibungen hinterlegt.

Mittels der im Data-Server vorgehaltenen Konfigurationsschnittstelle besteht die Möglichkeit zur Konfiguration der Daten, Modelle und Workflows durch zusätzliche externe Tools. Durch diese Schnittstelle ist es möglich, im Bedarfsfall das generische HTML-Interface im Internet Web-Server durch ein individuelles Präsentationstool zu ersetzen und hierüber die Manipulation der Daten, Modelle und Workflows vorzunehmen.

### 2.3.3 ORBIT-Server

Der ORBIT-Server wird mit den XML-generierten Auftragsdaten aus der Datenbank versorgt und verschickt sie über SOAP an den entsprechenden Tool-Server zur weiteren Verarbeitung. Nach Abschluss der Berechnungen ist der ORBIT-Server für die Speicherung der Ergebnisdaten in der Datenbank verantwortlich, damit diese später vom Internet Web-Server an den Benutzer übermittelt werden können. Darüber hinaus überwacht der ORBIT-Server die Einhaltung und Weichaltung von Workflows zur korrekten Abwicklung komplexer, zusammengesetzter Aufträge. Die Kommunikation zwischen dem jeweiligen Remote Tool und dem ORBIT-Server erfolgt durch die Remote-Tool-API. Zur Steuerung der Workflows bietet der ORBIT-Server ein Programmier-Interface für die Workflow-Engine.

## 3 Workflows in ORBIT

Anwendungsprozesse, die durch ORBIT

gesteuert werden, bestehen in der Regel aus mehreren Schritten, die jedoch nicht zwingend sequentiell ausgeführt werden müssen (Zyklen, paralleles Aufrufen, bedingte Ausführung, etc.). Um diesen Anforderungen gerecht zu werden, verfügt ORBIT über eine Workflow-Engine, die die Verwaltung der Anwendungsprozessstrukturen in Form von dynamischen Workflows zulässt. Workflows werden in ORBIT – vereinfacht ausgedrückt – als gerichtete Graphen mit Knoten und Kanten abgelegt. Ein Knoten eines Graphen definiert den jeweiligen Teilschritt des Prozesses, eine Kante definiert die Vorgänger-Nachfolger-Beziehung zwischen jeweils zwei Knoten.

### 3.1 Datenstruktur eines Workflows

Die datenstrukturelle Grundlage eines Workflows in ORBIT basiert auf Empfehlungen der Workflow Management Coalition (WfMC, <http://www.wfmc.org>). Dieses Gremium, mit über 300 Mitgliedsorganisationen weltweit, definiert eine Reihe von Standards um den Themenkreis Workflow. Im ORBIT-System wurde bei der Implementierung auf eine weitgehende Einhaltung der dort empfohlenen Standards geachtet, so dass die Kompatibilität mit anderen WfMC-konformen Systemen gegeben ist.

Eine Workflowdefinition in ORBIT besteht im Wesentlichen aus den Grundelementen

- Participant
- Data
- Activity
- Transition

Das Grundelement „Participant“ beschreibt den Teilnehmer eines Workflows. In Bezug auf das ORBIT-System werden hier die Daten eines Tool-Clients, bzw. Tool-Anwenders hinterlegt. Das Data-Element trägt Nutzdaten, die zwischen einzelnen Participants und Activities ausgetauscht werden. Mit Bezug auf ORBIT handelt es sich hierbei um die Modelldaten, Parametersätze und Ergebnisse. Durch das Element „Activity“ wird ein einzelner Prozessschritt näher beschrieben. In ihm werden die Konfigurations- und

Aufrufparameter der Remote-Tools abgelegt. Die Transition-Elemente entsprechen den Kanten im gerichteten Graph und definieren die Übergänge zwischen einzelnen Activities. Nachfolgende Abbildung zeigt die graphische Repräsentation eines typischen Workflows mit den Participants „Sim1“ und „Opt1“.

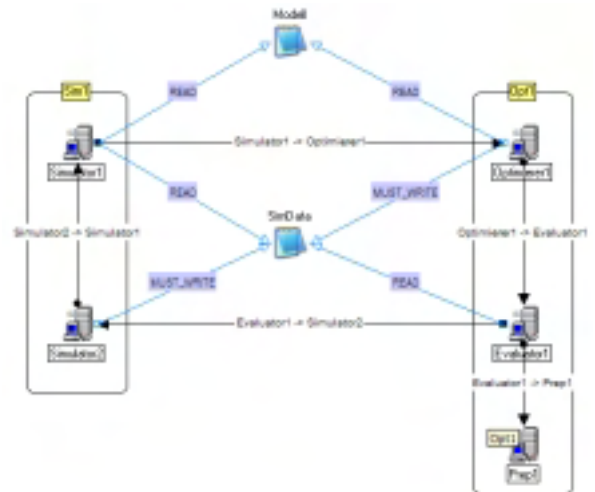


Abbildung 3: Typischer Workflow

### 3.2 Workflow-Editor

Der manuelle syntaktisch korrekte Entwurf von Workflowbeschreibungen erfordert vom Bearbeiter ein sehr detailliertes Wissen um den datenstrukturellen Aufbau der Datei. Um den Anwender bei dieser Arbeit zu entlasten, wurde ein graphischer Editor für den Entwurf von Workflows für das ORBIT-System entwickelt. Folgende Abbildung zeigt einen Screen-Shot des Workflow-Editors.



Abbildung 4: Der Workflow-Editor

Der Editor ermöglicht die graphische Eingabe

des Workflows mit seinen Komponenten. Über spezielle Dialog-Fenster werden die einzelnen Elemente des Workflows konfiguriert. Die Abbildung oben zeigt beispielhaft das Dialog-Fenster eines Participant-Elements.

### 3.3 Austauschformat der Workflowdaten

Zur Verwendung im ORBIT-System müssen die mit dem Editor erstellten Workflows exportiert und über das WF-Engine-Interface im ORBIT-Server importiert werden. Als Austauschformat für die Daten wird XML verwendet. Die zugrunde liegende Struktur ist ebenfalls WfMC-konform. Folgende Abbildung zeigt einen Ausschnitt einer Workflow-Definition in XML.

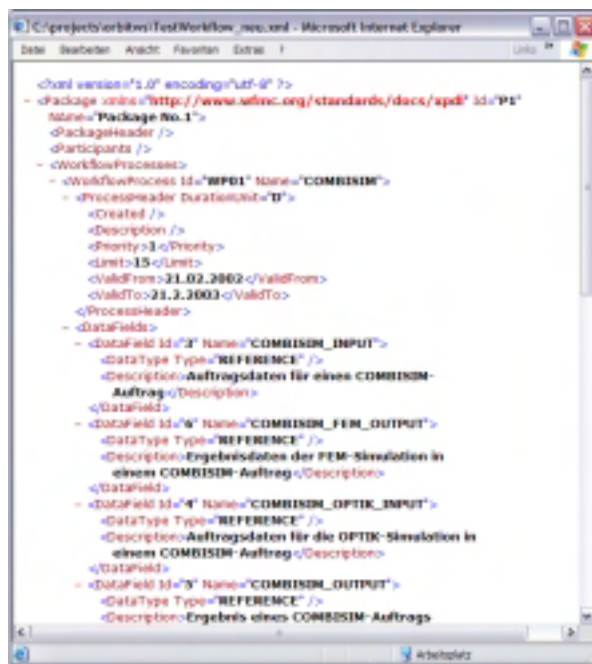


Abbildung 5: Workflow-Beschreibung in XML

## 4 Einsatz-Szenario

Ein typischer Ablauf eines Simulationsauftrags in ORBIT-System besteht aus vier Schritten:

- Schritt 1: Start der Simulation
- Schritt 2: Abfrage von Statusinformationen

- Schritt 3: Ergebnisabfrage
- Schritt 4: Abfrage statischer Ergebnisse

### 4.1 Schritt 1: Start der Simulation

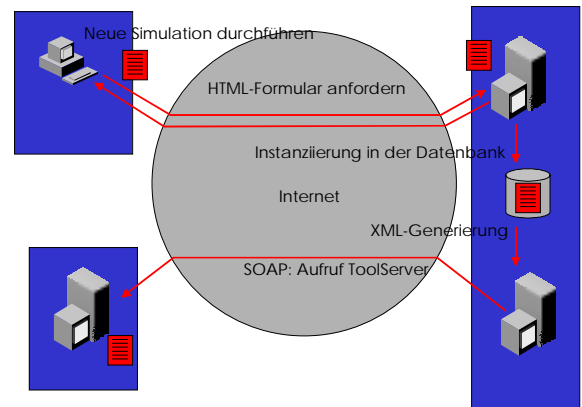


Abbildung 6

Der Tool-Anwender fordert vom http-Server ein HTML-Formular an, in das er die Modelldaten der Simulation einträgt. Dabei kann, je nach Art der Simulation ebenfalls aus einer Liste bereits bestehender Modelle ausgewählt werden. Die eingegebenen Daten werden dann vom http-Server in die Datenbank eingetragen und es erfolgt die Auftragsgenerierung und Übermittlung der XML-Daten an den ORBIT-Server. Der ORBIT-Server verbindet sich daraufhin mit dem passenden Toolserver und übermittelt ihm die Auftragsdaten zur Berechnung. Nach positiver Quittierung wird der Start in der Datenbank protokolliert.

## 4.2 Schritt 2: Abfrage von Statusinformationen

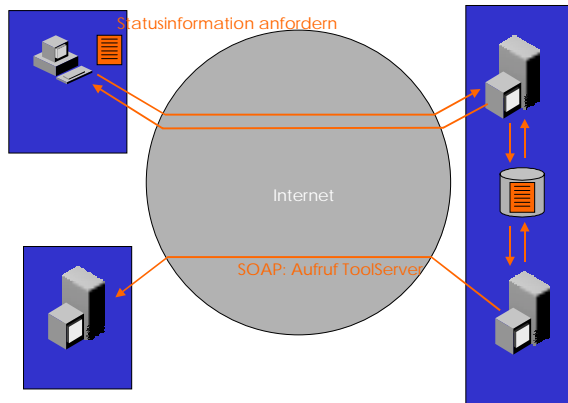


Abbildung 7

Bei den meist relativ zeitintensiven Berechnungen kann der Benutzer in der Regel keine sofortige Rückübermittlung der Ergebnisse erwarten. Daher ist es sinnvoll, dem Benutzer die Möglichkeit zum Abruf von Statusinformationen über den abgesetzten Auftrag zu geben. Der Ablauf ist ähnlich zu dem in Schritt 1: Nach Übermittlung der Anfrage vom Benutzer an den http-Server veranlasst dieser den Abruf von Statusinformationen des Tool-Servers über den ORBIT-Server. Diese Informationen werden dann wieder in der Datenbank abgelegt und werden zum Abschluss vom http-Server an den Tool-Anwender übermittelt.

## 4.3 Schritt 3: Ergebnisabfrage

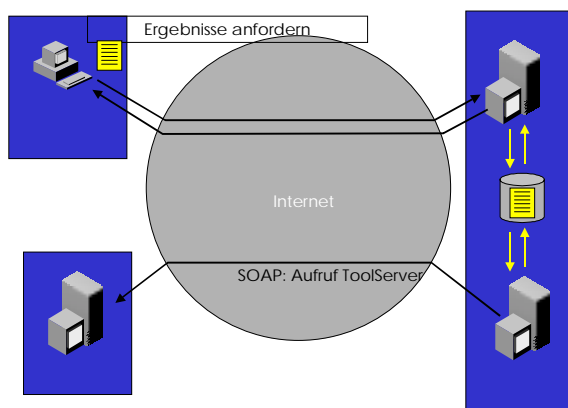


Abbildung 8

Nach Abschluss der Berechnungen im Tool-Server stehen die Ergebnisse zur Abfrage

durch den Tool-Anwender bereit. Die Vorgehensweise ist vergleichbar mit den Schritten 1 und 2. Der http-Server übermittelt die Ergebnisanforderung an den ORBIT-Server, der seinerseits für den Abruf der Daten beim Tool-Server sorgt. Die gelieferten Ergebnisdaten werden dann in der Datenbank abgelegt und als Links für den Benutzer abrufbar. Alternativ können die Ergebnisdaten auch auf dem Tool-Server verbleiben. Bei dieser Variante müssen die Ergebnisdaten jedoch für den Tool-Anwender direkt bei Tool-Server abrufbar sein (Zugriffsrechte).

## 4.4 Schritt 4: Abfrage statischer Ergebnisse

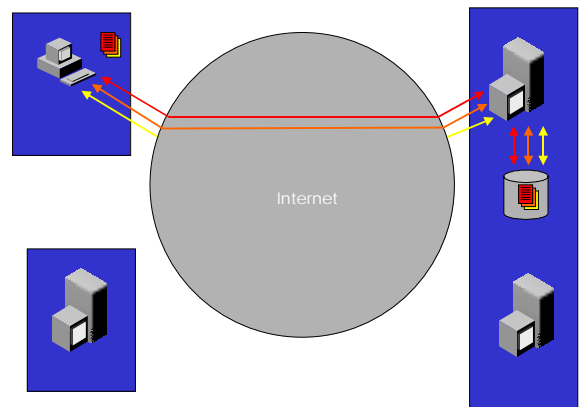


Abbildung 9

Ein wesentlicher Gesichtspunkt bei der Konzeption der ORBIT-Struktur ist die zentrale Rolle der Datenbank. Aufgrund der unter Umständen langen Berechnungszeiten für Aufträge ist es sinnvoll, Auftrags- und Ergebnisdaten in der Datenbank dauerhaft zu speichern. Dadurch ist auch ein späterer statischer Abruf von Daten durch den Benutzer möglich. Sollten vom Benutzer Auftragsdaten eines bereits berechneten Auftrags vorhanden sein, wird dies bereits beim Initialisierungsprozess durch den http-Server festgestellt und der Benutzer direkt mit den zugehörigen Ergebnisdaten versorgt. Die zeitaufwendige Neuberechnung auf dem Tool-Server entfällt.

## 5 Beispiel: Simulation der Aktorplatte

Im Rahmen der Verbundprojekts OMID wird Simulation und Optimierung einer Aktorplatte durchgeführt. Zur Simulation werden dabei Modellparameter an einen entsprechend konfigurierten Tool-Server im Forschungszentrum Karlsruhe Technik und Umwelt übermittelt. Dieser Tool-Server führt dann die Berechnungen durch und stellt die Ergebnisse zur Verfügung. Dieses Simulationsszenario ist als ORBIT-Anwendung konfiguriert und kann so von jedem zugelassenen Anwender per WWW-Browser aufgerufen werden. Abbildung 6 zeigt die HTML-Auswahlmaske von verschiedenen Modellen sowie den Konfigurationsdialog für ein Modell.

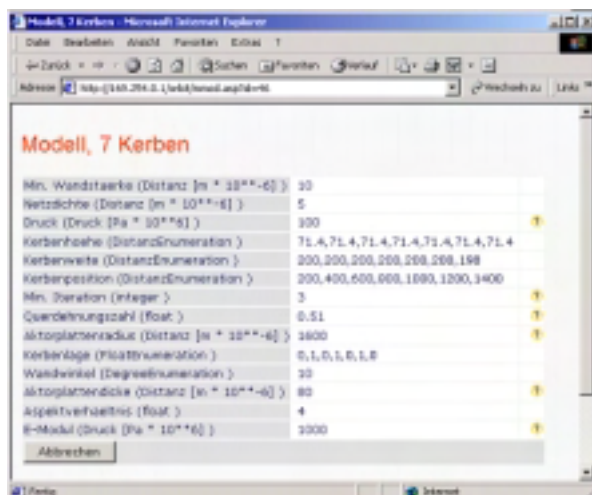
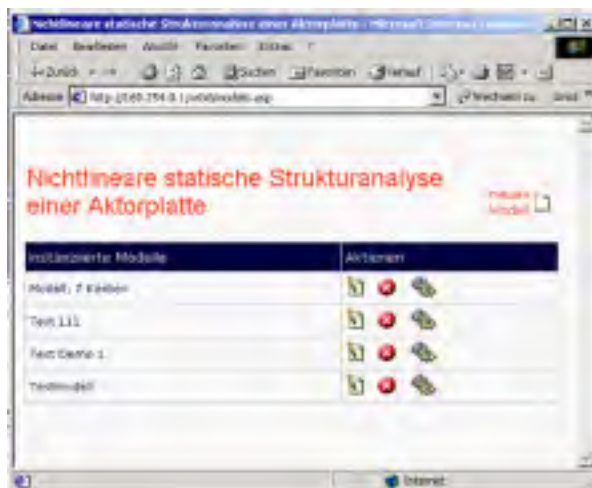


Abbildung 10

Nachdem die Berechnung durchgeführt wurde, können die Ergebnisdateien per

HTML-Link im WWW-Browsers des Anwenders angezeigt werden (Abbildung 7).

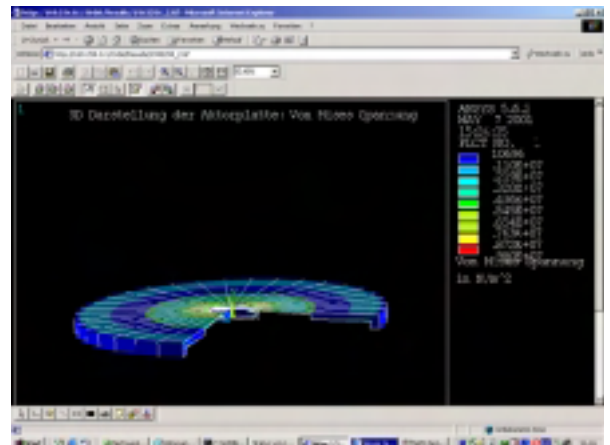


Abbildung 11

### 5.1 Weiteres Beispiel

Ein weiteres Beispiel zum Einsatz des ORBIT-Systems in OMID für die internetbasierte Designoptimierung findet man unter [6].

## 6 Einbinden von Tool-Servern in ORBIT

Beim Betreiben eines Workflow-Systems muss davon ausgegangen werden, dass das System in einer existierenden nicht einheitlichen IT-Infrastruktur eingesetzt wird. Bestimmte Schritte in einem Arbeitsablauf müssen mit Anwendungen durchgeführt werden, die auf den unterschiedlichsten Betriebssystemen mit unterschiedlichen Kommunikationsmechanismen (stdin/stdout, OLE, DDE, Corba, DCOM) installiert sind. Damit diese Arbeitsschritte im Workflow-System kontrolliert werden können und ein Datenaustausch mit solchen Arbeitsschritten gewährleistet ist, muss den Entwicklern von Workflow-fähigen Programmen eine einfache und flexible Schnittstelle zur Verfügung gestellt werden, die diese Anforderungen ermöglicht.

Existierende Anwendungen, die in einen Workflow integriert werden müssen und diese Schnittstelle nicht unterstützen, werden über spezielle Programme, den „Tool Agents“ eingebunden. Tool Agents kommunizieren mit dem Workflow-System

über die in diesem Dokument vorgestellte Schnittstelle und mit der zu integrierenden Anwendung mit deren optimalen Kommunikationsmechanismen. Somit ist die Kontrolle und der Datenaustausch zwischen Workflow-System und Anwendung sichergestellt.

## 6.1 Schnittstellenbeschreibung

Zur Einbindung von Workflow-fähigen Programmen bzw. „*Tool Agents*“ in ORBIT wird das Protokoll SOAP (Simple Object Access Protocol) eingesetzt. SOAP [1] ist ein öffentlicher Standard und geht auf eine Initiative von mehreren großen Herstellern zurück (z.B. IBM Microsoft, Lotus Development Corp.). SOAP definiert eine Art von RPC-Mechanismus, der HTTP als Transportprotokoll und XML als Kodierungsschema benutzt. Aufgrund der Verwendung der weitverbreiteten Standards HTTP und XML ist eine Implementierung auf den unterschiedlichsten Systemen einfach möglich. Des Weiteren lässt sich der Kommunikationsmechanismus problemlos in sicherheitskritische Umgebungen (Firewall-geschützte Netze) einfügen. Zur Funktionalität der Schnittstelle wurden die Spezifikationen der „Workflow Management Coalition“ [2], [3] berücksichtigt.

Aus der Sicht von ORBIT ist die einzubindende Anwendung ein Objekt, dessen Kontrolle vollständig in ORBIT liegt. ORBIT kontrolliert das Starten, Beenden und den Status dieser Anwendung und ermöglicht den Datenaustausch zwischen Anwendung und ORBIT.

Folgende Methoden sind notwendig, um mit ORBIT kommunizieren zu können:

- **OAConnect** verbindet ORBIT (Workflow Engine) mit dem Tool Agent
- **OAdisconnect** trennt ORBIT (Workflow Engine) von dem Tool Agent
- **OAinitApp** veranlasst den Tool-Client, seine Anwendung zu initialisieren

- **OAstartApp** veranlasst den Tool-Client, seine Anwendung zu starten, bzw. zu aktivieren, wenn das Starten schon beim Initialisieren durchgeführt wurde.
- **OATermApp** fordert die Anwendung auf, sich zu beenden
- **OAreqApp** erlaubt dem Workflow-System, den Status einer laufenden Anwendung abzufragen

## 7 ORBIT-Datenstrukturen

ORBIT verwendet im Kommunikationsserver eine relationale Datenbank, in der alle Auftragsrelevanten Daten abgelegt werden. Dazu zählen sowohl die Nutzdaten zur Abwicklung der Aufträge (Modelldaten, Parameterlisten, etc.) , als auch durch externe Tools berechnete Ergebnisdaten. Um einen einheitlichen Zugang für die möglichen Simulations- und Optimierungsszenarien zu bieten, wurde ein objekt-orientiertes Datenmodell auf Basis der relationalen SQL-Datenbank entwickelt. Im Folgenden wird die Struktur und Handhabung dieses Datenmodells näher beleuchtet.

### 7.1 Objekt-Orientiertes Datenmodell

Ein Objektorientiertes Datenmodell unterscheidet sich von einem herkömmlichen (relationalen) Datenmodell in der Strukturierbarkeit und Vererbbarkeit der einzelnen modellierten Entitäten. Die grundlegende Datenstruktur im objekt-orientierten Sinn ist die „Klasse“. Eine Klasse definiert eine Entität mit den dazugehörigen Eigenschaften. Diese Eigenschaften werden durch Attribute (Eigenschaftswerte) und Methoden der Klasse bestimmt. Im Rahmen der ORBIT-Konzeption ist die Modellierung von Methoden derzeit jedoch nicht vorgesehen. Ein weiteres Grundelement des objekt-orientierten Ansatzes ist die Vererbung: Klassen können die Eigenschaften von übergeordneten Klassen „erben“ und diese durch Hinzufügen spezifischer Eigenschaften weiter spezialisieren.

Zusätzlich zur Klassifizierung und Vererbung kennt das ORBIT-Datenmodell den Mechanismus der Delegation. Eine instanzierte Klasse (Objekt) kann ein weiteres Objekt als Default-Größe referenzieren. Nicht explizit angegebene Attributwerte im Objekt werden dann durch die angegebenen Attributwerte aus dem Default-Objekt ersetzt. Mit dem Mechanismus der Delegation können so Modellvarianten umgesetzt werden.

### 7.1.1 Klassen

Ein Klasse wird festgelegt durch

- eine eindeutige ID (Primärschlüssel),
- die ID der zugehörigen Elternklasse und
- ihren Namen.

Die ID dient der eindeutigen Identifikation der Klasse. Bei geschickter Wahl kann jedoch auch das Namensfeld zur Identifikation verwendet werden. Ist die ID der Elternklasse gleich NULL, handelt es sich um eine Basisklasse. Pro Klasse ist maximal eine Elternklasse zulässig (keine Mehrfachvererbung).

### 7.1.2 Attribute zu Klassen

Ein Attribut einer Klasse wird festgelegt durch

- eine eindeutige ID (Primärschlüssel)
- die ID der zugehörigen Klasse
- den Namen des Attributs
- die ID des zugeordneten Attributtyps (s.u.)
- einen beschreibenden Hilfetext und
- einen Vorgabewert.

Die ID des zugeordneten Attributtyps entspricht einem Verweis auf die Tabelle der Attributtypen. Mittels des optionalen Hilfetext können erläuternde Informationen über dieses Attribut für den Benutzer hinterlegt werden. Der optionale Vorgabewert legt fest, welchen Wert dieses Attribut automatisch erhält, sollte es bei der Klasseninstanzierung nicht explizit angegeben sein.

### 7.1.3 Definition Attributtypen

Jedes Attribut besitzt eine eindeutige Typzuweisung. Dieser Attributtyp wird in der Tabelle der Attributtypen modelliert. Zur Modellierung komplexer Strukturen kann diese Tabelle um eigene Attributtypen erweitert werden. Es werden folgende Attributtypen unterschieden:

- Skalare Typen (integer, float,string,url)
- Referenzen auf Klasseninstanzen (Pointer)
- Felder (Enumerationen von skalaren Typen oder Referenzen oder anderen Attributtypen)

Ein Attributtyp ist festgelegt durch

- eine eindeutige ID (Primärschlüssel)
- Name (Bsp. Integer, Float, etc.)
- Eigenschaft Skalar oder Referenz (Bei Referenz ist zusätzlich die Id der referenzierten Klasse angegeben)
- Eigenschaft Enumeration (d. h. das Attribut ist ein Array). Es ist dann zusätzlich die Id des Attributtyps, der in einem Feldelement verwendet wird, angegeben.
- Optional können Masseinheiten (m,Pa,l,etc.) und Umrechnungsfaktoren angegeben werden.

### 7.1.4 Klasseninstanzen

Durch die Entitäten Klasse, Attribut und Attributtyp lässt sich ein annähernd beliebiges Datenmodell entwickeln. Dieses Datenmodell steht nun als Grundlage für die Instanzierung in Klasseninstanzen zur Verfügung. Eine Klasseninstanz sind festgelegt durch

- eine eindeutige ID (Primärschlüssel)
- Name der Klasseninstanz
- Die Id der instanziierten Klasse
- Die optionale Angabe einer Referenzklasseninstanz (Id), aus der evtl. Default-Werte für Attributinstanzen ermittelt werden.

## 8 Zusammenfassung

Die hier vorgestellte internetbasierte Plattform für Remote-Anwendungen besteht aus den drei Grundkomponenten Internet-Webserver, Data-Server und ORBIT-Server, unterstützt durch den Workflow-Editor für den individuellen Prozessentwurf. Diese Teilkomponenten sind so konzipiert, dass sie sowohl als Einzelanwendungen eingesetzt werden, als auch im Komponentenverbund miteinander agieren können. Für kleinere Anwendungen ist der Einsatz einzelner Komponenten durchaus sinnvoll. So können beispielsweise Remote-Tools vergleichsweise einfach durch den ORBIT-Server aufgerufen werden. Der Data-Server kann ebenfalls losgelöst vom Gesamtsystem als permanenter Modell- und Parameterspeicher verwendet werden. Der eigentliche Mehrwert der Plattform kommt jedoch erst im kooperativen Zusammenspiel der Einzelkomponenten zum Tragen: Die anfallenden Daten und Informationen werden zwischen den Teilsystemen ausgetauscht, ausgewertet und weiterverarbeitet. Der Anwender erhält also neben den eigentlichen Toolein- und ausgabedaten zusätzliche Informationen über Historie, Anwendungsverläufe, Durchführungprotokolle, etc. Das ORBIT-Gesamtsystem bietet somit eine solide Grundlage für eine professionelle Plattform im Bereich der Remote-Anwendungen.

## Literatur

- [1] Simple Object Access Protocol (SOAP) 1.1 <http://www.w3.org/TR/SOAP/>
- [2] Workflow Management Coalition. „Workflow Management Application Programming Interface Specification,“ WfMC-TC-1009, July-1998, [http://www.aiim.org/wfmc/standards/docs/if2\\_v20.pdf](http://www.aiim.org/wfmc/standards/docs/if2_v20.pdf).
- [3] Workflow Management Coalition. „The Workflow Reference Model,“ WfMC-TC-1003, Version 1.1, Jan. 1995, <http://www.aiim.org/wfmc/standards/docs/tc03v11.pdf>.
- [4] Integration der OMID-Tools in das Workflow-System ORBIT, Einbindung von

externen Programmen in ORBIT, InterConnect Software GmbH, November 2000

[5] „Datenmodellierung in ORBIT“, Beschreibung der Formate für Benutzerdaten im Workflow-System ORBIT, InterConnect Software GmbH, August 2001

[6] „Internetbasierte Designoptimierung mit HyGLEAM“, W. Jakob, D. Peters, A. Reiffer, OMID-Optimierung von Mikrosystemen für Diagnose- und Überwachungsanwendungen (Abschlußseminar des BMBF-Verbundprojekts OMID), FZK, 2002.