

acatech DISKUTIERT

> PRODUKTIONSFAKTOR MATHEMATIK

WIE MATHEMATIK TECHNIK UND WIRTSCHAFT BEWEGT

MARTIN GRÖTSCHEL, KLAUS LUCAS UND VOLKER MEHRMANN
(Hrsg.)

acatech DISKUTIERT

> PRODUKTIONSFAKTOR MATHEMATIK

WIE MATHEMATIK TECHNIK UND WIRTSCHAFT BEWEGT

MARTIN GRÖTSCHEL, KLAUS LUCAS UND VOLKER MEHRMANN
(Hrsg.)

> INHALT

> Vorwort	15
> Einführung Martin Grötschel, Klaus Lucas und Volker Mehrmann	17

PROZESSE

> Vorausschauend planen, gezielt handeln – Über die Regelung und Steuerung technischer Prozesse Lars Grüne, Sebastian Sager, Frank Allgöwer, Hans Georg Bock und Moritz Diehl	27
1 Executive Summary	27
2 Eine lange Erfolgsgeschichte	28
2.1 Das Stabilitätskriterium von Hurwitz	28
2.2 Das Pontrjaginsche Maximumprinzip	30
2.3 Fazit	31
3 Status quo und aktuelle Entwicklungen: Das Beispiel „Modellprädiktive Regelung“	32
3.1 Eine kurze Einführung in die Steuerungs- und Regelungstechnik	32
3.2 Prinzip der modellprädiktiven Regelung	36
3.3 Stabilität von MPC	39
3.4 Anwendungsbeispiele	40
3.5 Direkte Optimalsteuerungsmethoden	42
4 Herausforderungen	48
4.1 Modellierung	49
4.2 Robustheit von Lösungen, Unsicherheiten	51
4.3 Weitere Herausforderungen	53
5 Visionen und Handlungsempfehlungen	54
5.1 Theorie und Praxis	55
5.2 Interdisziplinarität in der Ausbildung	55
6 Literatur	56

>	Datenkompression, Prozessoptimierung, Aerodynamik:	
	Eine Tour durch die Skalen	
	Wolfgang Dahmen und Wolfgang Marquardt	63
1	Einige Leitgedanken	63
2	Erfolgsgeschichten	66
2.1	Informationsgewinnung aus riesigen Datensätzen	66
2.2	Adaptiv optimiert	67
2.3	Aerodynamik unter dem mathematischen Mikroskop	71
3	Status quo	74
3.1	Multiskalen-Zerlegungen: Wavelets	74
3.2	Multiskalenverfahren zur Datenanalyse	79
3.3	Lösung von Optimalsteuerungsproblemen mit adaptiver Waveletdiskretisierung	82
3.4	Adaptive Verfahren für partielle Differentialgleichungen der Strömungsmechanik	86
4	Stärken/Schwächen-Analyse	93
5	Visionen und Handlungsempfehlungen	93
6	Literatur	94
>	Aktive Strömungsbeeinflussung -	
	Eine mathematische Herausforderung	
	Rudibert King, Volker Mehrmann und Wolfgang Nitsche	99
1	Executive Summary	99
2	Erfolgsgeschichten	100
3	Aktive Beeinflussung von Strömungen, Status quo	100
3.1	Modellierung	104
4	Stärken/Schwächen-Analyse, Herausforderungen	106
5	Visionen und Handlungsempfehlungen	107
6	Literatur	108
>	Data-Mining für die Angebotsoptimierung im Handel	
	Jochen Garcke, Michael Griebel und Michael Thess	111
1	Executive Summary	111
2	Angebotsoptimierung im Handel: Übersicht und Status quo	112
2.1	Kampagnenoptimierung	113
2.2	Cross- und Up-Selling	115
3	Ausblick	120
4	Visionen und Handlungsempfehlungen	122
5	Literatur	123

NETZE

>	Planungsprobleme im öffentlichen Verkehr	
	Ralf Borndörfer, Martin Grötschel und Ulrich Jaeger	127
1	Executive Summary	127
2	Erfolgsgeschichten	128
3	ÖV-Planungsprobleme: Übersicht und Status quo	131
3.1	Betriebsplanung	132
3.2	Steuerung des Betriebs	137
3.3	Angebotsplanung	138
3.4	Wettbewerbsbedingungen	140
4	Stärken/Schwächen-Analyse, Herausforderungen	141
4.1	Rahmenbedingungen für den Einsatz von Mathematik	141
4.2	Mathematische Modelle und Algorithmen	145
4.3	Transfer und Ausbildung	148
4.4	Fazit	149
5	Visionen und Handlungsempfehlungen	150
5.1	Diskrete Optimalsteuerung: Umplanung von Verkehrssystemen bei Störungen in Echtzeit	150
5.2	Modellintegration: Angebotsplanung im Bus- und Bahnbereich	152
6	Literatur	153
>	Mit Mathematik zu mehr Intelligenz in der Logistik	
	Rolf H. Möhring und Michael Schenk	157
1	Executive Summary	157
2	Erfolgsbeispiele	159
2.1	Logistikketten	159
2.2	Transport	160
2.3	Produktionslogistik	161
2.4	Steuerung logistischer Netze	164
3	Logistik und Mathematik: Status quo	167
3.1	Zur Entwicklung der Logistik	167
3.2	Mathematik in der Logistik	168
4	Zukünftige Herausforderungen	170
5	Visionen und Handlungsempfehlungen	171
6	Literatur	172

>	Optimierung von Kommunikationsnetzen	
	Jörg Eberspächer, Moritz Kiese und Roland Wessäly	175
1	Executive Summary	175
2	Ob Telefonnetz oder Internet: Mathematik ist immer dabei!	176
2.1	Routing im Internet	176
2.2	Qualitätskontrolle im Telefonnetz und im Internet	178
2.3	Kostenoptimierte Kommunikationsnetze	179
3	Netze, Planung und Methoden	179
3.1	Netzarchitektur und Planungsaspekte	179
3.2	Strategische Planung	181
3.3	Taktische Planung	182
3.4	Operative Planung	185
3.5	Stand der Methoden: Fazit	186
4	Einflussfaktoren und Herausforderungen	187
4.1	Markteinflüsse, Wettbewerbsbedingungen, Regulierung	187
4.2	Datenverfügbarkeit	187
4.3	Konvergenz und Komplexität der Systeme	188
4.4	Planungsprozesse und Planungswerkzeuge	188
4.5	Zusammenarbeit von Industrie und Wissenschaft, Ingenieuren und Mathematikern	189
4.6	Ausgewählte Herausforderungen	190
5	Was zu tun ist - Vorschläge und Initiativen	193
5.1	Ausbildung	193
5.2	Wirtschaft	193
5.3	Forschungsschwerpunkt	194
6	Literatur	194
>	Mathematik in der drahtlosen Kommunikation	
	Holger Boche und Andreas Eisenblätter	197
1	Executive Summary	197
2	Erfolgsgeschichten	197
2.1	Mobilkommunikation braucht Mathematik	198
2.2	Das Fundament der Kommunikationstheorie	199
2.3	Schlüsseltechniken	201
2.4	Optimierung von GSM-Netzen	202
3	Wissen schafft Fortschritt	204
3.1	Optimierte Ressourcenvergabe in zellularen Funknetzen	204
3.2	Steuerung von Sendeleistungen	207
3.3	Axiomatik für interferenzbeschränkte Funkssysteme	209
3.4	Kapazitätsplanung für UMTS-Funknetze	209
3.5	Hardwareentwicklung	211

4	Zusammenarbeit stärken	213
4.1	Interdisziplinäre Zusammenarbeit	213
4.2	Internationale Zusammenarbeit	214
4.3	Transfer	214
4.4	Forschung von Nachwuchstalenten fördern	215
5	Perspektiven	215
5.1	Informationstheorie und digitale Welt	215
5.2	Netzwerkinformationstheorie	216
5.3	Quanteninformationstheorie	216
5.4	Neue Ansätze der Frequenznutzung	217
5.5	Selbstorganisierende Netze	218
6	Literatur	218
>	Mathematik im Chip-Design	
	Jürgen Koehl, Bernhard Korte und Jens Vygen	223
1	Executive Summary	223
2	Erfolgsgeschichten	226
3	Status quo des Chip-Designs	237
3.1	Platzierung	241
3.2	Timing Optimierung	243
3.3	Routing	249
4	Stärken/Schwächen-Analyse, Herausforderungen	253
5	Visionen und Handlungsempfehlungen	254
6	Literatur	255

MATERIALIEN UND MECHANIK

>	Chancen und Visionen der modernen Mechanik	
	Wolfgang Ehlers und Peter Wriggers	259
1	Executive Summary	259
2	Erfolgsgeschichten	260
2.1	Kontinuumsmechanik von Mehrkomponenten-Mehrphasenmaterialien	260
2.2	Simulationstechniken zur Beschreibung von heterogenen Materialien (Multi-scale Modelling)	264
3	Übersicht und Status quo	272
3.1	Fakten	272
3.2	Zusammenfassung	277

4	Stärken/Schwächen-Analyse, Herausforderungen	277
4.1	Ausgangsbasis und Visionen	277
4.2	Wie kommt die Mathematik ins Spiel?	279
4.3	Stärken, Schwächen und Herausforderungen	281
5	Handlungsempfehlungen	282
6	Literatur	282
>	Mathematik für Werkzeugmaschinen und Fabrikautomatisierung	
	Berend Denkena, Dietmar Hömberg und Eckart Uhlmann	285
1	Executive Summary	285
2	Erfolgsgeschichten	286
2.1	Mathematik für die Entwicklung neuer Maschinenkonzepte: Parallelkinematiken	286
2.2	Mathematik für die Auslegung von Werkzeugmaschinen: Zerspanprozesse	286
3	Mathematische Konzepte in der Produktionstechnik	289
3.1	Modellierung von Prozessketten	289
3.2	Adaptive Numerik	290
3.3	Optimalsteuerung	294
4	Herausforderungen	296
4.1	Wechselwirkung zwischen Struktur und Prozess	296
4.2	Wechselwirkungen beim Fräsen	298
4.3	Rekonfiguration von Produktionsanlagen	299
5	Perspektiven	301
6	Literatur	303
>	Herstellung und Einsatz moderner Materialien	
	Wolfgang Dreyer	307
1	Executive Summary	307
2	Erfolgsgeschichten	308
3	Status quo	318
4	Stärken/Schwächen-Analyse, Herausforderungen	320
5	Visionen und Handlungsempfehlungen	321
6	Literatur	321

>	Topologie und Dynamische Netzwerke: Anwendungen der Optimierung mit Zukunft	
	Günter Leugering, Alexander Martin und Michael Stingl	323
1	Executive Summary	323
2	Von den Eigenschaften zu optimalen Strukturen	324
3	Erfolgsgeschichte: Struktur-, Topologie- und Materialoptimierung	328
3.1	Fazit/Empfehlungen	330
4	Optimierung von Transportnetzen	331
4.1	Fazit/Empfehlungen	335
5	Danksagung	337
6	Literatur	337

ENERGIE UND BAUEN

>	Ressourceneinsatzplanung in der Energie- und Verfahrenstechnik	
	Sebastian Engell, Edmund Handschin, Christian Rehtanz und Rüdiger Schultz	341
1	Executive Summary	341
2	Erfolgsgeschichten	343
2.1	Elektrische Energieversorgung	343
2.2	Ressourceneinsatzplanung in der verfahrenstechnischen Produktion	346
3	Status quo	350
3.1	Energieversorgung	350
3.2	Verfahrenstechnische Produktionsprozesse	354
3.3	Mathematische Methoden	357
4	Perspektiven und Herausforderungen	359
4.1	Energieversorgung	359
4.2	Verfahrenstechnische Produktionsprozesse	364
4.3	Mathematische Methoden	365
5	Visionen und Handlungsempfehlungen	367
6	Literatur	369

>	Simulationsbasierte Optimierung in der Bautechnik – Neue informatische Lösungskonzepte	
	Dietrich Hartmann, Matthias Baitsch und Van Vinh Nguyen	375
1	Executive Summary	375
2	Erfolgsgeschichten	377
3	Optimierung von Nichtstandardproblemen mit Multiagentensystemen (MAS): Status quo	387
3.1	Agentensysteme	388
3.2	Strategienverbund	389
4	Stärken/Schwächen-Analyse, Herausforderungen	393
4.1	Praxisgegebenheiten	394
4.2	Neue Optimierungsstrategien/ Algorithmisierung	395
4.3	Erweiterung der mathematischen Grundlagen für Ingenieure	395
5	Visionen und Handlungsempfehlungen	396
6	Literatur	397
>	Objektorientierte Modellierung zur Simulation und Regelung von Energiewandlungsprozessen	
	Dirk Abel	401
1	Einführung und Übersicht	401
2	Anwendung im Kraftwerksbereich: Projekt OXYCOAL-AC	402
2.1	Projektziele und -erfordernisse	402
2.2	Objektorientierte Modellierung von Kraftwerksprozessen	403
2.3	Objektorientierte Modellierung des OXYCOAL-Kraftwerks	405
2.4	Ergebnisse	408
2.5	Fazit	411
3	Anwendung im Bereich der Verbrennungsmotoren	411
3.1	Modellbasierte Regelungen im Verbrennungsmotor	411
3.2	Aufbau der Regelstrecke	412
3.3	Vergleich zwischen objektorientierter und signalorientierter Modellierung am Beispiel von Luftpfadkomponenten	414
3.4	Fazit	419
4	Resümee	420
5	Literatur	420

>	Planungswerkzeuge für energieeffiziente Gebäude	
	Dirk Müller	423
1	Executive Summary	423
2	Entwicklung eines neuen Kühlsystems mit Hilfe einer thermohydraulischen Gebäudesimulation	424
2.1	Einsatz der objektorientierten Programmiersprache Modelica	425
2.2	Ausbau eines komplexen Simulationsmodells	425
2.3	Ergebnisse der Berechnungen	428
3	Stärken/Schwächen-Analyse, Herausforderungen	430
4	Visionen und Handlungsempfehlungen	431
5	Literatur	431

MEDIZIN

>	Mehr Mathematik wagen in der Medizin	
	Peter Deuffhard, Olaf Dössel, Alfred K. Louis und Stefan Zachow	435
1	Executive Summary	435
2	Mathematik in der medizinischen Bildgebung	435
2.1	Geschichte eines Erfolges	435
2.2	Mathematik als Innovationsfaktor	439
2.3	Perspektive: Neue Bildgebungsmethoden	439
3	Mathematik in der Kardiologie und der Herzchirurgie	443
3.1	Erfolgsgeschichten: EKG und Biosignalverarbeitung	443
3.2	Mathematik als Innovationsfaktor	444
3.3	Perspektive: das virtuelle Herz	445
4	Mathematik in der Therapie- und Operationsplanung	449
4.1	Erfolgsgeschichte: MKG-Chirurgie	449
4.2	Mathematik als Innovationsfaktor	450
4.3	Perspektive: der virtuelle Patient	456
5	Visionen und Handlungsempfehlungen	456
6	Literatur	457

>	Wirkstoffe, Medikamente und mathematische Bildverarbeitung	
	Günter J. Bauer, Dirk A. Lorenz, Peter Maass, Hartwig Preckel	
	und Dennis Trede	461
1	Executive Summary: Vom Bild zum Medikament	461
2	Erfolgsgeschichten	463
2.1	Automatische Analyse von Zellbildern im High-Content- High-Throughput-Screening	465
2.2	Statistische Auswertung der Daten	467
3	Status Quo: Mathematische Methoden für die High-Content-High-Throughput-Analyse	469
3.1	Parameteroptimierung und Inverse Probleme	469
3.2	Verbesserte Zellsegmentierung	472
4	Stärken/Schwächen-Analyse: Das Dilemma disziplinären Denkens	473
5	Visionen und Handlungsempfehlungen	474
5.1	Konkrete Herausforderungen	474
5.2	Kooperation	475
6	Literatur	475
>	Autorenverzeichnis	479

> VORWORT

Die klassischen Produktionsfaktoren der Volkswirtschaftslehre sind *Arbeit, Kapital* und *Boden*. Auf den ersten Blick erscheint es daher vielleicht provokativ, auch die *Mathematik* als Produktionsfaktor zu bezeichnen. Aber diese Provokation hat das Ziel, die nach wie vor geringe Sichtbarkeit der Leistungen der Mathematik zu erhöhen und die Bedeutung der Mathematik für Wirtschaft und Technik in den Fokus zu rücken. Dieser Band will hierzu einen Beitrag leisten.

Die Betriebswirtschaftslehre differenziert Produktionsfaktoren detaillierter und fügt u. a. den *dispositiven Faktor* als weiteren Produktionsfaktor hinzu. Dieser hat für eine möglichst gute Kombination der übrigen Faktoren zu sorgen. Ziel aller Produktion ist, den Nutzen der Beteiligten zu maximieren, wobei Ressourcenbeschränkungen, technische Gegebenheiten, physikalische und chemische Gesetze, Budgetgrenzen und natürlich auch staatliche Vorgaben als Nebenbedingungen zu berücksichtigen sind.

Die Begriffe *Maximierung, möglichst gute Kombination, Berücksichtigung von Nebenbedingungen* deuten auf mathematische Ansatzpunkte hin. Aber der Einfluss der Mathematik ist keineswegs allein auf Optimierungsfragen beschränkt. Mathematik dringt tief in das Design von Produkten, die Gestaltung von Produktionsprozessen und Wertschöpfungsketten ein und trägt zum grundsätzlichen Verständnis von Sachverhalten bei. Sie liefert die Sprache zur abstrakten Beschreibung naturwissenschaftlicher, technischer und wirtschaftlicher Gegebenheiten und Vorgänge. Sie ermöglicht die Modellierung, Simulation und Optimierung von Produkten und Prozessen, bevor überhaupt mit deren Realisierung begonnen wird. Mathematik erweist sich nicht nur als Produktionsfaktor zur besseren Herstellung von Waren und Dienstleistungen, sondern auch als Schlüsselwissenschaft zur Beherrschung komplexer Technologien und als Basiswissenschaft für Innovationen. Ein Ziel dieses Buches ist, diese sehr allgemein formulierten Thesen durch die Schilderung konkreter Fälle substantiell zu belegen.

Dieser Band, herausgegeben zum Jahr der Mathematik 2008 von acatech, fokussiert insbesondere auf das Zusammenspiel von Mathematik und Ingenieurwissenschaften. Wie kann Mathematik zur Verbesserung von technischen Produkten und Prozessen beitragen? Wo tut sie es bereits? Wo gibt es Defizite? Was kann man in Zukunft erwarten? Was kann aus einer engeren Zusammenarbeit zwischen Ingenieuren und Mathematikern entstehen?

Mit diesen Fragen beschäftigt sich der vorliegende Band in 19 Artikeln zu besonders exemplarischen Themenfeldern, in der Mehrzahl geschrieben von Autorenteams aus Ingenieuren, Praktikern aus der Industrie und Mathematikern. Diese Mischung hat

den Vorzug, dass nicht eine einzige fachspezifische Sichtweise dominiert. Alle Autoren wurden gebeten, für ein breites Publikum zu schreiben. Meist ist das sehr gut gelungen, in einigen wenigen Fällen wird, um die konkreten Anwendungsfragen oder die mathematischen Beiträge erläutern zu können, etwas mehr Fachwissen vorausgesetzt.

Der Einsatz von Mathematik ist je nach Branche unterschiedlich tiefgehend. In manchen ist ingenieurwissenschaftlicher Fortschritt direkt mit mathematischer Spitzenforschung verknüpft. In anderen dagegen werden selbst jahrzehntelang erfolgreiche mathematische Instrumente nicht eingesetzt. Warum das so ist, wird in einigen Artikeln erläutert. Die Gründe hierfür sind vielfältig. Sprach- und Terminologiebarrieren spielen eine Rolle, auch Ausbildungsdefizite bei Ingenieuren und Mathematikern. Sogar ordnungspolitische Rahmenbedingungen (z. B. Monopolsituationen), Hierarchien in Firmen oder die Psychologie der Entscheidungsfindung führen gelegentlich dazu, dass Mathematik nicht angemessen genutzt wird. Alle Autoren sind sich darüber einig, dass eine adäquate Beherrschung von Mathematik ein klarer Wettbewerbsvorteil ist.

Dieser Band ist dank der bunten Autorenmischung ein authentischer Bericht über die Wechselwirkungen zwischen Mathematik und Technikwissenschaften. Er gibt auch unterschiedlichen Auffassungen Raum. So findet man Artikel, die sich in ihren Aussagen zur Herangehensweise an ingenieurwissenschaftliche Probleme widersprechen. Mathematiker etwa fordern tiefer gehende Theorieentwicklung, Ingenieure dagegen haben manchmal mit hemdsärmeligen Heuristiken eher Erfolg in der Praxis und stellen die Notwendigkeit der Theorieverfeinerung infrage. Deutliche branchenspezifische Unterschiede sind auch hier erkennbar und geben besonders klare Signale zu intensiverem Diskurs!

Deutschland hat für einen vertieften Einsatz von Mathematik eine hervorragende Ausgangssituation. Die Technikwissenschaften sind in vielen Bereichen Weltspitze, auch die Mathematik in Deutschland hat Weltgeltung, insbesondere in den Bereichen der Mathematik, die schon seit Jahren auf Anwender zugegangen sind und sich ihrer Probleme angenommen haben. In vielen anderen Ländern orientiert man sich an dieser positiven Entwicklung und erkennt die wissenschaftlichen und wirtschaftlichen Vorteile, die sich aus einer engen Kooperation der Mathematik mit der Industrie ergeben. Dennoch bleibt viel zu tun, und dies wird in diesem Band auch deutlich. In einigen Artikeln werden konkrete Empfehlungen gegeben, wie die Zusammenarbeit zwischen Mathematik und Technikwissenschaften verbessert werden kann.

Deutschland ist ein Hochlohnland, das hauptsächlich von seinen guten Ideen und deren Umsetzung lebt. Ein großes Plus der deutschen Industrie ist das Verständnis, die Planung und Steuerung von Systemen. Wenn wir in Deutschland unseren Lebensstandard halten wollen, müssen wir mit unseren Fähigkeiten zur Entwicklung neuer Technologien, diffiziler Produkte, schwieriger Produktionsprozesse und der Beherrschung komplexer Systeme weltweit führend sein. Ohne Mathematik ist das heute nicht mehr möglich.

Berlin, November 2008

Martin Grötschel, Klaus Lucas, Volker Mehrmann

> EINFÜHRUNG

MARTIN GRÖTSCHEL, KLAUS LUCAS UND VOLKER MEHRMANN

Der vorliegende Band ist der zentrale Beitrag von acatech – Deutsche Akademie der Technikwissenschaften zum Jahr der Mathematik, das 2008 deutschlandweit auf die Bedeutung und die vielfältigen Facetten dieser Wissenschaft hinweist. Dieses Buch diskutiert das Verhältnis von Mathematik und Ingenieurwissenschaften, es beleuchtet insbesondere den Beitrag der Mathematik zur industriellen Wertschöpfung und die Schlüsselstellung der Mathematik bei der Beherrschung komplexer Systeme, mit anderen Worten: die Rolle der Mathematik als Produktionsfaktor und Innovationsverstärker. Der Band gibt nicht nur einen Zustandsbericht, er weist auf Stärken und Schwächen hin, skizziert zukünftige Forschungs- und Entwicklungsthemen und regt Maßnahmen zur Verbesserung der Interaktion zwischen Mathematik und Technikwissenschaften sowie zwischen Forschung und Industrie an.

Es ist klar, dass sich ein solches Werk nicht mit allen Bereichen der Technikwissenschaften und allen Aspekten mathematischer Anwendungen beschäftigen kann. Die Herausgeber haben gleichwohl eine große Bandbreite unterschiedlicher Themenfelder ausgewählt, an denen sich das Spannungsfeld und die positiven Wechselwirkungen zwischen den Gebieten sehr gut deutlich machen lassen. So ist ein Buch entstanden, kohärent und heterogen zugleich, von dem die Herausgeber und Autoren hoffen, dass ihre Analysen und Anregungen weit über die hier konkret angesprochenen Fachgebiete hinauswirken.

Die *Kohärenz* des aus 19 Artikeln bestehenden Bandes liegt in der gemeinsamen Gliederung aller Beiträge. Für jeweils ein ausgewähltes technikwissenschaftliches Fachgebiet wird die Beziehung dieses Gebietes zur Mathematik beleuchtet. Es wird untersucht, inwieweit Mathematik in dieses Fach produktiv hineinwirkt. Hierbei werden wissenschaftliche Aspekte angesprochen, die Mathematik wird aber auch in ihrer Rolle als Produktionsfaktor im zugehörigen Industriezweig analysiert. Jeder Artikel beginnt mit einem *Executive Summary*, das auf rund einer Seite die Erkenntnisse des Aufsatzes zusammenfasst. Mit zwei bis vier kurzen Erfolgsgeschichten zur positiven Wechselwirkung zwischen dem Fachgebiet und der Mathematik wird jeder Artikel eingeleitet. Es folgt eine Übersicht über den Einfluss der Mathematik auf das betreffende Fachgebiet und eine Beschreibung des Status quo. Aus einer anschließenden Stärken/Schwächen-Analyse werden die für das Fachgebiet und die damit verbundenen mathematischen Disziplinen bestehenden besonderen Herausforderungen abgeleitet. Dies betrifft so-

wohl die Rahmenbedingungen für Kooperationen zwischen Ingenieuren und Mathematikern, den Transfer in die Industrie als auch die Ausbildung und Fortbildung an Hochschulen. Visionen möglicher Weiterentwicklungen und Handlungsempfehlungen für die beteiligten Akteure, Industriepartner und Förderorganisationen runden das Bild ab.

Die *Heterogenität* resultiert aus der unterschiedlichen Mischung der Autorentams und aus ihrer fachlichen Orientierung. Viele Artikel sind von Mathematikern, Ingenieuren und Praktikern gemeinsam geschrieben worden, in manchen dominieren Ingenieure, in manchen Mathematiker. Auch das technikwissenschaftliche Fachgebiet spielt hierbei eine Rolle. So ist die Herangehensweise eines Gebäudetechnikers eine ganz andere als die eines Verfahrenstechnikers, Nachrichtentechnikers oder Logistikers. So werden Interaktions- und Spannungsfelder sowohl innerhalb der Technikwissenschaften als auch zwischen Mathematik und Technikwissenschaften sichtbar, die eine positive Entwicklung des Zusammenspiels erwarten lassen. Viele Gespräche haben begonnen, sie werden sich hoffentlich fortsetzen und weit über den Kreis der Autoren des vorliegenden Buches hinausragen.

Die Gliederung der 19 Artikel des Bandes in die Themenbereiche

- Prozesse,
- Netze,
- Materialien und Mechanik,
- Energie und Bauen,
- Medizin

ist sehr grob. Manche Artikel müssten eigentlich mehreren Themenbereichen zugeordnet werden, weil nicht immer nur ein einziges, klar definiertes technikwissenschaftliches Fachgebiet angesprochen wird. In einigen Aufsätzen werden die Wirkungen einer mathematischen Methodik, die sich in einem Anwendungsfall als besonders geeignet erwies, auch in anderen Feldern beschrieben, siehe z. B. die Artikel Engell et al. und Leugering et al. Diese Beiträge zeigen die universelle Einsatzfähigkeit der Mathematik auf hervorragende Weise. Mathematik fördert nicht selten den Transfer von Erkenntnissen zwischen wissenschaftlichen Fachgebieten und industriellen Anwendungen, die inhaltlich kaum etwas miteinander zu tun haben.

In dieser Einführung wollen wir besonders auf die unterschiedliche Interpretation und Verwendung einiger Begriffe hinweisen, um die gegenseitige Verständigung zu befördern. Besonderer Klärungsbedarf besteht in Bezug auf die Begriffe *Modellierung*, *Simulation* und *Optimierung*. Diese kommen in fast allen Beiträgen vor, werden aber nicht immer in der gleichen Bedeutung benutzt.

Mathematische Modelle haben die Aufgabe, reale Sachverhalte oder physische Prozesse in der Sprache der Mathematik zu beschreiben. Sie werden aus *mathematischen Bausteinen* errichtet. Die Modellierung beginnt damit, dass man die Mengen festlegt,

die zur Beschreibung des Lösungsraumes benötigt werden. In manchen Fällen dürfen Parameter, Variablen und andere Größen beliebige reelle Zahlenwerte annehmen, gelegentlich müssen komplexe Zahlen benutzt werden, andere Werte müssen ganzzahlig sein, usw. Es werden Variablen definiert, mit deren Hilfe man die Realität beschreiben will. Sie operieren auf den ausgewählten Grundmengen. Die Beziehungen zwischen den Variablen werden durch Gleichungen und Ungleichungen dargestellt, bei Optimierungsproblemen kommen noch Zielfunktionen hinzu, die es zu maximieren oder minimieren gilt.

Jedem, der sich mit Modellierung beschäftigt, ist klar, dass ein mathematisches Modell niemals ein exaktes Abbild der Natur sein kann. Die *Kunst der Modellierung* besteht darin, die für das jeweilige Ziel der Untersuchung wichtigen Parameter und Beziehungen herauszufiltern, diese möglichst gut zu beschreiben und Unwichtiges wegzulassen. Daraus folgt automatisch, dass Modellierung allein durch Mathematiker und ohne fachliche Begleitung nicht funktionieren kann. Gute Modellierung spielt sich in einem Interaktionsprozess zwischen mehreren beteiligten Gruppen ab. Neben Ingenieuren und Mathematikern können hier Informatiker, Biologen, Chemiker, Physiker, Betriebswirte aber auch Juristen beteiligt sein. Am Ende eines Modellierungsprozesses sollte ein mathematisches Modell stehen, von dem alle Beteiligten überzeugt sind, dass es die „Wirklichkeit des betrachteten Problems“ einigermaßen genau wiedergibt.

Modellierung unterliegt auch anderen Einschränkungen. Man hat vielleicht ein perfektes Modell gefunden, kennt aber keine Methoden, die die Gleichungen des Modells lösen. Modelle müssen dann vereinfacht werden, um sie „rechenbar“ zu machen, siehe z. B. King et al. Es kann auch sein, dass die für ein gut passendes Modell notwendigen Daten nicht beschafft werden können. Dann muss das Modell so reduziert werden, dass dieses nur auf verfügbare Daten zurückgreift. Manchem mathematischen Theoretiker sträuben sich dabei die Haare, Ingenieure haben eher gelernt, mit der Unzulänglichkeit der realen Welt umzugehen.

Am Ende eines solchen Modellierungsprozesses steht dann ein mathematisches Modell für die vorliegende Fragestellung. Hier zeigt sich häufig bereits ein Unterschied zwischen Mathematik und Technikwissenschaften. Der Mathematiker erwartet ein Modell in Form einer analytischen Darstellung, also z. B. beschrieben durch lineare Ungleichungen, gewöhnliche oder partielle Differentialgleichungen. Das kann ein Ingenieur nicht immer liefern. Sein Modell ist u. U. in Form eines Computerprogramms formuliert, das möglicherweise viele if-then-else-Abfragen enthält. Ein solches Modellierungsergebnis lässt Mathematiker in der Regel verzweifeln, denn sie können komplexe Programme dieser Art nur sehr selten mit mathematischen Werkzeugen verarbeiten. Dies bedeutet auch, dass *prozedurale Modelle* (gegeben durch Codes) mit anderen Methoden bearbeitet werden müssen, als sie in der Mathematik üblich sind. Beispiele hierfür sind z. B. in der Energie- und Gebäudetechnik zu finden, siehe z. B. Hartmann et al. Ähnliches gilt, wenn ein System in viele Teilkomponenten zerlegt werden kann, siehe z. B. Abel oder Müller, für jede Teilkomponente zwar angemessene Modelle vorhanden sind, das Ge-

samtmodell aber außerhalb der (derzeitigen) Reichweite der mathematischen Analyse ist.

Dennoch, die intensive Zusammenarbeit von Mathematikern und Technikwissenschaftlern bei der Behandlung ganz konkreter praktischer Fragestellungen hat in den letzten Jahren zu enormen Fortschritten geführt. Dies gilt u. a. für die Entwicklung und Analyse von neuen Materialien, siehe z. B. Dreyer sowie Ehlers und Wriggers, bei denen auf der Basis von mathematischen Modellgleichungen versucht wird, das Materialverhalten vorherzubestimmen. Ähnliche Erfolge sind auch bei der Entwicklung von Produktionsprozessen zu beobachten, siehe z. B. Denkena et al. Aber nicht nur das Verständnis der physikalisch-technischen Prozesse wurde besser, auch die damit verbundene Mathematik wurde weiterentwickelt, neue mathematisch spannende Fragestellungen konnten herausdestilliert werden, und daraus konnten neue Lösungsalgorithmen entwickelt oder bekannte Verfahren extrem beschleunigt werden, siehe z. B. Borndörfer et al. Ohne den engen Kontakt zur Praxis wären viele dieser Fragen durch Mathematiker überhaupt nicht aufgegriffen worden, andererseits hat der Kontakt zur Theorie Ingenieuren vielfach neue und effizientere mathematische Werkzeuge verfügbar gemacht.

Häufig sind Modelle nicht nur für die Fragestellung geeignet, für die sie entworfen wurden. Die entwickelte Theorie und die zugehörigen Lösungsalgorithmen können nicht selten mit geringem Aufwand auf ganz andere Anwendungsfälle übertragen werden. Sichtbar wird dies zum Beispiel im Themenbereich Netze, wo Methoden der ganzzahligen Optimierung für Telekommunikationsnetze entwickelt wurden und dann in der Logistik und im Verkehr eingesetzt werden – und umgekehrt (Borndörfer et al., Möhring und Schenk, Eberspächer et al., Boche und Eisenblätter). Gleiches gilt für mathematische Methoden zur Behandlung von Mehrskaligkeit, die bei der Bildkompression und -kodierung ebenso zum Einsatz kommen wie in der Prozessindustrie und der Aerodynamik, siehe z. B. Dahmen und Marquardt.

Nach der Modellierung ist der nächste Schritt im mathematischen Ansatz die *Simulation*. Dies ist nichts anderes als die mehrfache Durchrechnung eines mathematischen Modells unter Variation von Parametern. Durch Simulation überprüft man zunächst ein Modell. Sind die Ergebnisse eines Simulationslaufs so wie erwartet? Treten Überraschungen auf? Liegen unerwartete Ergebnisse an Modellierungsfehlern oder an fehlerhaften Daten? Oder hat man wirklich neues Verhalten entdeckt? Durch Simulation verschafft man sich eine gewisse Sicherheit bzgl. der Korrektheit der Modellbildung. Einen Beweis für die Korrektheit eines Modells liefert numerische Simulation zwar nicht, sie unterstützt aber den Prozess des „allmählichen Verstehens“ von Prozessen auf sehr fruchtbare Weise. Die Simulation hat sich daher in den letzten Jahren neben dem Experiment und der Theorie als eine immer wichtiger werdende dritte Säule des Erkenntnisgewinns etabliert.

Falls Simulationsläufe gut und schnell funktionieren, macht man häufig umfangreiche Simulationsstudien, die durchaus Hinweise auf gute Parameterkombinationen für die reale Durchführung eines Prozesses (z. B. in der chemischen Verfahrenstechnik oder beim Bau eines Kraftwerkes) geben können. In der Praxis wird gelegentlich geglaubt,

dass mit Simulationsrechnungen optimiert werden kann. Dem widersprechen Mathematiker heftig.

Mit der *Optimierung* kann erst dann sinnvollerweise begonnen werden, wenn man von der Angemessenheit eines Modells überzeugt ist und sich dies z. B. in Simulationsstudien erwiesen hat. Praktiker sprechen häufig bereits von Optimierung, wenn eigentlich nur Verbesserung einer Situation (z. B. mit heuristischen Verfahren und Simulationstechniken) betrieben wird. Mathematiker sehen das enger. Sie möchten Modell-Lösungen finden, die beweisbar optimal sind. Falls das nicht erreicht werden kann, möchten sie zumindest Gütegarantien abgeben etwa der Form, dass der Zielfunktionswert einer gefundenen Lösung beweisbar nur um 5 % vom Wert einer Optimallösung abweicht. Hier ist ein weites Feld der Interpretation, denn in vielen Fällen, gerade bei extrem großen Anwendungsbeispielen, müssen auch Mathematiker auf Approximationsmethoden zurückgreifen; und sie bedienen sich dann auch der Heuristiken, die Ingenieure erfolgreich im praktischen Einsatz anwenden.

Mit Optimierung allein ist es jedoch nicht getan. Viele praktische Fragestellungen haben keine eindeutige Zielfunktion. Dies gilt z. B. dann, wenn man ökonomische (z. B. Gewinnmaximierung), ökologische (geringer negativer Einfluss auf die Natur) und soziale (hoher Beschäftigungsgrad) Ziele miteinander verbinden will. Man spricht dann von *multikriterieller Optimierung*. Dies ist ein Gebiet, dessen Forschungsstand noch nicht befriedigend ist und bei dem in der Praxis fast nur mit heuristischen Methoden gearbeitet wird.

Eine weitere Frage schließt sich an. Wie steht es um die *Stabilität* oder *Robustheit* von berechneten Optimallösungen? Führen leichte Parameteränderungen zu ganz anderen Lösungen? In allen Anwendungsfeldern treten Fragen dieser Art in immer stärkerem Maße auf. Überall ist man beim Entwurf von Systemen oder Prozessen gezwungen, potentielle Störungen bereits in der Designphase so zu berücksichtigen, dass sie im laufenden Betrieb „auf vernünftige Weise“ kompensiert werden können.

Zufällige und überraschende Veränderungen, Erwartungswerte und Risikoabschätzungen sind natürlich heiße Themen der Finanzmathematik, stochastische Einflüsse sind jedoch von genauso großer Bedeutung in der Medizin, der Biotechnologie und bei der Steuerung von Anlagen. Störungen können dabei von außen kommen (Wetter, Unfälle), aber auch Datenfehler (fehlerhafte Messinstrumente) oder falsche Modellannahmen können zur Unsicherheit beitragen. Bei der Auslegung von Verkehrs- und Logistiksystemen, siehe Borndörfer et al. sowie Möhring und Schenk, ist man z. B. daran interessiert, Fahrpläne und Fahrzeugumläufe so einzurichten, dass lokale Verkehrsstörungen und -staus sich nicht überregional verbreiten und die entstehenden Verzögerungen schnellstmöglich mit geringen Auswirkungen für die Kunden ausgeglichen werden können. Die Stochastik beschäftigt sich mit solchen Fragen und beginnt, in verschiedenen Anwendungsfällen derartige Dinge in den Griff zu bekommen. Trotz theoretischer Fortschritte bleibt jedoch Simulation weiterhin ein sehr wichtiges Werkzeug zur Untersuchung von Robustheit, Stabilität, etc.

Mit diesem Fragenkomplex berühren wir insbesondere das Gebiet *Regelung und Steuerung*, das im Themenbereich Prozesse intensiv behandelt wird (siehe z. B. Grüne et al., King et al. und Engell et al.). Regelung beschreibt dabei den fortlaufenden rückgekoppelten Eingriff in das dynamische Verhalten von Prozessen und Systemen mit dem Ziel, Veränderungen so zu bewirken, dass die erforderlichen Ressourcen weitestgehend geschont werden und der Ablauf stabil ist. Bei der Steuerung werden im Gegensatz dazu a priori Eingangsparameter so festgelegt, dass ohne weiteren Eingriff ein gewünschtes Systemverhalten erzielt wird. Es gibt kaum ein Teilgebiet der Mathematik, das zum Thema Regelung und Steuerung keinen Beitrag geleistet hat. Diese Methoden spielen auch eine zentrale Rolle in der modernen Nachrichtentechnik, siehe z. B. Boche und Eisenblätter.

So gut wie alles, was bisher skizziert wurde, ist ohne die Existenz leistungsfähiger Computer und Informationstechnik nicht denkbar. Schnelle und zuverlässige Rechner haben die in diesem Band beschriebene Entwicklung erst ermöglicht. Die Mathematik hat hierzu nicht nur durch den Entwurf und die Implementierung effizienter Codes für die verschiedensten Probleme beigetragen, auch die Entwicklung von Computerchips selbst ist ohne mathematische Methoden nicht mehr möglich, siehe z. B. Koehl et al. Aber auch in der wirtschaftlichen Datenanalyse und Aufbereitung von Massendaten sind mathematische Methoden heutzutage zentral, siehe z. B. Garcke et al.

Neben den beschriebenen Erfolgen und Perspektiven des Zusammenspiels von Mathematik und Technikwissenschaften sind insbesondere die Medizin, die Medizintechnik und die Medikamentenentwicklung Forschungsgebiete, die zunehmend auf mathematische Methoden angewiesen sind, siehe z. B. Deuffhard et al. und Bauer et al.

Die Zukunft von Forschung und Entwicklung liegt in fast allen diskutierten Bereichen in der integrierten Modellierung, Simulation und Optimierung. Diese kann nur durch interdisziplinäre Zusammenarbeit von Ingenieuren und Mathematikern (und anderen Beteiligten) erreicht werden. Dazu ist eine größere inhaltliche und sprachliche Annäherung der verschiedenen Fachdisziplinen notwendig. Eine wesentliche Handlungsempfehlung, die sich in fast allen Artikeln findet, ist daher die Reform der Ausbildungsprogramme von Ingenieuren und Mathematikern. Dies betrifft u. a. die fehlenden Komponenten zu diskreter Mathematik und Optimierung in Ingenieurstudiengängen, aber auch die sehr marginale Behandlung mathematischer Modellierung in mathematischen Studiengängen ist zu beklagen.

Eine andere zentrale Handlungsempfehlung, die sich durch den Band zieht, ist der Wunsch nach weiteren Förderprogrammen, insbesondere des Bundesministeriums für Bildung und Forschung (BMBF). Das seit einigen Jahren geförderte BMBF-Programm „Mathematik für Innovationen in Industrie und Dienstleistungen“ (inzwischen leider mit stark reduziertem Umfang) war und ist fraglos ein effektiver Katalysator für die bisherige Entwicklung, viele der Autoren dieses Bandes haben es erfolgreich genutzt. Der Bedarf und das vorhandene Potenzial sind jedoch so immens, dass eine Intensivierung der Förderung erfolgen sollte. Nur so kann eine Beschleunigung des in diesem Band skizzierten

möglichen wissenschaftlichen und technischen Fortschritts erfolgen, der ganz besonders durch eine engere Zusammenarbeit von Mathematik, Technikwissenschaften und Industrie erzielt werden kann.

In der 2007 erschienenen BMBF-Broschüre *Der Schlüssel zur Hochtechnologie, Mathematik für Innovationen in Industrie und Dienstleistungen* schreibt Dr. Annette Schavan, die Bundesministerin für Bildung und Forschung, in ihrem Grußwort: „Aber auch im Hinblick auf ganz konkrete wirtschaftliche und gesellschaftliche Herausforderungen kommt der Mathematik eine zentrale Rolle zu. So ermöglicht sie Innovationen im Industrie- und Dienstleistungssektor, die für zusätzliche Arbeitsplätze und eine wachsende Wettbewerbsfähigkeit Deutschlands sorgen“, und sie fährt fort: „So begleitet uns eine alte Wissenschaft in die Zukunft, mit einem Potenzial, das noch lange nicht erschöpft ist.“ Der Analyse der Ministerin schließen sich die Herausgeber dieses Bandes gerne an. Sie verbinden dies mit der Hoffnung, dass die Industrie, die Politik und die Förderorganisationen daraus die richtigen Konsequenzen ziehen. Die Mathematik kann einen noch viel größeren Beitrag zur Produktentwicklung, Produktionssicherheit, Wertschöpfung, Ressourcenschonung, etc. leisten, als sie es jetzt schon (leider vielfach beinahe unbemerkt) tut. Dieser Band gibt Zeugnis davon und regt gleichzeitig geeignete Maßnahmen zur Verbesserung der Situation an.

PROZESSE

> VORAUSSCHAUEND PLANEN, GEZIELT HANDELN – ÜBER DIE REGELUNG UND STEUERUNG TECHNISCHER PROZESSE

LARS GRÜNE, SEBASTIAN SAGER, FRANK ALLGÖWER, HANS GEORG BOCK
UND MORITZ DIEHL

1 EXECUTIVE SUMMARY

Seit Beginn der industriellen Revolution nimmt die Steuerungs- und Regelungstechnik eine Schlüsselstellung in vielen technischen Bereichen ein. James Watts Fliehkraftregler für Dampfmaschinen ist eines der frühen Beispiele eines extrem erfolgreichen Reglerkonzepts, von dem Ende der 1860er Jahre geschätzte 75 000 Exemplare allein in England im Einsatz waren [2, S. 24]. Etwa um diese Zeit begannen Ingenieure, motiviert durch die immer höhere Komplexität der zu regelnden Maschinen, sich systematisch mit theoretischen Grundlagen der Regelung zu beschäftigen. Dies führte unausweichlich zu der Einsicht, dass das dynamische Verhalten der geregelten Systeme nur mit Hilfe der Mathematik verstanden und weiterentwickelt werden konnte, oder wie Werner von Siemens, ein weiterer technischer Pionier in diesem Bereich es formulierte: „Ohne Mathematik tappt man doch immer im Dunkeln.“

In der Steuerungs- und Regelungstechnik stellt sich daher – heute ebenso wie vor mehr als einhundert Jahren – nicht die Frage, *ob* man Mathematik benutzen muss, sondern vielmehr, *welche* Mathematik man sinnvollerweise benutzen sollte. Tatsächlich gibt es von der Algebra über die Geometrie und die Theorie der Dynamischen Systeme bis hin zur Optimierung und Numerik kaum ein mathematisches Gebiet, das nicht bedeutende Beiträge zur Regelungs- und Steuerungstechnik und ihren mathematischen Nachbardisziplinen System- und Kontrolltheorie geleistet hätte.

In diesem Artikel wollen wir zum einen anhand zweier Beispiele darstellen, welche Beiträge die Mathematik in der langen Geschichte der Steuerungs- und Regelungstechnik geliefert hat und dabei aufzeigen, welche Faktoren grundlegend für den Erfolg dieser Methoden waren. Zum anderen wollen wir – neben einem Überblick über den Status Quo in der Steuerungs- und Regelungstechnik – eine moderne Regelungsmethode, die *Modellprädiktive Regelung* (kurz: MPC, vom englischen „model predictive control“), genauer vorstellen, die dahinterstehenden mathematischen Konzepte aus der Systemtheorie, der Numerik und der Optimierung kurz erläutern und darauf aufbauend einige der zukünftigen Herausforderungen an die Mathematik skizzieren.

In ihrer linearen Variante ist MPC speziell in der Verfahrenstechnik bereits heute ein Standardwerkzeug in industriellen Prozessen, in ihrer nichtlinearen Variante (NMPC), auf die wir hier besonders eingehen werden, gilt sie allgemein als eine der vielversprechendsten modernen Methoden zur Regelung komplexer technischer Prozesse, die sich durch die rasanten Fortschritte sowohl auf theoretischer als auch auf algorithmischer Seite zunehmend neue Anwendungsgebiete erschließt. Zudem ist sie ein Paradebeispiel für eine Methode, die nur durch die interdisziplinäre Zusammenarbeit auf den Gebieten Regelungstechnik, mathematische Systemtheorie, Numerik und Optimierung ihren heutigen Entwicklungsstand erreichen konnte und auch nur durch diese Zusammenarbeit weiter entwickelt werden kann. Diese Interdisziplinarität, die bereits jetzt erfreulicherweise an einer Reihe von Universitäten durch Exzellenzcluster, Forschungszentren und Graduiertenschulen in Forschung und Lehre gefördert wird, erscheint uns als ein wesentlicher Faktor, um die vielfältig vorhandenen Chancen des Produktionsfaktors Mathematik im Bereich der Steuerungs- und Regelungstechnik auch in Zukunft nutzen zu können.

2 EINE LANGE ERFOLGSGESCHICHTE

Mathematische Methoden spielen in der Regelungstechnik seit Beginn dieses Gebietes eine wesentliche Rolle. Zwei prominente Beispiele, die wir in diesem Abschnitt kurz erläutern, sollen dies belegen. Anhand dieser Beispiele – das Stabilitätskriterium von Hurwitz und das Pontrjaginsche Maximumprinzip – werden wir prototypisch aufzeigen, welche „Erfolgsfaktoren“ diese beiden mathematischen Entwicklungen so erfolgreich gemacht haben.

2.1 DAS STABILITÄTSKRITERIUM VON HURWITZ

Das Stabilitätskriterium von Hurwitz wurde 1893–1894 von dem am Züricher Polytechnikum (heute ETH Zürich) als Professor tätigen Mathematiker Adolf Hurwitz entwickelt [30]. Anlass für die Entwicklung war aber weniger Hurwitz' mathematische Neugier als eine konkrete Anfrage seines Züricher Kollegen Aurel Stodola, eines Maschinenbauingenieurs, der sich mit der Entwicklung von Regelungen hydraulischer Turbinen beschäftigte. Stodolas Problem lässt sich am einfachsten anhand eines alltäglichen Problems veranschaulichen. Stellt man sich eine Heizungsregelung vor, bei der die Raumtemperatur auf einen gewünschten Wert eingestellt werden soll, so ist eine naheliegende Strategie, das Ventil für den Warmwasserdurchfluss im Heizkörper abhängig von der aktuell gemessenen Temperatur einzustellen: Ist die Temperatur geringer als die gewünschte, wird man das Ventil aufdrehen, ist sie höher, wird man es zudrehen. Man benötigt eigentlich keine Mathematik, um es einleuchtend zu finden, dass dieses Vorgehen tatsächlich die gewünschte Temperatur liefern wird.

Allerdings gilt das nur im Idealfall. Komplizierter wird die Sache, wenn der Einfluss der Stellgröße (im Beispiel das Durchflussventil) auf den Wert (also die Temperatur)

weniger direkt wird, z. B. durch Verzögerungen im System. Wer schon einmal unter einer Dusche gestanden ist, deren Warmwasserzufuhr nur verzögert auf das Öffnen oder Schließen der Ventile reagiert, kennt diesen Effekt: Statt der gewünschten Temperatur erhält man ständig zwischen „zu heiß“ und „zu kalt“ schwankende Werte – ein klassisches Beispiel für ein instabiles Verhalten eines Regelkreises – und nur mit viel Mühe und Fingerspitzengefühl gelingt es manchmal, am Ende doch die richtige Temperatur einzustellen. Auch wenn dies nur eine Analogie ist, ähnlich sind die Effekte bei den mechanischen Turbinensystemen, die Stodola betrachtete, dort hervorgerufen durch die Vielzahl mechanischer Kopplungen, die zu berücksichtigen sind. Auch hier ist es kompliziert, eine Regelung zu entwerfen, die das System stabil auf einem vorgegebenen Wert hält, es also stabilisiert.

Stodola kannte und verwendete nun aber einige zu der Zeit knapp 20 Jahre alte mathematische Resultate [62], die es erlaubten, aus einem Modell des Regelkreises eine Bedingung für die Stabilität abzulesen. Diese Bedingung lautete, dass ein Polynom, also eine mathematische Funktion der Form

$$P(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$$

nur Nullstellen (also komplexe Zahlen x mit $P(x) = 0$) mit negativem Realteil besitzt. Wenn solch ein Polynom nur wenige Summanden besitzt, so ist diese Bedingung leicht nachzuprüfen. Solche einfachen Polynome gehören allerdings auch zu sehr einfachen Regelungssystemen, bei denen diese Theorie gar nicht nötig ist, um die Stabilität sicher zu stellen, da ein erfahrener Ingenieur dies den Systemen sowieso „ansieht“. Wirklich relevant für die Praxis sind Polynome mit vielen Termen und das Hauptproblem ist dann, dass man die Nullstellen – aus prinzipiellen algebraischen Gründen, die seit dem frühen 19. Jahrhundert durch die Arbeiten von Évariste Galois bekannt sind – nicht einfach ausrechnen kann.

Hurwitz beschäftigte sich also auf Anregung von Stodola eine Weile mit diesem Problem und fand schließlich ein Kriterium, mit dem sich ermitteln ließ, ob die Nullstellen negative Realteile besitzen, ohne dass man diese tatsächlich ausrechnen muss. Mathematisch ausgedrückt bestand seine Lösung darin, die Koeffizienten a_0, a_1, \dots, a_n des Polynoms in n verschiedene Matrizen unterschiedlicher Größe zusammenzufassen. Dann genügt es, lediglich die Determinanten dieser Matrizen zu berechnen, um die Frage nach dem Vorzeichen der Realteile der Eigenwerte zu beantworten.

Stodola war von dieser Lösung begeistert. In einem Brief an Hurwitz schrieb er: „Ich bin Ihnen außerordentlich verpflichtet für die geniale Lösung des Wurzelproblems, das mich so sehr plagte.“ [5]. Aber auch Hurwitz erwähnte in einer Fußnote zu seiner zugehörigen Veröffentlichung in den Mathematischen Annalen nicht ganz ohne Stolz die erfolgreichen praktischen Anwendungen seines Kriteriums: „Herr Stodola benutzt mein Resultat in seiner Abhandlung [...], deren Ergebnisse bei der Turbinenanlage des Badesortes Davos mit glänzendem Erfolge Anwendung gefunden haben.“ Bemerkenswert ist in diesem Zusammenhang noch die Tatsache, dass der englische Mathematiker Edward

John Routh bereits knapp 20 Jahre vor Hurwitz eine ähnliche Lösung entwickelte, weswegen Hurwitz' Kriterium heutzutage meist als Methode von Routh-Hurwitz bezeichnet wird. Den Weg in die Anwendungen hingegen hat Hurwitz' Resultat früher gefunden, weil Rouths Ergebnis Jahrzehnte lang nur einem kleinen akademischen Zirkel aber kaum einem in der Praxis arbeitenden Ingenieur bekannt war [2, S. 81f.].

2.2 DAS PONTRJAGINSCHES MAXIMUMPRINZIP

Das Pontrjaginsche Maximumprinzip hat trotz seines auf den russischen Mathematiker Lew S. Pontrjagin zurückgehenden Namens mehrere Urheber. Zunächst einmal hat Pontrjagin das im Jahre 1956 veröffentlichte Prinzip nicht alleine, sondern mit wesentlichen Beiträgen der oft nicht genannten Mathematiker V. G. Boltjansky und R. L. Gamkrelidze veröffentlicht und bewiesen [8] (siehe auch [51]). Zum anderen, vgl. [49, S. 55f.], findet sich das Prinzip bereits in Veröffentlichungen des amerikanischen Mathematikers M. R. Hestenes aus dem Jahre 1950 und – sogar in verallgemeinerter, allerdings nicht vollständig bewiesener Form – des ebenfalls amerikanischen Mathematikers R. P. Isaacs aus dem Jahre 1954/55. Wichtige grundlegende Ideen zum Maximumprinzip findet man sogar noch früher in Carathéodorys Buch von 1935, wie in [50] ausgeführt wird.

Beim Maximumprinzip geht es um die Frage, wie man die Bewegung eines Systems – etwa die Flugbahn einer Rakete oder die Bewegung des Arms eines Industrieroboters – optimal vorausplant. „Optimal“ ist dabei immer auf ein vorher zu definierendes Kriterium mit zugehörigen Nebenbedingungen bezogen, also zum Beispiel: wie kommt eine Rakete mit minimaler Energie (Kriterium) in einer vorgegebenen Zeit (Nebenbedingung) von der Erde in eine spezielle Umlaufbahn, oder wie kann man den Arm eines Industrieroboters schnellstmöglich (Kriterium) bei konstantem Energieaufwand (Nebenbedingung) von einer bestimmten Position in eine andere steuern? Das Beispiel der Rakete ist dabei nicht zufällig gewählt, weil die Raumfahrt ebenso wie die militärische Raketentechnik im beginnenden kalten Krieg zwischen der UdSSR und den USA einer der Hauptgründe für die Entwicklung dieser Art von Mathematik war. Die für die industrielle Entwicklung sicherlich ungleich bedeutendere Anwendung auf Industrieroboter kann daher – ähnlich wie die berühmte Teflon-Pfanne – durchaus als ein Abfallprodukt der Weltraumforschung angesehen werden.

Optimierungsprobleme ähnlicher Art wurden bereits etwa 250 Jahre vor Pontrjagin und den anderen Entwicklern des Maximumprinzips mit Hilfe der sogenannten Variationsrechnung gelöst. Die vollständige Darstellung dieser Methode überschreitet sicherlich den Umfang dieses Beitrags, zumindest eine kurze Beschreibung der prinzipiellen Funktionsweise soll aber hier gegeben werden (für eine detailliertere Darstellung siehe z. B. [49]). Ausgangspunkt für den Lösungsansatz ist zunächst ein Modell, mit dessen Hilfe eine Bewegung durch mathematische Gleichungen beschrieben werden kann. Dies geschieht üblicherweise mit Hilfe eines Systems von Differentialgleichungen. Zur

Lösung des Problems muss nun für jeden Zeitpunkt die Position auf der optimalen Bahn berechnet werden. Das Problem dabei ist nun, dass die Wahl der Werte zu verschiedenen Zeitpunkten voneinander abhängt: Beschleunigt man z. B. am Anfang sehr stark, so muss man am Ende stärker bremsen und umgekehrt. Die Variationsrechnung löste diese komplexen Abhängigkeiten durch das Aufstellen eines erweiterten Differentialgleichungssystems (die Euler-Lagrange-Gleichungen), dessen erster Teil „normal“ vorwärts, dessen zweiter Teil aber rückwärts gelöst wird – ein sogenanntes Randwertproblem. Für jeden Zeitpunkt kann aus der Lösung dieses Systems dann die optimale Bahn ermittelt werden.

Das Hauptproblem der Anwendung der Variationsrechnung auf praktische Probleme besteht nun darin, dass dieses direkt die optimalen Bahnen berechnet, nicht aber die eigentlich relevanten optimalen Steuerwerte, die an die Raketentriebwerke oder die Roboterarmmotoren übermittelt werden müssen – und weil diese Werte in der Rechnung nicht auftauchen, ist es auch unmöglich, Wertebereiche vorzugeben, in denen diese Steuerwerte schon aus physikalischen Gründen liegen müssen. So kann die Variationsrechnung zwar durchaus optimale Flugbahnen ermitteln, es ist aber unmöglich auszuschließen, dass diese eine Schubkraft der Triebwerke benötigen, die weit jenseits der physikalischen Möglichkeiten liegt. Dies ist ganz offensichtlich eine Unzulänglichkeit, die nicht nur in der Raumfahrt, sondern auch in fast jeder denkbaren industriellen Anwendung ernsthafte praktische Probleme bereitet.

Hier bietet nun das Maximumprinzip Abhilfe: Da hier – bei konzeptionell durchaus ähnlicher Verwendung von Randwertproblemen – der Fokus nicht auf der Berechnung optimaler Bahnen, sondern auf der direkten Berechnung der zugehörigen optimalen Steuerstrategien liegt, ist es nun leicht möglich, physikalisch oder auch ökonomisch notwendige Beschränkungen direkt in die Rechnung einzubeziehen. Dabei liefert das Maximumprinzip ein durchaus auch intuitiv verständliches Kriterium, da die optimalen Steuerwerte in jedem Zeitpunkt nichts anderes als die Lösung eines neuen „kleinen“ Optimierungsproblems sind, dessen Parameter sich gerade aus dem gegebenen Randwertproblem ergeben. Da sich solche Probleme auch numerisch am Rechner gut lösen lassen, wird das Prinzip nicht nur für die theoretische Analyse, sondern auch als Basis für Algorithmen verwendet.

2.3 FAZIT

Die in diesem Abschnitt beschriebenen mathematischen Konzepte zur Analyse und Lösung von Steuerungs- und Regelungsprobleme sind nur zwei von vielen möglichen Beispielen. Sie besitzen aber eine Reihe von Eigenschaften, die in mehr oder weniger starker Ausprägung typisch für erfolgreiche mathematische Konzepte sind:

- Sie sind anwendbar, ohne dass man die zu ihrer Herleitung nötigen tieferliegenden mathematischen Theorien verstehen muss,

- sie haben für reale Anwendungen einen deutlichen Fortschritt gebracht und damit neue industrielle Entwicklungen erst ermöglicht,
- sie sind konstruktiv in dem Sinne, dass sie sich leicht als Algorithmus formulieren lassen und daher mit der Verfügbarkeit digitaler Rechner auch leicht implementiert werden konnten.

Bei all den offensichtlichen Vorteilen, die diese mathematischen Konzepte den industriellen Anwendungen gebracht haben, wären deren Entwicklungen aber trotzdem unmöglich gewesen, wenn bei den in den Steuerungs- und Regelungstechnik tätigen Ingenieuren nicht von vornherein ein Bewusstsein für die Notwendigkeit mathematischer Methoden – und damit einhergehend eine solide mathematische Grundbildung – vorhanden gewesen wäre. Ohne Stodolas Kenntnisse über die Bedeutung der Nullstellen wäre Hurwitz vermutlich nie auf die Idee gekommen, sein Kriterium zu entwickeln und ohne die Verfügbarkeit geeigneter Differentialgleichungsmodelle wäre die Entwicklung des Maximumprinzips bereits im Ansatz unmöglich gewesen.

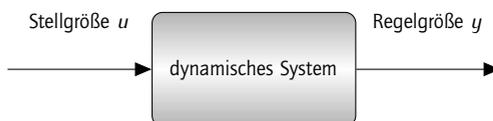
3 STATUS QUO UND AKTUELLE ENTWICKLUNGEN: DAS BEISPIEL „MODELLPRÄDIKTIVE REGELUNG“

3.1 EINE KURZE EINFÜHRUNG IN DIE STEUERUNGS- UND REGELUNGSTECHNIK

Das Ziel von Steuerungen und Regelungen ist die Beeinflussung dynamischer Systeme, so dass deren Verhalten gewünschte Eigenschaften besitzt. Die dynamischen Systeme werden dabei durch sogenannte *Stellgrößen* beeinflusst. Beim dynamischen System „Auto“ sind dies z. B. der Lenkradeinschlag oder die Stellung des Gaspedals. Durch geeignete Veränderung der Stellgröße über der Zeit kann dann die Systemdynamik in der gewünschten Weise beeinflusst werden. Das interessierende Verhalten des zu beeinflussenden Systems wird oft in der sogenannten *Regelgröße* zusammen gefasst. Abb. 1 zeigt schematisch eine solche *Regelstrecke*.

Man unterscheidet nun zwischen Steuerung und Regelung. Wie in Abb. 2 zu sehen ist wird bei einer *Steuerung* das dynamische Verhalten durch eine offene Wirkungskette

Abb. 1: Schematische Darstellung einer Regelstrecke mit Stellgröße u und Regelgröße y



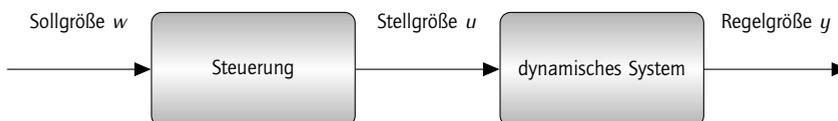
beeinflusst. Der Sollverlauf für die Regelgröße y wird durch die *Sollgröße* w vorgegeben, die den Eingang für die Steuerung bildet. Die Steuerung generiert aus diesem Signal den Verlauf der Stellgröße, die dann das dynamische System so beeinflusst, dass die Regelgröße dem Verlauf der Sollgröße möglichst exakt folgt.

Steuerungen sind in der industriellen Praxis sehr wichtig und erfüllen ihre Aufgabe zumeist sehr gut. Als Beispiele seien hier Robotersteuerungen oder Steuerungen für Werkzeugmaschinen und Fertigungseinrichtungen genannt. Grundlegende Voraussetzung für die korrekte Funktion einer Steuerung ist aber, dass das Verhalten des dynamischen Systems mit Hilfe des vorhandenen Modells hinreichend genau vorhergesagt werden kann; ist dies nicht der Fall, so kann das tatsächliche Verhalten weit von der zur Berechnung der Stellgröße u verwendeten theoretischen Vorhersage abweichen.

Dies kann z. B. dann passieren, wenn große externe Störungen auf das dynamische System wirken, wie z. B. starker Seitenwind bei Automobilen, oder wenn das Verhalten des dynamischen Systems nicht exakt bekannt ist, weil das mathematische Modell nicht alle Aspekte des realen Systems abbildet – in der Regelungstechnik spricht man in diesem Fall von *unsicheren Systemen*. Die Auswirkungen solcher Unsicherheiten sind immer dann besonders gravierend, wenn das Systemverhalten *instabil* ist, wenn also wie beim Beispiel der Dusche in Abschnitt 2.1 kleine Änderungen in der Stellgröße (Ventil) zu großen Änderungen in der Regelgröße (Wassertemperatur) führen. In diesen Fällen müssen *Regelungen* eingesetzt werden, um die Stabilität des Systems – also die Tatsache, dass die Regelgröße stabil in der Nähe des Sollwertes bleibt – zu erreichen.

Eine Regelung besitzt im Gegensatz zu einer Steuerung eine Rückführung (engl. feedback), wie in Abb. 3 gezeigt. Die Regelgröße wird gemessen, zurückgeführt und mit ihrem gewünschten Verlauf, also der Sollgröße, verglichen. Die Differenz zwischen Soll- und Istwert, auch Regelfehler genannt, wird nun dem Regler als Eingang zur Verfügung gestellt. Durch den geschlossenen Wirkungskreis kann der Einfluss von externen Störungen und Unsicherheiten durch die Regelung explizit erkannt und korrigiert werden, bevor diese zu großen Abweichungen vom Sollwert führen. Daher funktionieren Regelungen auch bei Vorhandensein von Störungen, Unsicherheiten und bei instabilen Regelstrecken.

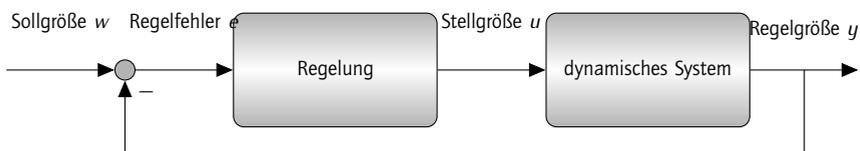
Abb. 2: Wirkprinzip der Steuerung



Natürlich führt die Rückführungsstruktur nicht automatisch zur Stabilität des geschlossenen Wirkungskreises: Entscheidend ist, nach welcher Vorschrift die Stellgröße aus dem Regelfehler berechnet wird. Wie in Abschnitt 2 bereits beispielhaft erläutert, kann das dynamische Verhalten eines geregelten oder gesteuerten mit den mathematischen Methoden der System- und Regelungstheorie untersucht werden. Die wichtigsten Analyseverfahren untersuchen dabei die Stabilität und die sogenannte *Robustheit der Stabilität* des geregelten oder gesteuerten Systems. Bei der Untersuchung der Robustheit der Stabilität wird die Frage beantwortet, ob die Stabilität erhalten bleibt, wenn sich das wirkliche dynamische System von den im Modell gemachten Annahmen unterscheidet. Diese Eigenschaft ist für den Einsatz der Methoden in der praktischen Anwendung von zentraler Bedeutung.

Von ganz entscheidender Relevanz ist die Mathematik beim *Entwurf* der Steuerungen und vor allem der Regelungen, d. h. bei der Herleitung der Vorschrift, mit der die Stellgröße aus der Sollgröße oder dem Regelfehler berechnet wird. Moderne Regler werden heute *modellbasiert* entworfen. Das heißt, es wird angenommen, dass ein mathematisches Modell des zu regelnden Systems, zumeist in Form von Differentialgleichungen, vorliegt. Weiterhin werden die Ziele der Regelung ebenfalls mathematisch formuliert. Das ist zum Einen die bereits mehrfach genannte Stabilität, aber auch die sogenannte *Regelgüte*, die zumeist in Form von Optimierungskriterien formuliert wird. Unterschiedliche Klassen von Reglerentwurfsverfahren unterscheiden sich, erstens, durch unterschiedliche Annahmen an die dem Entwurf zugrundeliegenden Modelle der Regelstrecken und, zweitens, durch unterschiedliche Zielvorgaben. In Bezug auf die Modellannahmen ist die wichtigste Unterscheidung diejenige in lineare bzw. nichtlineare Systeme. Lineare Systeme zeichnen sich – etwas vereinfacht gesagt – dadurch aus, dass der Zusammenhang zwischen Stellgröße u und Regelgröße y proportional ist, dass also zum Beispiel doppelt so große Stellgrößen genau doppelt so große Regelgrößen nach sich ziehen (mathematisch sagt man, sie erfüllen das Superpositionsprinzip). Sie werden durch lineare Diffe-

Abb. 3: Wirkprinzip einer Regelung mit Rückführung und Soll-/Istwert-Vergleich



rentialgleichungen, zumeist in der sogenannten Zustandsraumdarstellung, beschrieben

$$\dot{x} = Ax + Bu \quad (1a)$$

$$y = Cx + Du. \quad (1b)$$

Der Entwurf von Reglern für lineare Systeme ist sehr gut verstanden, und es stehen eine große Anzahl von äußerst leistungsfähigen linearen Reglerentwurfverfahren zur Verfügung. An dieser Stelle seien beispielhaft die LQ-Verfahren erwähnt (LQR, LQG, LQG/LTR, etc.) [23, 36, 42, 47], bei denen für lineare Systeme der Form (1) ein Regler entworfen wird, so dass der geschlossene Kreis optimales Verhalten in Bezug auf die Minimierung eines quadratischen Integralkriteriums

$$J = \int_0^{\infty} x^T(\tau)Qx(\tau) + u^T(\tau)Ru(\tau) d\tau$$

besitzt.

Bei den LQ-Verfahren werden keine externen Störungen, wie zum Beispiel der Seitenwind beim Auto, berücksichtigt. Beim moderneren H_{∞} -Reglerentwurf [24, 64] werden die Regler dagegen genau so entworfen, dass sie externe Störgrößen optimal unterdrücken, d. h. dass deren negative Auswirkungen auf den Regelkreis minimiert werden. Diese Minimierung kann über unterschiedliche mathematische Optimierungskriterien erfolgen und neben den H_{∞} -Verfahren können hier beispielhaft auch die L_1 -Methoden [14] genannt werden.

Auch wenn kaum ein in der industriellen Praxis auftretendes System tatsächlich linear ist, können lineare Modelle trotzdem oft eingesetzt werden. Der Grund dafür liegt darin, dass sich die allermeisten Systeme „fast“ wie lineare Systeme verhalten, wenn man nur einen kleinen Wertebereich für die Regelgröße betrachtet: Will man die Raumtemperatur von 20° auf 22° Celsius erhöhen, so muss man das Heizungsventil in etwa zweimal so weit öffnen, als wenn man die Temperatur nur auf 21° erhöhen möchte – der Zusammenhang zwischen Ventilöffnung und Temperaturzunahme ist linear. In größeren Temperaturbereichen gilt das aber nicht mehr, schon allein deswegen, weil sich das Ventil irgendwann nicht mehr weiter aufdrehen lässt und die Maximaltemperatur erreicht ist.

Wenn das zu regelnde System im relevanten Wertebereich kein lineares Verhalten aufweist, was in der Praxis häufig der Fall ist, müssen nichtlineare Reglerentwurfmethoden angewandt werden. Das dem Entwurf zugrunde liegende dynamische System wird hier durch nichtlineare Differentialgleichungen der Form

$$\dot{x} = f(x, u)$$

$$y = h(x, u).$$

beschrieben. Auf dem Gebiet der nichtlinearen Regelung wurden in den letzten ca. 20 Jahren sehr große Fortschritte gemacht und viele aktuelle Forschungsprojekte befassen

sich mit diesem Thema. Beispielhaft sollen hier die differentialgeometrischen Verfahren der exakten Linearisierung [33], die flachheitsbasierte Regelung [22, 58], die passivitätsbasierte Regelung [9, 55], oder die Backstepping-Methoden [37, 41] genannt sein. Diese Methoden liefern im einfachsten Fall¹ eine statische Funktion

$$u = k(x), \quad (3)$$

d. h. eine mathematische Formel $k(x)$, die abhängig vom aktuellen Systemzustand x angibt, wie die Stellgröße u gewählt werden muss.

Im Nichtlinearen steht häufig das Stabilisierungsproblem im Vordergrund. Selbstverständlich ist aber auch hier die Regelgüte von Interesse. Das nichtlineare Äquivalent zu den oben erwähnten LQ-Verfahren ist die optimale Regelung, bei der wieder ein Optimierungskriterium und Nebenbedingungen – wie bereits in Abschnitt 2.2 beschrieben – vorgegeben werden. Im Falle der nichtlinearen optimalen Regelung ist die Berechnung einer expliziten Formel (3) für u in der Praxis aber selbst mit Hochleistungscomputern oft unmöglich. Einfacher ist es hingegen, nichtlineare optimale Steuerungen (also Stellgrößen u in Abhängigkeit von der Zeit) mit dem Computer zu berechnen, z. B. mit Hilfe des Pontrjaginschen Maximumprinzips, vgl. Abschnitt 2.2 oder der moderneren direkten Methoden, vgl. Abschnitt 3.5.

Deshalb wurde in den letzten Jahren eine neue Klasse von Verfahren zur optimalen Regelung nichtlinearer Systeme entwickelt, bei denen keine explizite Formel der Form (3) berechnet wird, sondern bei denen das Rückführgesetz $k(x)$ online in Echtzeit aus optimalen Steuerungen berechnet wird. Obwohl solch eine online-Rechnung auf den ersten Blick wesentlich schwieriger aussieht, zeigt es sich, dass diese Vorgehensweise deutliche Vorteile hat und zur Lösung von großen nichtlinearen Regelproblemen in der Praxis geeignet ist. Der prominenteste Vertreter dieser neuen Klasse von Regelungsverfahren ist die modellprädiktive Regelung, die in den folgenden Abschnitten genauer betrachtet und diskutiert werden soll.

3.2 PRINZIP DER MODELLPRÄDIKTIVEN REGELUNG

Wenn man einen Regler entwirft, der z. B. die Temperatur eines Hauses auf einem möglichst konstanten Wert halten soll, so besteht die einfachste Idee darin, die aktuelle Temperatur zu messen und darauffolgend die Warmwasserdurchflussmenge passend zu erhöhen oder zu verringern – wie in Abschnitt 2.1 erläutert, kann man bei geschickter Wahl der Regelparameter dadurch ein stabiles geregeltes System erreichen, das mehr oder weniger gut die durch das Wetter verursachten äußeren Temperaturschwankungen auffängt. Da die Heizung nicht beliebig schnell Wärme nachliefern kann, und auch die

¹ Oft ist der Regler nicht durch eine statische Gleichung, sondern selbst wieder durch eine Differentialgleichung gegeben.

Abkühlung nicht beliebig schnell folgen kann, wird die Raumtemperatur dabei immer leicht schwanken.

Was wäre aber, wenn wir den Wetterbericht nutzen würden? Dann könnten wir, wenn eine Kaltfront naht, die Wohnung bereits etwas vorwärmen, oder, wenn warmes Wetter erwartet wird, bereits einige Stunden im Voraus die Heizung abstellen, so dass das Haus sich langsam abkühlen kann. Es ist intuitiv klar, dass die Verwendung dieses „prädiktiven“ Prinzips uns erlaubt, das Regelungsziel „möglichst kleine Abweichungen vom Sollwert“ unter Beachtung unserer begrenzten Heiz- oder Kühlleistung besser einzuhalten.

Wenn wir selbst Auto fahren, ist es sogar absolut unabdingbar, dass wir „vorausschauend“, also „prädiktiv“ fahren, denn würden wir nach schneller gerader Fahrt ohne vorheriges Abbremsen in eine Kurve rasen, flögen wir aus dieser. Dass wir das nicht tun, ist allein der fortgeschrittenen Regelungstechnik in unserem Gehirn zu verdanken, die die Kurvenfahrt bereits antizipiert, lange bevor wir die Fliehkraft in der Kurve am eigenen Leib spüren, siehe Abb. 4.

Die modellprädiktive Regelung (model predictive control, MPC) verwirklicht genau dieses Prinzip für die automatische Regelung technischer Systeme, vgl. z. B. [11,46]. Wir erläutern dies nun etwas detaillierter am Beispiel einer typischen Stabilisierungsaufgabe: der Systemzustand soll zu einem Sollwert geregelt und dort gehalten werden, indem der Zustand regelmäßig gemessen wird und gegebenenfalls festgestellte Abweichungen durch geeignete Anpassung der Stellgrößen ausgeglichen werden.

Ausgehend vom aktuellen Zustand x_0 des zu steuernden Systems und den aktuellen Vorhersagen über externe Einflüsse, berechnet ein modellprädiktiver Regler dazu eine (hier der Einfachheit halber stückweise konstante) optimale Steuerung für die nahe

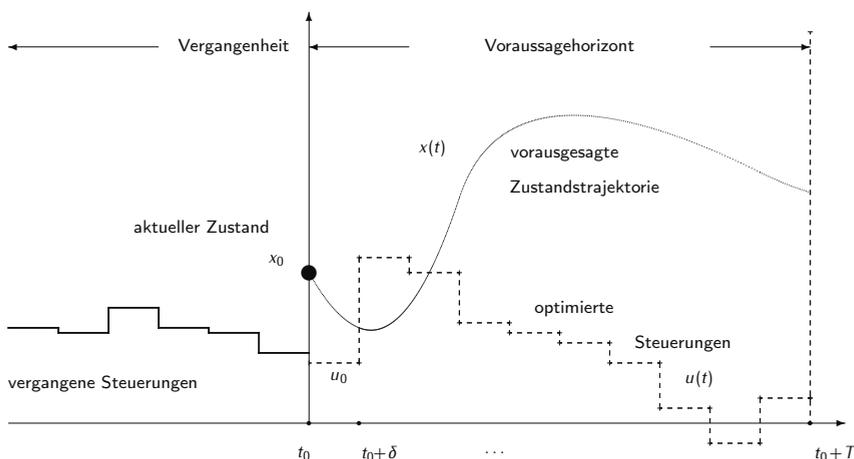
Abb. 4: Vorausschauendes Fahren – so, wie es sein soll ...



Zukunft, den „Voraussagehorizont“ der Länge T . Das dabei verwendete Optimalitätskriterium minimiert typischerweise den Abstand zum Sollwert, damit die berechnete optimale Lösung einen möglichst geringen Abstand von diesem hat. Von dieser Steuerung wird dann nur der erste Abschnitt u_0 an die reale Anlage gegeben, mit dem das System für eine kurze Zeit – die Abtastzeit δ – gesteuert wird. Nach Ablauf dieser kurzen Zeit wird der neue aktuelle Zustand ermittelt, der Vorhersagehorizont wird um δ nach vorne geschoben, und eine neue optimale Steuerung auf diesem Horizont wird berechnet, von dem wieder nur der erste kurze Teil verwirklicht wird. Auf diese Art und Weise werden die online berechneten optimalen Steuerungen zu einer Regelung „zusammengesetzt“. Abbildung 5 illustriert einen Schritt dieses Verfahrens.

Die zwei wesentlichen Vorteile der modellprädiktiven Regelungsmethode sind die Fähigkeit, vorausschauende Informationen zu berücksichtigen und Optimalitätskriterien und Beschränkungen einzubeziehen. Diese Vorteile kommen z. B. bei energieintensiven oder zeitkritischen Anwendungen zum Tragen, da Ziele wie „mit minimalem Energieaufwand“ oder „innerhalb einer gegebenen Zeit“ direkt in das Optimierungskriterium oder die Nebenbedingungen einbezogen werden können.

Abb. 5: Schematische Darstellung der modellprädiktiven Regelung



Die zwei wichtigsten Fragen sind:

- Wie kann Stabilität des mit MPC geregelten Systems garantiert werden?
- Wie können die Optimalsteuerungsprobleme auf dem Vorhersagehorizont schnell und zuverlässig gelöst werden?

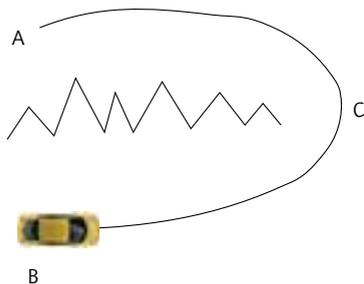
Beide Fragen sind zutiefst mathematischer Natur, die erste fällt in den Bereich der Systemtheorie, die zweite in den Bereich der Numerik. Wir wollen diese Fragen – sowie eine Vorstellung von Anwendungen von MPC – in den folgenden Abschnitten behandeln.

3.3 STABILITÄT VON MPC

Da wir exakte mathematische Definitionen vermeiden wollen, soll „Stabilität“ hier einfach die Tatsache bezeichnen, dass der Regelalgorithmus sein Ziel erfüllt, dass der Systemzustand also in den Sollwert geregelt wird. Intuitiv erscheint es plausibel, dass man durch Minimierung des Abstandes vom Sollwert eine in diesem Sinne stabile Regelstrategie erhält. Dass das aber durchaus schiefgehen kann, wollen wir an einem kleinen Beispiel veranschaulichen, vgl. Abb. 6. In dieser Abbildung soll die Position des Autos (Systemzustand, zu Beginn im Ort B) in den Ort A (Sollwert) gebracht werden, wobei die Geschwindigkeitsbeschränkungen auf der Straße eingehalten werden sollen. Da ein Gebirge im Weg ist, führt die Straße allerdings nicht auf direktem Weg von A nach B, sondern auf einem Umweg über den Ort C.

Nimmt man als Optimierungskriterium nun den Luftlinien-Abstand des Autos von Punkt A, so muss man diesen Abstand auf dem Weg der Straße von B nach A zunächst vergrößern, um ihn letztendlich verringern zu können. Eine MPC-Strategie, deren Vorhersagehorizont z. B. nur bis zum Punkt C in der Kurve reicht, wird dies nicht erkennen, denn jede Bewegung des Autos im Rahmen dieses Horizonts würde den Abstand ja nur ver-

Abb. 6: Beispiel eines Regelungsproblems



größern, weswegen es optimal ist, einfach im Punkt B stehen zu bleiben. Die geregelte Lösung wird für alle Zeiten im Punkt B verharren und nie im Punkt A ankommen: der Regelalgorithmus ist nicht stabil.

Dieses Problem ist natürlich seit langer Zeit bekannt und es wurden verschiedene Möglichkeiten zur Abhilfe vorgeschlagen. Ein in der Literatur oft betrachteter Ansatz ist es, in der Optimierung auf dem Horizont T nur solche Lösungen zu berücksichtigen, die im gewünschten Sollwert enden. Diese Methode funktioniert theoretisch (was mathematisch rigoros bewiesen wurde, vgl. [38]) und wird in der Praxis auch gelegentlich eingesetzt, besitzt aber den Nachteil, dass man eine weitere Beschränkung zu dem zu lösenden Optimierungsproblem hinzufügt und erfordert zudem i. A. einen sehr großen Horizont T um sicherzustellen, dass es unter den gegebenen Beschränkungen (also z. B. den Geschwindigkeitsbeschränkungen in unserem Straßenbeispiel) überhaupt möglich ist, bis zur Zeit T im Sollwert anzukommen. Beides kann die Lösung des Optimierungsproblems deutlich erschweren. Ein weiterer Ansatz ist die Verwendung von Korrekturtermen in der Optimierung, die den Effekt von zu kurzen Horizonten T ausgleichen, siehe z. B. [13]. Die Berechnung dieser Terme ist aber i. A. recht aufwendig, weswegen diese Methode in der Praxis eher selten angewendet wird.

Die in der Praxis zumeist verwendete Lösung ist die naheliegendste: Die Horizontlänge T wird einfach so lange erhöht, bis der Algorithmus stabil wird. Interessanterweise wird diese Lösung im Großteil der theoretischen MPC-Literatur *nicht* betrachtet und tatsächlich wurde für nichtlineare MPC-Methoden unter relativ allgemeinen Voraussetzungen erst vor kurzem mathematisch rigoros nachgewiesen, dass dies auch wirklich funktioniert [26, 34]. Nachteil dieser Methode ist, dass im vorhinein i. A. nicht klar ist, wie groß T denn nun gewählt werden muss – und je größer T ist, desto mehr Zeit benötigt die Lösung des Optimierungsproblems, die ja online während der Laufzeit des Systems durchgeführt werden muss. Aktuelle Forschungsansätze [27, 28] versuchen daher, den benötigten Horizont T aus den Systemeigenschaften abzuschätzen und dabei insbesondere zu ermitteln, wie das Optimalitätskriterium gewählt werden muss, damit man mit möglichst kleinem T Stabilität erhält. Über die Praktikabilität dieses Vorgehens für industriell relevante Prozesse kann derzeit nur spekuliert werden. Für unser Straßenbeispiel liefert dieser Ansatz allerdings eine ebenso natürliche wie effiziente Lösung, auf die der eine oder die andere unter den Leserinnen und Lesern vielleicht schon gekommen ist: Misst man den Abstand vom Ort A nicht über die Luftlinie, sondern über die Straßenlänge, so ist es bei beliebig kurzem Horizont T immer besser, sich auf A zuzubewegen, da der Abstand in jedem Fall verringert wird. Wird über diesen Abstand optimiert, so ist das Verfahren daher für beliebige Horizonte T stabil.

3.4 ANWENDUNGSBEISPIELE

Die Modellprädiktive Regelung (MPC) wurde ursprünglich in der Verfahrenstechnik entwickelt, wo sie z. B. zum Regeln großer und langsamer Destillationskolonnen verwendet

wird – diese Systeme sind extrem energieintensiv, und die langsame Zeitskala lässt viel Zeit für die Berechnung der optimalen Steuerungen. Die Verfahrenstechnik ist auch heute noch das Hauptanwendungsgebiet von MPC: Dittmar und Pfeiffer [20] haben 2005 in einer Erhebung 9456 MPC-Anwendungen ermittelt (gehen aber von einer deutlich höheren tatsächlichen Zahl aus), von denen mehr als 80% im Bereich der Verfahrenstechnik liegen. Schwerpunktmäßig wird MPC dabei bei der Regelung kontinuierlicher Reaktoren eingesetzt, bei denen der Zustand des Prozesses (Temperatur, Konzentration der Chemikalien ...) kontinuierlich auf einem Sollwert gehalten werden muss. Seltener wird MPC bei sogenannten Batch-Reaktoren verwendet, bei denen die Reaktion nicht kontinuierlich, sondern in einem vorgegebenen Zeitrahmen stattfindet und bei denen die Regelungsaufgabe darin besteht, einem vorausberechneten Verlauf der Prozesszustände zu folgen.

Wenngleich es Beispiele erfolgreicher Anwendungen nichtlinearer MPC (NMPC) in der Verfahrenstechnik gibt [52], werden hier überwiegend lineare MPC Verfahren eingesetzt, bei denen das zu Grunde liegende Differentialgleichungsmodell linear ist. Dies hat den großen Vorteil, dass die resultierenden Optimierungsprobleme linear sind, wodurch schnellere und zuverlässigere Lösungsverfahren zur Verfügung stehen. Zwar hat eine vergleichende Studie an einem industriellen Batch-Reaktor [48] ergeben, dass NMPC-Verfahren durchaus zu einer besseren Regelung führen, die Verbesserungen sind bisher aber noch nicht so ausgeprägt, dass sich die aufwendigere Implementierung des Verfahrens ökonomisch lohnen würde – es bleibt abzuwarten, ob sich dies mit knapper werden Rohstoffen und teurerer Energie in Zukunft ändern wird. Einen weiteren Hinderungsgrund für die Anwendung von NMPC stellt in vielen Fällen das Fehlen geeigneter nichtlinearer Modelle in der Verfahrenstechnik dar [52], ein Bereich, in dem in jüngster Zeit allerdings mit Hilfe geeigneter mathematischer Methoden deutliche Fortschritte erzielt wurden, siehe Abschnitt 4.1.

Durch effizientere Algorithmen und theoretische Fortschritte erschließen sich MPC und insbesondere NMPC-Methoden seit einiger Zeit immer weitere Anwendungsfelder außerhalb der Verfahrenstechnik. Beispiele sind die Regelung von Schleusen in Kanalsystemen, die Regelung von Heizungs- und Klimasystemen in großen Gebäuden, oder die optimale Regelung von saisonalen Wärmespeichern. Letztere nutzen die Wärme im Sommer, um den Grund unter einem Gebäude aufzuheizen und dabei gleichzeitig das Gebäude zu kühlen, und gebrauchen diese Wärme im Winter mit Hilfe einer Wärmepumpe wieder, um Heizkosten zu sparen. Nur die prädiktive Regelung erlaubt es hier, zu entscheiden, wann die Wärmepumpe verwendet werden soll und wann optimalerweise zugeheizt werden soll, basierend auf den saisonalen Witterungsvorhersagen. Zunehmend schnellere Optimierungsalgorithmen erlauben immer schnellere Anwendungen der modellprädiktiven Regelung. So ist heute das punktgenaue Regeln der Fahrt von Zügen möglich und experimentell getestet [25], ebenso wie die Regelung der Einspritz- und Luftmenge bei Automotoren im 50 Millisekudentakt [21]. Auch Roboterarme, die komplexe Manöver in möglichst kurzer Zeit ausführen sollen, und dies in ständig wechselnden

Abb. 7: Roboter an der Universität Leuven, an denen zur Zeit mit MPC im Millisekundentakt experimentiert wird.



der Umgebung, sollen demnächst mit MPC geregelt werden, siehe Abbildung 7. Eine der visionärsten Anwendungen der MPC ist die automatische Steuerung von Flugdrachen, die in Zukunft Windenergie aus großen Höhen liefern sollen und unter wechselnden Windbedingungen stabil auf Kreisbahnen am Himmel fliegen sollen [12, 32], siehe Abbildung 8. Eine weitere zur Zeit untersuchte Anwendung der MPC ist die automatische Steuerung der Insulinzugabe bei Diabetikern.

3.5 DIREKTE OPTIMALSTEUERUNGSMETHODEN

Wie weiter oben angekündigt, wollen wir nun einen kurzen Einblick in den aktuellen Stand der Forschung im Bereich der Methoden zur Lösung des rechenintensivsten Teilschritts der modellprädiktiven Regelung geben, der Lösung der einzelnen Optimalsteuerungsprobleme auf dem betrachteten Zeithorizont. Wir konzentrieren uns dabei auf den für die aktuelle Forschung ungleich relevanteren Fall der nichtlinearen Systeme.

In Abschnitt 2.2 wurde das Maximumprinzip angesprochen. Auf diesem mathematischen Theorem basierende, sogenannte *indirekte* Methoden sind eine Möglichkeit, optimale Lösungen für Steuerungsprobleme zu berechnen. Diese werden in der Tat auch heute noch eingesetzt. Dies gilt insbesondere für die Luft- und Raumfahrt, siehe beispielsweise [10], aber auch für die Analyse von Problemen in der Verfahrenstechnik,

Abb. 8: Automatisch betriebener Flugdrachen zur Energieerzeugung [12]



siehe [59]. Das Maximumprinzip nimmt auch nach wie vor eine wichtige Rolle in der Analyse von Eigenschaften von Optimalsteuerungsproblemen ein. Für die Konstruktion von Algorithmen, die am Computer umgesetzt werden können, haben diese indirekten Methoden allerdings einige Nachteile:

- der erhebliche analytische Aufwand schon bei kleinen Änderungen des Modells oder von Parametern,
- die komplizierte Behandlung von Pfad- und Steuerungsbeschränkungen, die zu zustandsabhängigen Sprüngen in den differentiellen Variablen führen,
- die häufig numerisch sehr schwierig zu lösenden Randwertprobleme mit extrem kleinen Konvergenzbereich, die nur durch den Einsatz von Homotopien lösbar sind.

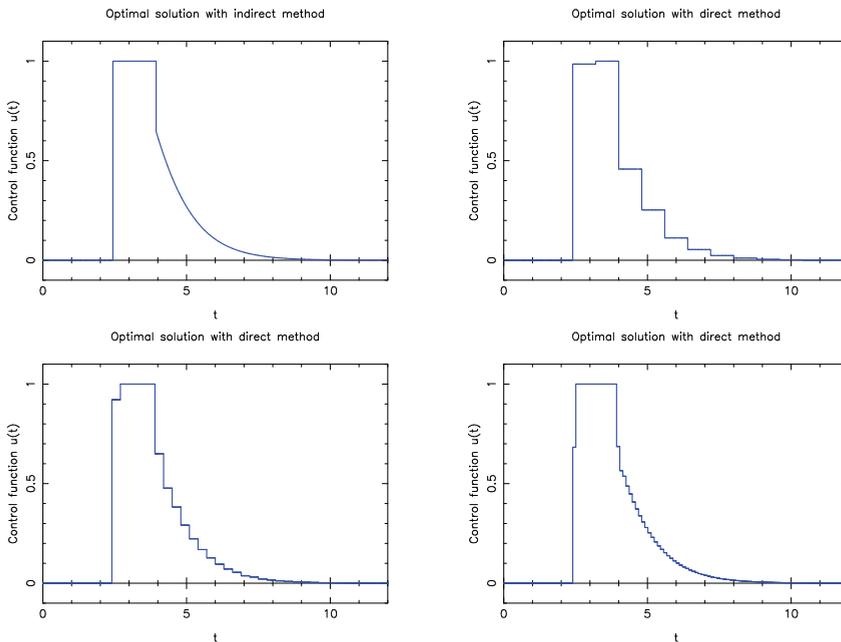
Auf Grund dieser Nachteile haben sich für die konkrete Berechnung von Regelungsstrategien in der Praxis zumeist die auch unter dem Schlagwort „first discretize, then optimize“ bekannt gewordenen *direkten* Methoden durchgesetzt, siehe auch die Diskussion in [4].

Eine weitere wichtige Frage ist der zeitliche Verlauf der berechneten optimalen Steuerfunktionen. Beim Maximumprinzip ist die optimale Steuerung eine fast beliebige Funktion, strukturelle Einschränkungen an ihren Verlauf können dort nicht gemacht werden. Von einem praktischen Gesichtspunkt aus betrachtet macht dies nicht unbedingt Sinn. Man stelle sich beispielsweise einen Temperaturregler vor, dessen Steuergröße sich aus technischen Gründen nicht kontinuierlich, sondern nur in regelmäßigen Abständen ändern lässt. Solch ein Regler kann dann nicht jede beliebige Steuerfunktion exakt nachfahren, sondern diese nur approximieren. Eine solche Approximation ist auch das Grundprinzip der direkten Methoden.

Bei diesen Methoden werden die Steuerfunktionen durch Funktionen, die durch endlich viele Steuergrößen bestimmt sind, approximiert. Im einfachsten Fall sind dies stückweise konstante Funktionen, die also auf einem gegebenen Zeitintervall $[t_i, t_{i+1}]$ den konstanten Wert q_i annehmen. In Abb. 9 wird dies veranschaulicht.

Diese Transformation auf ein mathematisches Optimierungsproblem, das nun nur noch von endlich vielen Freiheitsgraden abhängt (nämlich genau diesen q_i 's), erlaubt

Abb. 9: Links oben: Steuerung mit indirekter Methode berechnet. Die anderen Diagramme zeigen Approximationen mit der direkten Methode auf unterschiedlichen Diskretisierungsgittern.



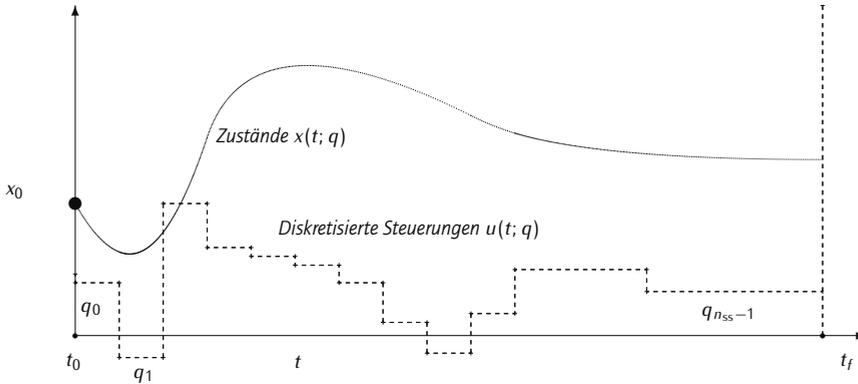
den Einsatz der weit entwickelten Methoden der nichtlinearen Optimierung (nonlinear programming). Hier konkurrieren im wesentlichen zwei Typen von Algorithmen miteinander: Innere Punkt- und active set basierte Methoden. Während bei letzteren die Menge der aktiven Ungleichungen (diese wird als active set bezeichnet) mitgeführt und von Iteration zu Iteration modifiziert wird, bewegen sich Innere Punkt Methoden durch Hilfe von logarithmischen Barrierefunktionen innerhalb des zulässigen Bereiches auf dem sogenannten zentralen Pfad auf die Lösung zu. Beide Algorithmientypen haben gewisse Vor- und Nachteile, die dazu führen, dass sie für eine bestimmte Klasse von Optimierungsproblemen schneller die optimale Lösung bestimmen. Eine Grundregel ist dabei, dass Innere Punkt Methoden häufig besser bei Problemen mit vielen Ungleichungsbeschränkungen funktionieren (weil sie aus dem Inneren kommen und die vielen Schnittpunkte der Ungleichungsbeschränkungen also gar nicht „wahrnehmen“, die bei anderen Algorithmen der Reihe nach „besucht“ werden müssen), während active set basierte Methoden effizientere Warmstarts in der Lösung verwandter Optimierungsprobleme erlauben. Dieses liegt daran, dass sich die Mengen der aktiven Ungleichungen häufig nur geringfügig unterscheiden, wenn sich beispielsweise durch neue Messwerte ein leicht verändertes Optimierungsproblem ergibt. Ein Start in der schon berechneten Lösung des Ursprungsproblems führt dann häufig schon nach wenigen Iterationen zum Ziel.

Bei der Diskretisierung von Optimalsteuerungsproblemen mit direkten Methoden werden gemeinhin drei Algorithmientypen unterschieden, siehe [4]: direktes *single shooting*, direktes *multiple shooting* und direkte *Kollokation*. *Single shooting* wurde in den 70er Jahren entwickelt, siehe beispielsweise [29] und [54]. Die Grundidee besteht darin, die differentiellen Variablen als abhängige Größen der unabhängigen Steuergrößen zu betrachten und in jeder Iteration des Optimierungsalgorithmus ein Anfangswertproblem zu lösen, um die Zielfunktion und Nebenbedingungen sowie die Ableitungen nach den Steuergrößen zu berechnen. Anders ausgedrückt: wir haben einen äußeren Optimierungsalgorithmus, in dem unsere Freiheitsgrade optimiert werden, und einen inneren, in dem wir für festgehaltene Lösungen „schießen“, also durch Integration die Differentialgleichung lösen um den Wert der Zustände zu allen Zeitpunkten zu berechnen. Abb. 10 visualisiert diese Grundidee.

Bei Optimierungsproblemen ist vor allem der Zustandswert am Ende des Zeithorizontes häufig von Interesse. Wenn man an den Wurf eines Balles denkt, so wird schnell klar, wie wichtig die Feinabstimmung aller Steuergrößen ist, möchte man ein Ziel in großer Distanz exakt treffen. Um diese Schwierigkeit über den Horizont zu verteilen, wurde das Konzept des *single shooting* zum direkten *multiple shooting*, der direkten Mehrzielmethode, erweitert. Diese Methode wurde erstmals 1981 veröffentlicht, [6]. Seither wurde die Methode ständig weiterentwickelt oder neu implementiert, siehe beispielsweise [43].

Wie beim *single shooting* auch, werden die Kontrollen auf einem Zeitgitter diskretisiert. Um den Ballwurf über eine weite Strecke aber besser koordinieren zu können, werden weitere Mitspieler in Form von zusätzlichen Variablen eingesetzt. Hier muss der

Abb. 10: Schematische Illustration der direkten single shooting Methode. Die Steuerungen sind durch stückweise konstante Funktionen q_i gegeben, die zugehörigen Zustände werden durch Integration ermittelt. Die Intervalllängen müssen nicht notwendigerweise äquidistant sein, sondern können beispielsweise auch auf den letzten Intervallen länger sein, wie angedeutet.



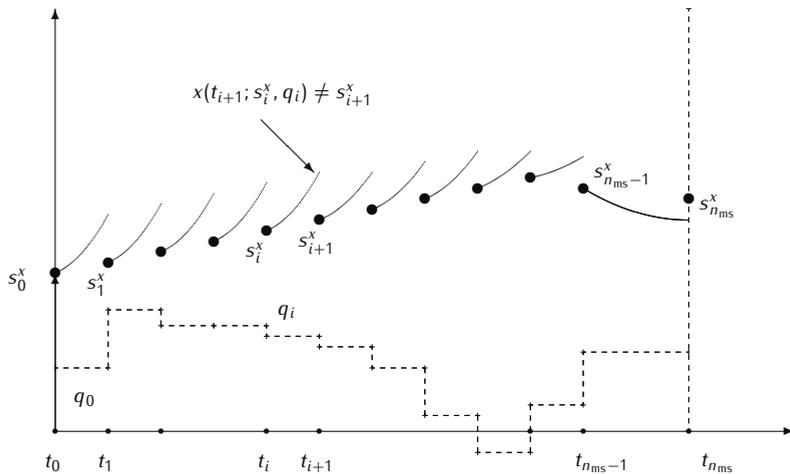
Ball nun nur noch zum nächsten Mitspieler „geworfen“ werden, ähnlich wie bei einer Menschenkette zum Löschen eines Feuers. Da Nichtlinearitäten und eventuelle Instabilitäten des Problems auf einem kurzen Zeithorizont nicht so stark zum Tragen kommen wie auf einem langen, ist es nun deutlich einfacher, das anvisierte Ziel zu treffen. Diese Herangehensweise ist ein gutes Beispiel für die Strategie, ein schweres Problem in mehrere kleine aufzuteilen. Dieses Prinzip *divide et impera*, teile und herrsche, findet sich in vielen mathematischen Algorithmen wieder.

Als neue „Mitspieler“ werden für die Zustände auf einem gegebenen Zeitgitter $t_0 \leq t_1 \leq \dots \leq t_{n_{ms}} = t_f$ multiple shooting Variablen $s_i^x \in \mathbb{R}^{n_x}$ mit $0 \leq i < n_{ms}$ eingeführt, die bei der entkoppelten Integration auf den Zeitintervallen $[t_i, t_{i+1}]$ als Anfangswerte fungieren. Wie in Diagramm 11 zu sehen, ist eine derart zusammengesetzte Trajektorie auf dem gesamten Zeithorizont nur dann stetig, wenn die Bedingungen

$$s_{i+1}^x = x(t_{i+1}; s_i^x, q_i, p) \quad (4)$$

erfüllt sind. Die linke Seite der Gleichung (4) entspricht dem Anfangswert des Zeitintervalls $[t_i, t_{i+1}]$, die rechte Seite dem aufintegrierten Wert des Zustandes zum Zeitpunkt t_{i+1} in Abhängigkeit vom Startwert s_i^x und den Steuergrößen q_i und p . Diese Gleichungen werden dem Optimierungsproblem hinzugefügt und müssen von einer optimalen Lösung erfüllt werden. Aus diesem Grund wird die direkte Mehrzielmethode häufig als *all-at-once* Ansatz bezeichnet, da Simulation und Optimierung zur gleichen Zeit betrieben werden im Gegensatz zum sequentiellen Ansatz des single shooting.

Abb. 11: Schematische Illustration der direkten Mehrzielmethode (multiple shooting). Die Steuerungen sind durch stückweise konstante Funktionen q_i gegeben, die zugehörigen Zustände werden durch stückweise Integration ermittelt. Die Anschlussbedingungen sind in diesem Beispiel nicht erfüllt, die resultierende Trajektorie ist noch unstetig.



Anschaulich gesprochen muss jeder Mitspieler sicherstellen, dass sein Wurf auch den jeweiligen Nachfolger erreicht, damit der Ball auch wirklich vom Anfangspunkt zum Endpunkt befördert wird. Die Entscheidungen, die jeder Mitspieler zu treffen hat, unsere Steuerungen q_i , müssen natürlich auch im single shooting getroffen werden. Nun kümmert sich aber jeder erstmal vorrangig um seinen eigenen Zeitabschnitt, und darum, wie er den Ball zum Nachbarn bekommt und sich dabei trotzdem optimal verhält. Dieses „sich um seinen Kram kümmern“ ist eine der größten Stärken dieses Ansatzes. Neben dem Zugewinn an Stabilität führt es nämlich dazu, dass sich Strukturen im Optimierungsproblem ergeben, die algorithmisch ausgenutzt werden können. Die Matrix, die sich bei der Analyse der notwendigen Optimalitätsbedingungen ergibt, hat Blockgestalt, wobei die Blöcke alle Variablen umfassen, die zu einem Zeitintervall gehören.

Aufgrund der Parametrisierung des Zustandsraumes sind sich *direkte Kollokation* und die direkte Mehrzielmethode recht ähnlich. Kollokation geht zurück auf [61] und wurde unter anderem in [7], [3] und [56] erweitert. Die Grundidee hier ist, gar keine unabhängigen Integratoren für die Lösung der Differentialgleichung zu nutzen, sondern sämtliche Gleichungen zu diskretisieren und mit in das Optimierungsproblem aufzunehmen.

Kollokation und Mehrzielmethode teilen mehrere Vorteile im Vergleich zum single shooting. So

- ist es möglich, Vorwissen über den Prozess (das beispielsweise in Form von Messdaten vorliegt) zur Initialisierung der Zustandsvariablen zu nutzen,
- ist garantiert, dass die auftretenden Differentialgleichungssysteme auch für nichtlineare Systeme überhaupt lösbar sind, während single shooting für schlecht gewählte Steuerungen in eine Singularität laufen kann,
- können auch instabile Systeme gelöst werden, wenn sie gut gestellt sind, da sich Störungen nicht über den kompletten Zeithorizont ausbreiten, sondern durch die Toleranzen in den Parametrisierungsvariablen gedämpft werden,
- zeigt sich in der Praxis ein wesentlich besseres Konvergenzverhalten im höherdimensionalen Raum.

Die jeweils resultierenden Optimierungsprobleme sind größer als im Fall des single shooting. Dies wird aber durch spezielle strukturausnutzende Algorithmen kompensiert. Im Falle der Kollokation sind dies Lösungsalgorithmen, die die spezielle Bandstruktur der dünnbesetzten Matrizen ausnutzen. Bei der direkten Mehrzielmethode werden Kondensierungsalgorithmen dem Quadratischen Programm vorgeschaltet, während für die Approximation der Hessematrix sogenannte high-rank updates genutzt werden. Diese tragen der speziellen Blockstruktur durch die Shooting-Intervalle Rechnung und akkumulieren so in jeder Iteration mehr Krümmungsinformation als dies bei normalen Quasi-Newton Methoden der Fall ist.

Zusammenfassend lässt sich sagen, dass sich aus unserer Sicht die direkte Mehrzielmethode und die Kollokation durchgesetzt haben, um nichtlineare Steuerungsprobleme effizient zu lösen. Darüberhinaus können diese Methoden effizient an die spezielle Struktur von NMPC-Anwendungen angepasst werden: Insbesondere die Tatsache, dass hier iterativ eine Serie ähnlicher Optimalsteuerungsprobleme gelöst wird, wurde in den letzten Jahren ausgenutzt, um speziell an NMPC angepasste Varianten dieser Methoden zu entwickeln [18,44,57], für die auch mit inakkuraten numerischen Lösungen noch die Stabilität des geregelten Systems nachgewiesen werden kann [17,19,63].

4 HERAUSFORDERUNGEN

Wir hoffen, geneigte Leserinnen und Leser bis hierher von der bedeutenden Rolle der Mathematik als Schlüsselfaktor für regelungstechnische Produktionsprozesse überzeugt zu haben, und einen – natürlich subjektiven – Einblick in den aktuellen Stand der Forschung und die Vielfalt der Anwendungsgebiete gegeben zu haben. In diesem Kapitel wollen wir uns nun einigen der vor uns liegenden Aufgaben und Herausforderungen widmen. Dazu betrachten wir in Abschnitt 4.1 Aspekte, die sich bei der Modellierung der betrachteten Prozesse ergeben und schauen, inwieweit die Mathematik auch auf dieser Ebene einen gewichtigen Beitrag leisten kann. In Abschnitt 4.2 diskutieren wir die Robustheit von MPC-Regelungen und im abschließenden Abschnitt 4.3 beleuchten wir einige weitere aktuelle Themen.

4.1 MODELLIERUNG

Die in den vorgehenden Abschnitten beschriebenen Methoden zur modellgestützten Steuerung von Prozessen haben eine entscheidende Prämisse: *„dass sich für den betrachteten Prozess eine adäquate mathematische Beschreibung in Form eines Differentialgleichungssystems finden lässt.“*

Die Schlüsselwörter in diesem Satz lassen schon erahnen, wo die Schwierigkeiten stecken.

- „für den betrachteten“ deutet an, dass es um problemspezifische Modellierungen geht. Von einer Anlage zur nächsten können einzelne Komponenten geringfügig anders sein – aber damit kann der gleiche Prozess ein komplett anderes mathematisches Modell erfordern.
- „adäquat“ impliziert mindestens dreierlei: zum einen, dass das Modell korrekt ist, also die wesentlichen Eigenschaften, die einen im speziellen Zusammenhang interessieren, beinhaltet. Zum zweiten aber auch, dass nicht wesentlich mehr modelliert ist, als unbedingt nötig ist, um Rechenzeiten nicht explodieren zu lassen. Und zum dritten, dass das resultierende Modell überhaupt numerisch geeignet ist für eine Simulation und Optimierung.
- Mit „mathematischer Beschreibung“ sind die unterliegenden Gesetzmäßigkeiten gemeint – ebenso aber auch die Anpassung an das spezifische Problem in Form von konkreten Zahlenwerten, die in das Problem eingehen (häufig als Parameter bezeichnet) und einem Vergleich von Simulationsdaten mit Messwerten standhalten müssen.
- Das sich etwas „finden lässt“, legt eine gewisse Zufälligkeit nahe. Die Frage stellt sich also, wie man solche „Findungsprozesse“ automatisieren kann.

Die mathematische Modellierung beschäftigt sich damit, ein mathematisches Modell für einen realen Vorgang zu erstellen. Dieses basiert auf Wissen aus dem jeweiligen Anwendungsgebiet. So ist spätestens seit Newton bekannt, dass die Beschleunigung die zweite Ableitung der zurückgelegten Strecke ist, welche sich daher bei festem Anfangszustand und gegebener Beschleunigung berechnen lässt. Ein anderes Beispiel sind bekannte Erhaltungssätze, die oftmals physikalischen oder chemischen Modellen zugrunde liegen. Alle solchen Modelle sind natürlich nur Näherungen der Wirklichkeit und vereinfachen an der ein oder anderen Stelle. Von der Berücksichtigung von Reibungskräften bis zu Wechselwirkungen auf Atomebene kann ein solches Modell beliebig verfeinert werden. Eine wichtige Aufgabe des Modellierers ist es, die richtige Mischung zwischen Einfachheit des mathematischen Modells und einer genügend genauen Beschreibung der Wirklichkeit zu finden.

Für viele Probleme der Praxis sind die betrachteten Prozesse leider häufig noch zu komplex, oder nicht ausreichend verstanden. Dies gilt insbesondere für den Bereich

der Systembiologie, der sich den Methoden des Wissenschaftlichen Rechnens erst seit einigen Jahren ernsthaft öffnet. Besser ist die Lage in der Robotik, Mechanik, oder Verfahrenstechnik, wo schon seit einigen Jahrzehnten Methoden der Modellierung, Simulation und auch Optimierung Anwendung finden. Die grundlegenden Prinzipien sind hier häufig sogenannte *first principles*, also grundlegende und gut verstandene Erhaltungssätze der Naturwissenschaften. Leider bleibt trotzdem festzuhalten, dass eine geeignete Modellierung – insbesondere im Bereich der nichtlinearen Modelle – nur durch sehr arbeitsintensive Bemühungen aufbauend auf vorhandenem Expertenwissen gelingt.

Die Verwirklichung der eingangs genannten Prämisse für relevante technische Prozesse bildet daher in der Tat eine große Herausforderung für die Zukunft. Es gibt aber durchaus erste Erfolge auf dem Weg zu einer algorithmischen Modellierung, die wir im Folgenden kurz exemplarisch anreißen wollen – nicht ohne zu betonen, dass auch diese Bewegung weg von dem mühsam und durch trial and error erreichten Modellierungserfolgen im stillen Kämmerlein hin zu standardisierten und verifizierbaren Methoden nicht ohne unterliegende mathematische Methoden denkbar wäre.

Parameterschätzung. Auch wenn die grundlegenden Gesetzmäßigkeiten bekannt sind, so sind doch häufig bestimmte Modellparameter unbekannt. Man denke beispielsweise an Massen von Komponenten in Mehrkörpersystemen, an Aktivierungsenergien, oder an Reaktionsgeschwindigkeiten. Zur Bestimmung dieser Parameter sind Messungen vonnöten, wobei häufig nicht die Parameter direkt gemessen werden können, sondern nur Zustände des Systems, oder Funktionen dieser Zustände. Die Parameterschätzung zielt nun darauf ab, den Abstand zwischen einer Modellantwort und erhobenen Messwerten zu minimieren durch ein *Fitten* der unbekannt Parameter. Die Methode der kleinsten Quadrate (least squares) wurde schon von Carl-Friedrich Gauß und Adrien-Marie Legendre Anfang des neunzehnten Jahrhunderts entwickelt. Sie erfuhr aber in den letzten Jahrzehnten noch bedeutende Weiterentwicklungen, insbesondere was eine effiziente Numerik, Beschränkungen an das Parameterschätzproblem, und robuste Parameterschätzung durch Verwendung von Zielfunktionalen, die Ausreißer zulassen, angeht, siehe beispielsweise [40].

Versuchsplanung. Die Parameterschätzung liefert nicht nur Parameterwerte, für die eine Simulation des Differentialgleichungssystems einen minimalen Abstand zu gemessenen Werten hat. Darüber hinaus erlaubt eine statistische Analyse die Berechnung der Kovarianzmatrix, welche ein Maß für die Genauigkeit liefert, mit der die Parameter bestimmt wurden. Vereinfacht gesagt entspricht dies Fehlerbalken, die häufig zu den Messwerten in Diagramme eingezeichnet werden.

Nun ist es so, dass verschiedene experimentelle Versuchsanordnungen zu unterschiedlichen Genauigkeiten führen. Das Ziel der Versuchsplanung ist es, Versuchsanordnungen zu bestimmen, die zu möglichst kleinen Konfidenzgebieten (Fehlerbalken) führen. Unter Versuchsanordnung wird hierbei alles verstanden, was ein Experimentator

an der Durchführung eines Versuches beeinflussen kann. Dies schließt Steuerungen genauso ein wie die Entscheidung, zu welchen Zeitpunkten (bzw. bei räumlich verteilten Prozessen: an welchen Orten) gemessen werden soll.

Während die Versuchsplanung in der Statistik schon seit langen Jahren untersucht wird, blieb der Transfer in die technischen Anwendungen weitestgehend aus. Dies liegt vor allem darin begründet, dass die Versuchsplanung lange Zeit als theoretisches Konstrukt untersucht wurde, und sich konkrete Rechnungen auf akademische Beispielprobleme beschränkten. Erst in den letzten Jahren wurde an einer effizienten numerischen Umsetzung und an einer Verallgemeinerung auf den nichtlinearen, dynamischen Fall gearbeitet. Die hierbei erzielten Durchbrüche in der Weiterentwicklung der statistischen Analyse erlauben nunmehr den Einsatz selbst innerhalb der industriellen Praxis, wo nichtlineare, dynamische Systeme von bis zu hundert differentiellen Zuständen, oder zunehmend räumlich verteilte Prozesse betrachtet werden.

Insbesondere für Prozesse, deren Dynamik durch Nichtlinearitäten geprägt ist, sind die berechneten optimalen Versuchspläne häufig alles andere als intuitiv, können aber erhebliche Verbesserungen bringen, wie das Beispiel der in Heidelberg am Interdisziplinären Zentrum für Wissenschaftliches Rechnen entwickelten und inzwischen regelmäßig beim Industriepartner BASF in Ludwigshafen eingesetzten Methoden zeigt: Nach Erfahrungswerten in mehr als zwei Dutzend konkreten Projekten zeichnet sich ab, dass gegenüber den traditionellen, von Experten ausgelegten Versuchsanordnungen, Einsparungen von über 80 % der Versuche bei gleichbleibender oder gar besserer statistischer Güte möglich sind [1]. In [39] wird von einem optimierten Versuchsplan mit nur 2 Experimenten berichtet, der die zu schätzenden Parameter auf 1 % genau bestimmt, während ein zum Vergleich von einem Experten erstellter Versuchsplan mit 15 Experimenten einen Parameter lediglich auf plus minus 30 % bestimmt. Gerade in dem hochkompetitiven Umfeld der Verfahrenstechnik tun sich dadurch enorme Einsparpotentiale auf, was Kosten, Umweltbelastung, und time-to-market angeht.

Weitere wichtige Forschungsfelder sind die *Modelldiskriminierung*, die sich mit der automatischen Erstellung von Versuchsplänen beschäftigt, um miteinander konkurrierende Modelle möglichst signifikant voneinander abgrenzen zu können, sowie die *Modellreduktion*, bei der es darum geht, ein schon bestehendes Modell zu vereinfachen, ohne die relevanten Eigenschaften zu verlieren. Ein Ansatz dazu ist die Spektralanalyse der Moden, die langsame Komponenten von den schnellen trennt, welche für die Steuerung oftmals nur eine untergeordnete Rolle spielen.

4.2 ROBUSTHEIT VON LÖSUNGEN, UNSICHERHEITEN

Während (N)MPC ein Konzept zur Regelung von Prozessen ist, das Unsicherheiten implizit Rechnung trägt, so ist insbesondere bei den in Abschnitt 3.5 vorgestellten Methoden zur offline-Optimierung vorausgesetzt, dass der zu optimierende Prozess sich deterministisch so verhält, wie es das mathematische Modell vorhersagt. In der Praxis wird dies nicht der Fall sein, denn es treten verschiedene Arten von Unsicherheiten auf:

- *Modellungenauigkeiten.* Trotz der in Abschnitt 4.1 vorgestellten Bemühungen, ein adäquates Modell für einen betrachteten Prozess zu erlangen, wird es doch immer einen Modellfehler geben, der zu (hoffentlich nur geringfügigen) Abweichungen der Simulationsergebnisse von dem realen Prozess führt.
- *Externe Störungen.* Durch zufallsbedingte Störungen (Rauschen) werden Einflüsse auf das System ausgeübt, die zu dem Zeitpunkt der Optimierung nicht bekannt sein konnten. Beispiele sind Seitenwind beim Auto, Regenschwankungen bei Abwasserkläranlagen, Konsumentenverhalten in der Wirtschaft, oder externe Temperaturschwankungen.
- *Unterschiedliche Szenarien.* Diese Art von Unsicherheit entsteht dadurch, dass bestimmte Parameter des untersuchten Prozesses a priori offen sind. Beispielsweise sollen Turbinenschaufeln oder Flugzeuge gute aerodynamische Eigenschaften für eine ganze Bandbreite von Anströmwinkeln haben. Oftmals unterscheidet man hier noch weiter zwischen einer diskreten Anzahl von Szenarien, einer Wahrscheinlichkeitsverteilung in einem kontinuierlichen Bereich, und einem worst-case Szenario.

Für alle drei Typen von Unsicherheit stellt sich sowohl dem Anwender, wie dem zuarbeitenden Mathematiker die Frage, wie man mit dieser Herausforderung umgeht.

Berechnete optimale Lösungen haben oftmals eine Eigenschaft, die ihren Einsatz in der Praxis unattraktiv macht: sie sind sehr sensitiv gegenüber Unsicherheiten. Dies liegt daran, dass die nichtlinearen Effekte so weit es geht ausgereizt werden. Auch Beschränkungen werden an vielen Stellen aktiv, das heißt maximal ausgereizt. Wenn man zeitoptimal von A nach B fahren möchte, so wird man an einigen Stellen genau die maximal erlaubte Geschwindigkeit fahren. Wenn nun das Modell ungenau ist, oder aber externe Störungen (man denke an Gegen- bzw. Rückenwind) auftreten, so werden diese Beschränkungen unter Umständen verletzt. Bei besagter Geschwindigkeitsrichtlinie mag das akzeptabel sein, aber bei sogenannten *runaway*-Prozessen, bei denen ab einem gewissen Punkt nicht mehr rückgesteuert werden kann, ist es dies nicht. Wer wollte verantworten, dass in einem Kernkraftwerk die maximale Energieeffizienz erzielt wird, wenn man dafür ständig kurz vor einer unumkehrbaren Kettenreaktion steht?

Für einen praktischen Einsatz hätte man daher viel lieber eine Lösung, die *fast* genau so gut ist wie die optimale, aber weniger störungssensitiv. Ansätze, die in diese Richtung gehen, arbeiten mit (a) Feedback, so wie es auch MPC durch das Einarbeiten von Messwerten in einem Echtzeit-Kontext tut, oder (b) mit Sicherheitsmargen, die auf die Beschränkungen addiert werden, oder (c) mit gewichteten Summen, multikriterieller Optimierung oder spieltheoretischen Ansätzen (hier gibt es erste Ansätze im Rahmen von NMPC [45], die aber algorithmisch deutlich aufwendiger als die „normale“ NMPC sind), oder aber (d) durch die Integration höherer Ableitungen in der Optimierungsroutine. Die Sensitivität der Lösung nach unsicheren Parametern lässt sich formal als Ableitung berechnen. Moderne Ansätze in der robusten Optimierung erweitern daher

das mathematische Modell oftmals um Terme mit höheren Ableitungen der Zielfunktion bzw. der Beschränkungen nach diesen Parametern. Trotz erfreulicher Fortschritte in diesem Bereich werden Unsicherheiten und die Bestimmung robuster Lösungen weiterhin eine große Herausforderung bleiben.

4.3 WEITERE HERAUSFORDERUNGEN

Während die Theorie und Algorithmik der Optimalen Steuerung und Regelung für dynamische Systeme entwickelt wurde, hat sich durch algorithmische Fortschritte und die Verfügbarkeit von schnellerer Hardware nach dem Mooreschen Gesetz die Tür geöffnet für die Betrachtung von *räumlichen Effekten*. Hiermit ist gemeint, dass die Temperatur in einem Raum nicht mehr anhand eines repräsentativen Messwertes an einer speziellen Stelle geregelt werden soll, sondern das Verhalten an jeder Stelle im Raum relevant wird.

Während dies für einfache Heizungen im Normalfall kaum nötig sein wird, gibt es viele Prozesse, bei denen die räumliche Komponente wichtig ist. Bei der Produktion von Stahlträgern beispielsweise ist strengstens darauf zu achten, dass die Temperaturunterschiede innerhalb des Trägers in der Abkühlphase nicht zu groß werden, um spätere Risse zu vermeiden, [60]. Im Bereich der Aerodynamik spielt das Fluidverhalten eine entscheidende Rolle. An welchen Stellen im Raum sich Luftwirbel bilden und wo die Schockfronten verlaufen, beeinflusst direkt die Flugeigenschaften, [35].

Die Optimierung von Prozessen, die durch partielle Differentialgleichungen beschrieben werden, hat in den letzten Jahren einen großen Zuwachs an Aufmerksamkeit erfahren, beispielsweise durch ein Schwerpunktprogramm der Deutschen Forschungsgemeinschaft [15]. Trotzdem bleibt festzuhalten, dass die Kluft zwischen mathematischer Grundlagenforschung und praktischem Einsatz an dieser Stelle noch besonders groß zu sein scheint, nicht nur im Bereich der Algorithmen zur Lösung von Optimalsteuerungsproblemen, sondern auch in den systemtheoretischen Grundlagen der MPC-Methodik.

Eine Herausforderung ganz anderer Art stellt die Kombination von kontinuierlichen und diskreten Ereignissen dar. Mit kontinuierlich ist hier ein zusammenhängender Bereich gemeint, aus dem eine Variable oder eine Steuerung einen Wert annehmen kann. Man denke beispielsweise an die Geschwindigkeit eines Fahrzeugs oder die Position des Gaspedals. Mit diskret ist ein gerade nicht zusammenhängender Bereich gemeint, der mathematisch aus einer endlichen Menge möglicher Werte besteht. Hier ist die Gangschaltung ein gutes Beispiel für eine diskrete Steuerung, während „der Laufroboter hat Bodenkontakt oder nicht“ ein Zustand ist, bei dem entweder das eine oder das andere zutrifft. Ebenso führen digitale Kopplungen in vernetzten Systemen, die aus einer Vielzahl kleinerer physikalischer Teilsysteme bestehen, auf diskrete Zustandskomponenten. Systeme, die beide Typen von Zuständen und/oder Steuerungen vereinen, bezeichnet man als *hybride Systeme*. Oftmals ergeben sich hybride Systeme durch eine Mehrskalensmodellierung. Hier werden schnelle transiente Übergänge als instantan angenommen.

Ein gutes Beispiel hierfür sind Ventile, die nur offen oder geschlossen sein können, in der Praxis aber von einem Zustand in den anderen überführt werden müssen.

Mathematisch stellen die ganzzahlig oder diskret genannten Variablen, die sich bei der Modellierung ergeben, eine erhebliche Erschwerung der Optimierung dar. Dies mag auf den ersten Blick überraschen, stehen doch nur begrenzt viele Möglichkeiten zur Verfügung, die man theoretisch alle ausprobieren kann. Doch die Anzahl der Möglichkeiten wächst enorm schnell, wenn sich die Anzahl der Variablen erhöht. Dies ist bei hybriden Systemen der Fall, insbesondere um genau entscheiden zu können, wann und wie oft umgeschaltet werden soll. Eine Behandlung dieser Variablen durch Ausprobieren scheidet also wegen der immensen Anzahl an Möglichkeiten aus. Eine Übersicht möglicher Methoden für ganzzahlige Steuerfunktionen und weiterführende Literatur zu hybriden Systemen findet man beispielsweise in [53].

Auch dieses Forschungsgebiet stellt sich als extrem aktiv dar, und wird beispielsweise durch das europäische Forschungsnetzwerk HYCON [31] und – für den Spezialfall digital vernetzter Systeme – durch einen DFG Forschungsschwerpunkt [16] gefördert. Insbesondere das Entwickeln verlässlicher und effizienter Algorithmen und eine Stabilitätsanalyse im Sinne des Abschnittes 3.3 werden in den nächsten Jahren viel Aufmerksamkeit erfordern, um das große Optimierungspotential bei hybriden Systemen ausschöpfen zu können.

5 VISIONEN UND HANDLUNGSEMPFEHLUNGEN

MPC und insbesondere NMPC gehören nach unserer Überzeugung zu den vielversprechendsten Methoden zur Lösung komplexer nichtlinearer Regelungsprobleme, weswegen sie in Zukunft weiter an Bedeutung gewinnen werden. Ihre Fähigkeit, Regelungsaufgaben nicht „nur“ lösen zu können, sondern über Optimalitätskriterien und Nebenbedingungen auch technische und ökonomische Beschränkungen einbeziehen zu können, liefert – gerade in Zeiten knapper werdender Ressourcen – deutliche Vorteile gegenüber anderen Methoden. Der rasante Fortschritt im Bereich der Optimierungsalgorithmen erlaubt dabei fortlaufend neue Anwendungen mit zunehmend schnellerer Dynamik.

Ganz im Sinne der Erfolgsfaktoren der klassischen Methoden von Hurwitz und Pontrjagin hat MPC darüberhinaus den entscheidenden Vorteil, intuitiv leicht verständlich zu sein und dadurch auch ohne tieferes Verständnis der dahinterstehenden Mathematik – seien es die systemtheoretische Analyse des Regelverhaltens oder die aus Numerik und Optimierung stammenden Grundlagen der verwendeten Algorithmen – anwendbar zu sein.

Wir wollen die Ausführungen im Folgenden daher auf diese Regelungsmethode beschränken, wenngleich sinngemäße Aussagen sicherlich auch für andere moderne Verfahren in der Steuerungs- und Regelungstechnik gelten.

5.1 THEORIE UND PRAXIS

Der aus theoretischer Sicht größte konzeptionelle Vorteil von (N)MPC – nämlich die explizite Einbeziehung eines mathematischen Modells zur Vorhersage des Systemverhaltens – ist zugleich einer der wesentlichen Nachteile in der Praxis, denn MPC verlangt dadurch zwingend nach hinreichend genauen mathematischen Modellen für die zu regelnden komplexen Prozesse. Wie in Abschnitt 4.1 ausgeführt, gibt es hier inzwischen auch im nichtlinearen Bereich vielversprechende mathematische Ansätze. Trotzdem wird das Erstellen geeigneter Modelle aller Voraussicht nach zumindest im nichtlinearen Bereich vorerst aufwendig bleiben, so dass die dafür notwendigen Investitionen nur dann von der Industrie getätigt werden dürften, wenn die erzielbaren Vorteile zur Optimierung der Produktionsprozesse durch konkrete Beispiele belegt werden. Vielen der in Abschnitt 3.4 aufgeführten neueren Anwendungen kommt daher Pilotcharakter zu, deren Erfolg oder Misserfolg (und natürlich auch die Kommunikation entsprechender Erfolge) mittelfristig über die Akzeptanz speziell von NMPC in der industriellen Anwendung mitentscheiden wird.

Aber auch auf der theoretischen Seite gibt es noch eine Reihe offener Fragen. Insbesondere unterscheiden sich, wie in Abschnitt 3.3 ausgeführt, die Annahmen im größeren Teil der theoretischen MPC-Literatur deutlich von den in der Praxis verwendeten Verfahren. Hier erscheint es wünschenswert, diese Lücke zwischen Theorie und Praxis zu schließen. Gerade theoretische Arbeiten, die auf konkrete Verbesserungen von MPC-Verfahren abzielen – sei es nun verbesserte Stabilität, schnellere Algorithmen, bessere Regelgüte, höhere Robustheit – sollten versuchen, ihre Ergebnisse und Verbesserungen unter realistischen Annahmen zu erzielen.

Auf der algorithmischen Seite schließlich ist weiterer Fortschritt vor allem durch verstärkte Kooperationen von System- und Regelungstheoretikern auf der einen Seite und Optimierern und Numerikern auf der anderen Seite zu erhoffen. So kann z. B. nur eine systemtheoretische Stabilitätsanalyse des gesamten Regelalgorithmus Aufschluss darüber geben, in welchem Sinne die in Abschnitt 4.2 erläuterte Integration höherer Ableitungen in der Optimierungsroutine die Robustheit tatsächlich erhöht – und wie genau das Optimalitätskriterium dafür aussehen sollte. Auf der anderen Seite sollte z. B. die systemtheoretische Analyse von MPC-Algorithmen stets die Eigenheiten (und die in der Praxis vielfältig angewendeten „Tricks“) der Optimierungsroutinen berücksichtigen, um verlässliche Aussagen über die Effizienz der Regelung treffen zu können. In beiden Beispielen ist also die Kompetenz aus beiden Gebieten nötig.

5.2 INTERDISZIPLINARITÄT IN DER AUSBILDUNG

All diese Ziele sind unseres Erachtens nur dann zu erreichen, wenn – ganz im Sinne der Zusammenarbeit von Hurwitz und Stodola – bereits in der Ausbildung Wert auf die Kooperation zwischen Mathematikern und Ingenieuren und zwischen den verschiedenen

Gebieten der Angewandten Mathematik gelegt wird. Nur so können Mathematiker herausfinden, welche methodischen Herausforderungen die Praxisprobleme liefern, nur so können Ingenieure von den neuesten mathematischen Methoden profitieren und nur so können sich auch innerhalb der Mathematik Verfahren und Techniken aus verschiedenen Teildisziplinen sinnvoll ergänzen.

Erfreulicherweise gibt es hier eine ganze Reihe sehr vielversprechender aktueller interdisziplinärer Projekte, wie z. B.

- das Internationale Doktorandenkolleg „Identification, Optimization and Control with Applications in Modern Technologies“ der Universitäten Bayreuth, Erlangen-Nürnberg und Würzburg im Elitenetzwerk Bayern
- die Graduiertenschule „Mathematical and Computational Methods for the Sciences“ der Universität Heidelberg
- das Center of Excellence „Optimization in Engineering“ (OPTEC) an der Universität Leuven
- das Exzellenz-Cluster „Simulation Technology“ (SimTech) der Universität Stuttgart mit Doktorandenschule sowie Bachelor- und Master-Studiengängen

um nur diejenigen universitären Einrichtungen zu erwähnen, an denen die Autoren dieses Beitrags beteiligt sind. Die formalen Rahmenbedingungen für weitere Fortschritte sind derzeit gerade in den Bereichen Steuerungs- und Regelungstechnik also durchaus günstig. Es bleibt zu hoffen, dass diese Anstrengungen fortgesetzt werden, damit viele der dort entwickelten mathematischen Verfahren den Sprung in die industrielle Anwendung schaffen und die vorhandenen Chancen des Produktionsfaktors Mathematik auch in Zukunft genutzt werden.

6 LITERATUR

- [1] BASF SE. BASF und Universität Heidelberg entwickeln gemeinsam neue Mathematik-Software für die Forschung. Press release P-08-308, June, 16th 2008.
- [2] S. Bennett. *A history of control engineering 1800-1930*. Peter Peregrinus Ltd., London, 1979. Paperback reprint 1986.
- [3] L.T. Biegler. Solution of dynamic optimization problems by successive quadratic programming and orthogonal collocation. *Computers and Chemical Engineering*, 8:243-248, 1984.
- [4] T. Binder, L. Blank, H.G. Bock, R. Bulirsch, W. Dahmen, M. Diehl, T. Kronseder, W. Marquardt, J.P. Schlöder, and O.v. Stryk. Introduction to model based opti-

- mization of chemical processes on moving horizons. In M. Grötschel, S.O. Krumke, and J. Rambau, editors, *Online Optimization of Large Scale Systems: State of the Art*, pages 295–340. Springer, 2001.
- [5] C. C. Bissell. Stodola, Hurwitz and the genesis of the stability criterion. *Int. J. Control*, 50:2313–2332, 1989.
- [6] H.G. Bock. Numerical treatment of inverse problems in chemical reaction kinetics. In K.H. Ebert, P. Deuflhard, and W. Jäger, editors, *Modelling of Chemical Reaction Systems*, volume 18 of *Springer Series in Chemical Physics*, pages 102–125. Springer, Heidelberg, 1981.
- [7] H.G. Bock. Recent advances in parameter identification techniques for ODE. In P. Deuflhard and E. Hairer, editors, *Numerical Treatment of Inverse Problems in Differential and Integral Equations*, pages 95–121. Birkhäuser, Boston, 1983.
- [8] V. G. Boltyanski, R. V. Gamkrelidze, and L. S. Pontryagin. On the theory of optimal processes (russian). *Dokl. Akad. Nauk SSSR*, 110:7–10, 1956.
- [9] C.I. Byrnes, A. Isidori, and J.C. Willems. Passivity, feedback equivalence, and the global stabilization of minimum phase nonlinear systems. *IEEE Transactions on Automatic Control*, 36(11):1228–1240, 1991.
- [10] J.-B. Caillau, J. Gergaud, T. Haberkorn, P. Martinon, and J. Noailles. Numerical optimal control and orbital transfers. In *Proceedings of the Workshop Optimal Control, Sonderforschungsbereich 255: Transatmosphärische Flugsysteme, Heronymus München, ISBN 3-8979-316-X*, pages 39–49, Greifswald, Germany, 2002.
- [11] E. F. Camacho and C. Bordons. *Model predictive control*. Springer-Verlag, London, 2nd edition, 2004.
- [12] M. Canale, L. Fagiano, M. Ippolito, and M. Milanese. Control of tethered airfoils for a new class of wind energy generators. In *Proceedings of the 45th IEEE Conference on Decision and Control, San Diego, California*, pages 4020–4026, 2006.
- [13] H. Chen and F. Allgöwer. A quasi-infinite horizon nonlinear model predictive control scheme with guaranteed stability. *Automatica*, 34(10):1205–1217, 1998.
- [14] M.A. Dahleh and I.J. Diaz-Bobillo. *Control of Uncertain Systems : A Linear Programming Approach*. Prentice Hall, Englewood Cliffs, N.J, 1995.

- [15] DFG Schwerpunktprogramm 1253, Optimization with Partial Differential Equations. <http://www.am.uni-erlangen.de/home/spp1253/>.
- [16] DFG Schwerpunktprogramm 1305, Regelungstheorie digital vernetzter dynamischer Systeme. <http://spp-1305.atp.rub.de/>.
- [17] M. Diehl, H.G. Bock, and J.P. Schlöder. A real-time iteration scheme for nonlinear optimization in optimal feedback control. *SIAM Journal on Control and Optimization*, 43(5):1714–1736, 2005.
- [18] M. Diehl, H.G. Bock, J.P. Schlöder, R. Findeisen, Z. Nagy, and F. Allgöwer. Real-time optimization and nonlinear model predictive control of processes governed by differential-algebraic equations. *J. Proc. Contr.*, 12(4):577–585, 2002.
- [19] M. Diehl, R. Findeisen, F. Allgöwer, H.G. Bock, and J.P. Schlöder. Nominal stability of the real-time iteration scheme for nonlinear model predictive control. *IEE Proc.-Control Theory Appl.*, 152(3):296–308, 2005.
- [20] R. Dittmar and B.-M. Pfeiffer. Modellbasierte prädiktive Regelung in der industriellen Praxis. *at – Automatisierungstechnik*, 54:590–601, 2006.
- [21] H.J. Ferreau, P. Ortner, P. Langthaler, L. del Re, and M. Diehl. Predictive control of a real-world diesel engine using an extended online active set strategy. *Annual Reviews in Control*, 31(2):293–301, 2007.
- [22] M. Fliess, J. Lévine, P. Martin, and P. Rouchon. Flatness and defect of nonlinear systems: introductory theory and examples. *International Journal of Control*, 61(6):1327–1361, 1995.
- [23] O. Föllinger. *Optimierung dynamischer Systeme: Eine Einführung für Ingenieure*. Oldenbourg, München, 2. edition, 1988.
- [24] B. Francis, J. Helton, and G. Zames. H_∞ -optimal feedback controllers for linear multivariable systems. *IEEE Transactions on Automatic Control*, 29(10):888–900, 1984.
- [25] R. Franke, M. Meyer, and P. Terwiesch. Optimal control of the driving of trains. *at – Automatisierungstechnik*, 50(12):606–614, 2002.
- [26] G. Grimm, M. J. Messina, S. E. Tuna, and A. R. Teel. Model predictive control: for want of a local control Lyapunov function, all is not lost. *IEEE Transactions on Automatic Control*, 50(5):546–558, 2005.

- [27] L. Grüne. Computing stability and performance bounds for unconstrained NMPC schemes. In *Proceedings of the 46th IEEE Conference on Decision and Control, New Orleans, Louisiana*, pages 1263–1268, 2007.
- [28] L. Grüne and A. Rantzer. On the infinite horizon performance of receding horizon controllers. *IEEE Transactions on Automatic Control*, 2008. To appear.
- [29] G.A. Hicks and W.H. Ray. Approximation methods for optimal control systems. *Can. J. Chem. Engng.*, 49:522–528, 1971.
- [30] A. Hurwitz. Über die Bedingungen, unter welchen eine Gleichung nur Wurzeln mit negativen reellen Theilen besitzt. *Math. Annalen*, 46:273–284, 1895. Nachgedruckt in: R. Jeltsch (ed.) et al., *Stability theory*. Birkhäuser, Basel, 1996, 239–249.
- [31] HYCON, European Network of Excellence on Hybrid Control. <http://www.ist-hycon.org/>.
- [32] A. Ilzhoefer, B. Houska, and M. Diehl. Nonlinear MPC of kites under varying wind conditions for a new class of large scale wind power generators. *International Journal of Robust and Nonlinear Control*, 17(17):1590–1599, 2007.
- [33] A. Isidori. *Nonlinear control systems*, volume 1. Springer, Berlin, 3. edition, 2002.
- [34] A. Jadbabaie and J. Hauser. On the stability of receding horizon control with a general terminal cost. *IEEE Transactions on Automatic Control*, 50(5):674–678, 2005.
- [35] A. Jameson. Aerodynamics. In E. Stein, R. De Borst, and T.J.R. Hughes, editors, *Encyclopedia of Computational Mechanics*, volume 3, pages 325–406. Wiley, 2004.
- [36] R. E. Kalman. When is a linear control system optimal? *Trans. ASME, Series D, Journal of Basic Engn.*, 86:51–60, 1964.
- [37] I. Kanellakopoulos, P.V. Kokotovic, and A.S. Morse. Systematic design of adaptive controllers for feedback linearizable systems. *IEEE Transactions on Automatic Control*, 36(11):1241–1253, 1991.
- [38] S. S. Keerthy and E. G. Gilbert. Optimal infinite horizon feedback laws for a general class of constrained discrete-time systems: stability and moving horizon approximations. *J. Optimiz. Theory Appl.*, 57:265–293, 1988.

- [39] S. Körkel. *Numerische Methoden für Optimale Versuchsplanungsprobleme bei nicht-linearen DAE-Modellen*. PhD thesis, Universität Heidelberg, Heidelberg, 2002.
- [40] S. Körkel, E. Kostina, H.G. Bock, and J.P. Schlöder. Numerical methods for optimal control problems in design of robust optimal experiments for nonlinear dynamic processes. *Optimization Methods and Software*, 19:327–338, 2004.
- [41] M. Krstic, I. Kanellakopoulos, and P.V. Kokotovic. *Nonlinear and Adaptive Control Design*. Wiley, New York, 1995.
- [42] H. Kwakernaak and R. Sivan. *Linear Optimal Control Systems*. Wiley, New York, 1972.
- [43] D.B. Leineweber, I. Bauer, A.A.S. Schäfer, H.G. Bock, and J.P. Schlöder. An efficient multiple shooting based reduced SQP strategy for large-scale dynamic process optimization (Parts I and II). *Computers and Chemical Engineering*, 27:157–174, 2003.
- [44] W.C. Li, L.T. Biegler, C.G. Economou, and M. Morari. A constrained pseudo-Newton control strategy for nonlinear systems. *Computers and Chemical Engineering*, 14(4/5)(451-468), 1990.
- [45] D. Limon, T. Alamo, F. Salas, and E. F. Camacho. Input to state stability of min-max MPC controllers for nonlinear systems with bounded uncertainties. *Automatica*, 42(5):797–803, 2006.
- [46] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. M. Scokaert. Constrained model predictive control: stability and optimality. *Automatica*, 36:789–814, 2000.
- [47] V. L. Mehrmann. *The autonomous linear quadratic control problem. Theory and numerical solution*, volume 163 of *Lecture Notes in Control and Information Sciences*. Springer-Verlag, Berlin, 1991.
- [48] Z. Nagy, B. Mahn, R. Franke, and F. Allgöwer. Evaluation study of an efficient output feedback nonlinear model predictive control for temperature tracking in an industrial batch reactor. *Control Engineering Practice*, 15:839–850, 2007.
- [49] H. J. Pesch. *Schlüsseltechnologie Mathematik*. Teubner, Stuttgart, 2002.
- [50] H.J. Pesch and R. Bulirsch. The maximum principle, Bellman's equation and Caratheodory's work. *Journal of Optimization Theory and Applications*, 80(2):203–229, 1994.

- [51] L.S. Pontryagin, V.G. Boltyanski, R.V. Gamkrelidze, and E.F. Miscenko. *The Mathematical Theory of Optimal Processes*. Wiley, Chichester, 1962.
- [52] S. J. Qin and T. A. Badgwell. An overview of nonlinear model predictive control applications. In J. C. Cantor, C. E. Garcia, and B. Carnahan, editors, *Nonlinear model predictive control*. Birkhäuser, Basel, 2000.
- [53] S. Sager, G. Reinelt, and H.G. Bock. Direct methods with maximal lower bound for mixed-integer optimal control problems. *Mathematical Programming*, published online at <http://dx.doi.org/10.1007/s10107-007-0185-6> on 14 August 2007, 2008.
- [54] R.W.H. Sargent and G.R. Sullivan. The development of an efficient optimal control package. In J. Stoer, editor, *Proceedings of the 8th IFIP Conference on Optimization Techniques (1977), Part 2*, Heidelberg, 1978. Springer.
- [55] A.J. van der Schaft. *L₂-gain and Passivity Techniques in Nonlinear Control*. Springer, London, 2. edition, 2000.
- [56] V.H. Schulz. Solving discretized optimization problems by partially reduced SQP methods. *Computing and Visualization in Science*, 1:83–96, 1998.
- [57] Y. Shimizu, T. Ohtsuka, and M. Diehl. A real-time algorithm for nonlinear receding horizon control using multiple shooting and continuation/krylov method. *International Journal of Robust and Nonlinear Control*, 2008. (in print).
- [58] H. Sira-Ramírez and S.K. Agrawal. *Differentially Flat Systems*. Marcel Dekker, New York, 2004.
- [59] B. Srinivasan, S. Palanki, and D. Bonvin. Dynamic Optimization of Batch Processes: I. Characterization of the nominal solution. *Computers and Chemical Engineering*, 27:1–26, 2003.
- [60] F. Tröltzsch and A. Unger. Fast solution of optimal control problems in the selective cooling of steel. *Zeitschrift für Angewandte Mathematik und Mechanik*, pages 447–456, 2001.
- [61] T.H. Tsang, D.M. Himmelblau, and T.F. Edgar. Optimal control via collocation and non-linear programming. *International Journal on Control*, 21:763–768, 1975.
- [62] J. Wischnegradski. Sur la théorie générale des régulateurs. *Comptes Rendus de l'Académie des Sciences de Paris*, 83:318–321, 1876.

- [63] V. M. Zavala and L.T. Biegler. The advanced step NMPC controller: Optimality, stability and robustness. *Automatica*, 2008. (accepted for publication).
- [64] K. Zhou, J.C. Doyle, and K. Glover. *Robust and Optimal Control*. Prentice Hall, Upper Saddle River, NJ, 1996.

> DATENKOMPRESSION, PROZESSOPTIMIERUNG, AERODYNAMIK: EINE TOUR DURCH DIE SKALEN

WOLFGANG DAHMEN UND WOLFGANG MARQUARDT

EXECUTIVE SUMMARY

Enorm gestiegene Rechnerkapazitäten ermutigen zu einer immer genaueren Erschließung komplexer Prozesse in Natur- und Ingenieurwissenschaften durch numerische Simulation. Die Mathematik als Schnittstelle zwischen realer und digitaler Welt liefert einerseits die Grundlagen zur Formulierung notwendigerweise vereinfachter Modelle der Realität. Andererseits bietet sie aber auch die methodischen Grundlagen für den Entwurf effizienter Algorithmen, die aus solchen Modellen quantitative Information gewinnen können. Eine zentrale Herausforderung liegt dabei in der Tatsache, dass reale Prozesse meist durch relevante Anteile auf einer riesigen Spannweite von Längenskalen getrieben werden. Die Entwicklung mathematischer Methoden zur Behandlung von Mehrskaligkeit ist daher von wesentlicher Bedeutung. In diesem Artikel wird dies anhand jüngster Entwicklungen in ganz unterschiedlichen Anwendungsbereichen erklärt und illustriert, auch um deutlich zu machen, wieviele „Fliegen“ man mit einer „mathematischen Klappe“ schlagen kann. Insbesondere werden grundlegende, auf Waveletzerlegung beruhende Prinzipien zunächst im Rahmen der Bildkompression und -kodierung auf weitgehend elementare Weise erklärt. Es wird dann aufgezeigt, dass diese Konzepte in ganz ähnlicher Weise bei der Datenanalyse im Zusammenhang mit verfahrenstechnischen Anwendungen, bei großskaligen Optimierungsproblemen der Prozessindustrie sowie bei komplexen Strömungsproblemen wie in aerodynamischen Anwendungen zum Tragen kommen. Stabile Zerlegungen in Anteile unterschiedlicher Längenskalen eröffnen dabei einen Zugang zu adaptiven Lösungskonzepten, die in der Lage sind, automatisch Rechnerressourcen dort zu plazieren, wo sie zu einer ökonomischen Realisierung der gewünschten Lösungsqualität, etwa in Form von Genauigkeitstoleranzen, benötigt werden.

1 EINIGE LEITGEDANKEN

Unter den vielfältigen Aspekten, unter denen sich die Rolle der Mathematik in einer modernen, technologiebasierten Gesellschaft betrachten lässt, wollen wir hier die Fähigkeit

zum Leitgedanken machen, oft ungeahnte *Abkürzungen* aufdecken zu können, die eine zunehmend schon aus ökonomischer Sicht unumgängliche Erschließung der realen Welt durch eine digital, virtuelle Welt immer besser ermöglicht.

Mathematik als Schnittstelle zwischen realer und virtueller Welt. Der Einfluss digitaler Informationstechnologie ist heutzutage augenscheinlich. Bei oberflächlicher Betrachtung entsteht bisweilen der Eindruck, dass schiere Leistungssteigerung von Rechenanlagen zur Lösung „aller“ Probleme in Naturwissenschaft und Technik führt. Allerdings werden noch so viele Windräder nicht in der Lage sein, Strom zu produzieren, wenn kein Wind bläst. Der umfangreichste Park von Baugeräten kann kein Gebäude errichten, wenn es keinen Konstruktions- und Einsatzplan gibt. Die reale Welt ist in ihrer Komplexität nicht erfassbar. Ein zunehmend besseres Verständnis der grundlegenden Mechanismen kann immer nur über vereinfachende *Modelle* der Realität gewonnen werden. Die Formulierung solcher Modelle drückt bereits eine Priorisierung der zu erschließenden Phänomene aus, wobei die Kunst darin besteht, all das weitestmöglich zu vereinfachen, was keine wesentliche Beeinträchtigung des Zielphänomens erwarten lässt. Die Mathematik bietet den Rahmen, solche Modelle zu formulieren und ihre Bestandteile zu quantifizieren¹. Ein typisches Beispiel ist die Modellierung von Strömungsprozessen in nahezu allen Technologiebereichen wie Aerodynamik im Flug- und Kraftfahrzeugbereich, reaktive und disperse Strömungen in der Verfahrenstechnik, Strömungen in porösen Medien bei der Analyse von Schadstoffausbreitung oder Reservoirexploration. Den Modellrahmen liefert in diesen Fällen die Kontinuumsmechanik, wobei der Teilchencharakter der betrachteten Medien vernachlässigt wird und die Größen von Interesse wie Geschwindigkeits-, Dichte- oder Druckfelder als Funktionen von kontinuierlichen Orts- und Zeitvariablen aufgefasst werden. Auf der Grundlage klassischer Bilanzgesetze für Masse, Impuls und Energie, lassen sich dann die örtlichen und zeitlichen Änderungsraten dieser Größen in Beziehung setzen und man erhält Systeme entsprechender *partieller Differentialgleichungen* oder auch von Integralgleichungen. Fragen nach der Lösbarkeit solcher Gleichungen und der Eigenschaften der Lösungen sind seit jeher eine wesentliche Quelle mathematischer Entwicklungen und selbstredend relevant für Anwendungen, die auf derartigen Modellen basieren. Nun ist es gerade die rasante Steigerung von Rechnerkapazitäten, die abgesehen von natürlicher Neugier die Wissenschaftler ermutigt, immer realitätsnähere Modelle zu formulieren, die dann naturgemäß immer komplexer werden und entsprechend neue Herausforderung an die Mathematik stellen. So ist es oft notwendig, den Rahmen der Kontinuumsmechanik zu verlassen, wie etwa in der modernen Halbleitertechnik oder in der Quantenchemie. Die Effekte einer im Rahmen des bisherigen Modells nicht darstellbaren *Skale* werden für tiefere Einsichten wichtig.

¹ Dies ist auch eine zentrale Fragestellung im gerade gestarteten Schwerpunktprogramm der DFG, SPP 1324: „Mathematische Methoden zur Extraktion quantifizierbarer Information aus komplexen Systemen“

Ein Universum von Skalen. Hier deutet sich an, dass ein wichtiger Parameter für die Komplexität der Realität eben die Spannweite der für einen realen Prozess relevanten Längenskalen ist, die man betrachten müsste, um ihn korrekt zu erfassen. Ein systematischer Umgang mit Mehrskaligkeit ist deshalb ein zentrales Ordnungsprinzip, das die Mathematik bieten kann, gangbare Wege – geeignete Abkürzungen – durch das Universum der Skalen zu finden, indem nur die relevanten Anteile herauskristallisiert werden. Diese Fragestellung prägt heutzutage viele Bereiche der Mathematik und eröffnet durchaus unterschiedliche Ansätze. Wir werden uns hier auf einen Blickwinkel konzentrieren, der insbesondere der Tatsache Rechnung trägt, dass man selbst innerhalb des Modellrahmens der Kontinuumsmechanik etwa im Zusammenhang mit turbulenten Strömungen mit einer Skalenspannweite konfrontiert wird, die auch in absehbarer Zeit nicht durch reine Rechenleistung über konventionelle Methoden handhabbar sein wird. Dies eröffnet einen wichtigen Blickwinkel, nämlich, dass die Mathematik nicht nur bei der Formulierung und Analyse des Modells stehen bleibt, sondern dass die dabei gewonnenen Einsichten direkt in den Entwurf von Lösungsalgorithmen – die ja erst die Verarbeitung auf Rechnern erlauben – einfließen müssen, um auch auf dieser Ebene bereits im Ansatz eine Komplexitätsreduktion zu erwirken. Eine Ausprägung ist das Prinzip *adaptiver Methoden*, die dafür sorgen, dass im Laufe des Lösungsprozesses Rechnerressourcen so verteilt werden, dass eine gewünschte Lösungsqualität etwa in Form von Genauigkeitstoleranzen, über eine möglichst geringe Zahl von Freiheitsgraden erzielt wird. Diese Rolle der Schnittstelle zwischen realer und digitaler Welt und die entsprechenden Konsequenzen sollen in diesem Artikel herausgearbeitet werden.

In diesem Prozess der Wissenschaftsentwicklung haben sich weite Bereiche der Mathematik gewandelt. Sie sind viel interdisziplinärer geworden. In der Tat kann sich das gesamte Potenzial nur in direkter Kooperation mit Informatikern, Ingenieuren und Naturwissenschaftlern entfalten. Dies schlägt sich allerdings nicht zuletzt auch in der *Art* der Erfolgsgeschichten nieder. Zweifellos gibt es zahllose Beispiele, wo es möglich ist, ein Anwendungsproblem geschlossen, sozusagen an die Mathematik abzugeben und nach begrenzter Zeit gelöst zurück zu bekommen. Aber auch hier zeigt sich Mehrskaligkeit im Wissenschaftsentwicklungsprozess. Oft weniger prägnant beschreibbar, aber langfristig umso wichtiger, sind methodologische Fortschritte, deren Manifestierung in der Alltagspraxis, wie am Beispiel Mehrgitterverfahren, erst nach manchmal vielen Jahren erkennbar wird, einfach weil das Anwendungsumfeld sehr komplex ist, weil oft eine einzige zündende Idee nicht reicht und vielfältige interdisziplinäre Zuarbeiten notwendig sind. Dieser Artikel ist deshalb auch ein Apell, sich nicht immer nur von kurzfristigen Zielsetzungen leiten zu lassen, sondern ein wenig auch den Blick für die Notwendigkeit längerfristiger Entwicklungen zu schärfen, die gerade aus dem Verständnis der Universität und in Ergänzung zu den immer kürzer werdenden Taktzeiten der modernen Industrie in unseren Augen unerlässlich sind. Die Stärke methodologischer Fortschritte äußert sich nicht zuletzt in einem weiteren Attribut der Mathematik, nämlich bisweilen viele Fliegen mit einer Klappe schlagen zu können, aus ökonomischer Sicht ein fraglos interessanter Aspekt.

2 ERFOLGSGESCHICHTEN

Im Folgenden wird sich zeigen, dass ein und daselbe Grundprinzip in ganz unterschiedlichen Anwendungsszenarien zum Tragen kommt, nämlich in der Signalverarbeitung, der Optimalsteuerung und in der numerischen Lösung partieller Differentialgleichungen. Wohlgermerkt, es ist nicht derselbe Algorithmus, der in den unterschiedlichen Feldern zum Einsatz kommt, sondern dasselbe mathematische Abkürzungskonzept, das die anwendungsangepassten Algorithmen treibt.

2.1 INFORMATIONSGEWINNUNG AUS RIESIGEN DATENSÄTZEN

In industriellen Anlagen zur Stoff- und Energiewandlung (wie beispielsweise Kraftwerke, Raffinerien oder Chemieanlagen) werden heute umfangreiche Datenarchive zur langfristigen Aufbewahrung der während des Betriebs erfassten Messdaten angelegt. Diese Messdatenarchivierung wird einerseits vom Gesetzgeber gefordert, um eine bestimmungsgemäße und sichere Fahrweise der Anlagen entsprechend der Errichtungsgenehmigung überprüfen zu können. Andererseits werden diese Datenarchive vom Anlagenbetreiber genutzt, um die Leistungsfähigkeit einer Anlage hinsichtlich Verfügbarkeit, Durchsatz, Produktqualität und Herstellkosten kontinuierlich zu verbessern. Außerdem wird eine Analyse der Daten in Echtzeit von den Anlagenbetreibern vorgenommen, um beispielsweise fehlerhafte Betriebszustände zu erkennen und die Fehlerursachen zu diagnostizieren, um eine vorbeugende Instandhaltung zu ermöglichen oder um die Betriebsbedingungen in Echtzeit so nachzustellen, dass eine wirtschaftlich optimale Fahrweise garantiert werden kann. Die zu behandelnden Datensätze sind sehr groß. Die simultane Erfassung und Archivierung von mehreren 10 000, in der petrochemischen Industrie sogar 100 000 und mehr Messgrößen bei sekundlicher Abtastung über einen mehrjährigen Zeitraum soll als Anhaltspunkt zur Charakterisierung der Komplexität des Problems dienen.

Bei der Archivierung der in der Anlage erfassten Zeitreihen wird zunächst eine Datenkompression vorgenommen. Sie zielt heute weniger auf die Reduktion der erforderlichen Speicherkapazität ab, weil mittlerweile sehr große Speicher kostengünstig realisiert werden können. Vielmehr ist man an der Abtrennung der grobskaligen Trends von den überlagerten Störsignalen als Grundlage der Interpretation der Signale interessiert. Die Trends spiegeln dabei die dynamischen Phänomene wieder, welche das Anlagenverhalten bestimmen, während die überlagerten Störsignale nicht nur Mess- und Übertragungsfehler, sondern auch das Prozessrauschen umfassen. Für Kompression und Trendanalyse werden heute eine Fülle von unterschiedlichen heuristischen Verfahren eingesetzt. Diese Verfahren entbehren einer vereinheitlichenden theoretischen Fundierung, sind nur mit viel Erfahrung nutzbringend anzuwenden, erfordern einen hohen Wartungs- und Pflegeaufwand und führen häufig zu nicht befriedigenden Ergebnissen.

Das Kernproblem der beschriebenen Aufgabenstellung ist die Zerlegung der multivariaten Zeitreihen in Signalanteile, die durch unterschiedliche Zeitskalen gekennzeichnet sind. Die in den letzten Jahren entwickelten Wavelet-basierten Multiskalenverfahren zur Signalanalyse [37] bilden eine hervorragende Basis zur Entwicklung einer theoretisch fundierten Sammlung von Methoden zur Komprimierung und Interpretation der multivariaten Anlagendaten.

Kürzlich wurden Methoden zur automatisierten Einstellung von industriell weit verbreiteten heuristischen Datenkompressionsverfahren entwickelt [39], die hervorragende Kompressionseigenschaften bei Zeitreihen mit unterschiedlichster Signalstruktur aufweisen. Mit diesen Methoden kann nicht nur sichergestellt werden, dass die archivierten Messdaten auch nach der Kompression eine hohe Aussagekraft bezüglich des Prozessverhaltens haben. Vielmehr erlauben die Methoden eine sichere Einstellung der Kompressionsverfahren bei minimalem Arbeitsaufwand, so dass auch bei einer sehr großen Zahl von Zeitreihen eine zuverlässige Anwendung gewährleistet werden kann [3].

In multivariaten Datensätzen lassen sich mit einem speziell zu diesem Zweck entwickelten Multiskalenverfahren [24] Zeitbereiche identifizieren, in denen alle Messgrößen in einem Toleranzband vorgegebener Breite um einen polynomialen Trend liegen. Nur die konsequente Nutzung der Eigenschaften von Multiskalen-Verfahren ermöglicht die sichere Bestimmung des Trendverlaufs einschließlich Anfangs- und Endzeitpunkt und damit eine Segmentierung des Datensatzes in Zeitintervalle mit qualitativ unterschiedlichem Verhalten. Auch dieses Verfahren konnte in industriellen Anwendungen mit hervorragenden Ergebnissen im Vergleich zu herkömmlichen Ansätzen [14] validiert werden [2]. Multiskalen-Methoden zur Kompression und Erkennung von Trends in multivariaten Messdaten bilden eine hervorragende Grundlage für die Entwicklung neuer Messdatenanalyse-Werkzeuge (siehe Abb. 1), die aufgrund ihrer theoretischen Fundierung nicht nur bessere und zuverlässigere Ergebnisse liefern, sondern auch den Engineering-Aufwand deutlich reduzieren lassen.

2.2 ADAPTIV OPTIMIERT

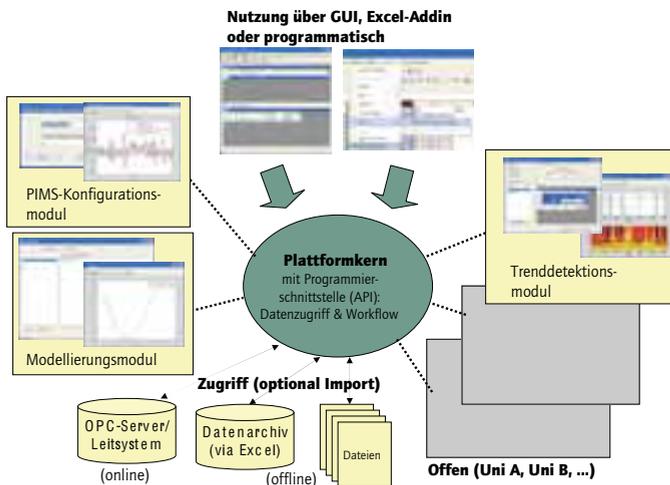
Optimierung kennt jeder aus dem täglichen Leben. Man versucht ständig, das Optimum aus gegebenen Möglichkeiten zu erzeugen – den maximalen Ertrag aus Aktien, die maximale Reichweite mit einer Tankfüllung, den maximalen Profit aus einer Fabrik. Eine bestimmte Problemklasse in dieser speziellen Disziplin der Mathematik bilden die Optimalsteuerungsprobleme. Sie lösen unendlich-dimensionale Optimierungsprobleme, bei denen die Entscheidungsvariablen Funktionen der Zeit sind. Ein einfaches Beispiel ist das Zurücklegen einer definierten Strecke mit einem Auto in minimaler Zeit, wobei das Auto beim Start und am Ziel stehen soll. Gleichzeitig sind Beschleunigung und Geschwindigkeit limitiert. Die Beschleunigung des Autos ist der Freiheitsgrad der Optimierung, die geeignet diskretisiert werden muss. Abb. 2 illustriert dieses Beispiel. Das linke Diagramm zeigt den optimalen Verlauf der Beschleunigung, das rechte Diagramm

zeigt die Verläufe von Geschwindigkeit und Distanz. Im Rahmen der Optimierung sollen mindestens 250 m in kürzest möglicher Zeit zurückgelegt werden, was durch die Wahl der Steuerung – hier der Beschleunigung – erreicht wird. Das Steuerungsprofil lässt sich in drei Teile unterteilen. Im ersten Teil wird das Auto maximal beschleunigt (10 s), dann fährt das Auto mit maximaler Geschwindigkeit (25 s) und wird anschließend so stark wie möglich abgebremst (35 s), um am Ziel zum Stehen zu kommen.

Extrapoliert man diese Problemstellung auf industrielle Anlagen der Prozessindustrie, so wird direkt klar, dass auch große Prozesse, die dynamisch betrieben werden, kontinuierliche Entscheidungsgrößen besitzen, wie z. B. den Energieeinsatz und die Rohstoffzufuhr pro Zeiteinheit. Auch hier ist sofort einsichtig, dass man mit diesen Stellgrößen gleich sehr wichtige Parameter des Prozesses beeinflussen kann, die über Ökonomie und Ökologie entscheiden. Gerade hier gebietet es die Vernunft, möglichst optimal vorzugehen. Voraussetzungen des gewünschten Vorgehens sind ein ausreichend genaues Modell des Prozesses und eine leistungsfähige Methode gegossen in eine effiziente Implementierung. Die Herausforderungen hierbei sind die enorme Größe der Modelle, die zur Modellierung von Kraftwerken oder chemischen Anlagen nötig sind, und die sehr langen Zeithorizonte, auf denen optimiert werden muss.

Die Formulierung und Lösung von Optimalsteuerungsproblemen mit differentialalgebraischen Nebenbedingungen zur Behandlung von Entwurfs- und Betriebsproblemen in der Prozessindustrie wurde bereits in den 1970er Jahren als eine zukunftssträh-

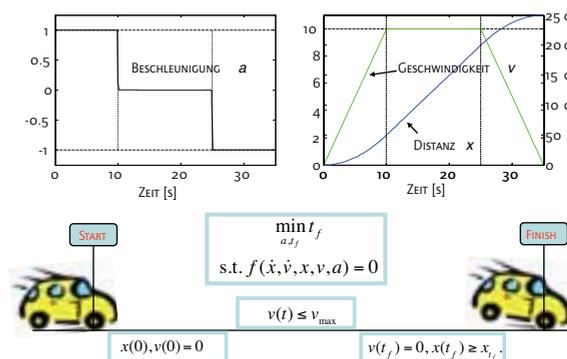
Abb. 1: Alanda, eine modulare, offene Plattform zur Datenanalyse in Offline- wie Onlineanwendungen



tige Forschungsrichtung erkannt [25]. Erste Arbeiten mit diesem Anwendungsbezug werden von [38] und [4] publiziert. Es wurde schnell klar, dass die enorme Komplexität der Problemstellungen, die häufig auch in Echtzeit zu lösen sind [6], nur mit strukturausnutzenden adaptiven Verfahren zu bewältigen sind. In einem gemeinsamen Forschungsvorhaben der Autoren wurde Mitte der 1990er Jahre ein radikal neues Konzept vorgeschlagen, das eine adaptive Verfeinerung der Zustands- und Steuergrößen des Optimierungsproblems anstrebte, um schließlich große Zustandsschätzprobleme in Echtzeit lösen zu können. Der verfolgte mehrskalige Lösungsansatz [7] zielte auf die Repräsentation der Messdaten, die Diskretisierung der Zustandsgrößen und die Parametrisierung der Steuerungen mit Hilfe eines einheitlichen mathematischen Rahmens ab, der durch biorthogonale Wavelet-Basen gegeben ist. Als herausfordernde Problemklasse wurde die Zustandsschätzung bei unsicheren dynamischen Modellen, das sogenannte dynamische Datenausgleichsproblem, ausgewählt. Die gemeinsamen Forschungsarbeiten führten zu völlig neuen Verfahrenskonzepten, welche die Überlegenheit des adaptiven Multiskalenansatzes eindrucksvoll demonstrierten [8,9]. Insbesondere konnte auch gezeigt werden, dass die Multiskalendiskretisierung sich sehr gut eignet, um den notwendigen Kompromiss zwischen Bias und Varianz bei den schlecht-gestellten inversen (dynamischen Datenausgleichs-)Problemen über eine adaptiv gewählte Diskretisierungstiefe zu identifizieren [10].

Trotz der Erfolge zeigten sich bei dem gewählten konzeptionellen Ansatz der adaptiven Volldiskretisierung mit Wavelet-Basen eine Reihe von grundsätzlichen technischen Aufgabenstellungen, deren Lösung umfangreiche Forschungsarbeiten erfordern. Um vielleicht schneller zu einer praktikablen, wenngleich konzeptionell weniger eleganten Lösung zu kommen, wurden zeitgleich erste Arbeiten zu alternativen Verfahren

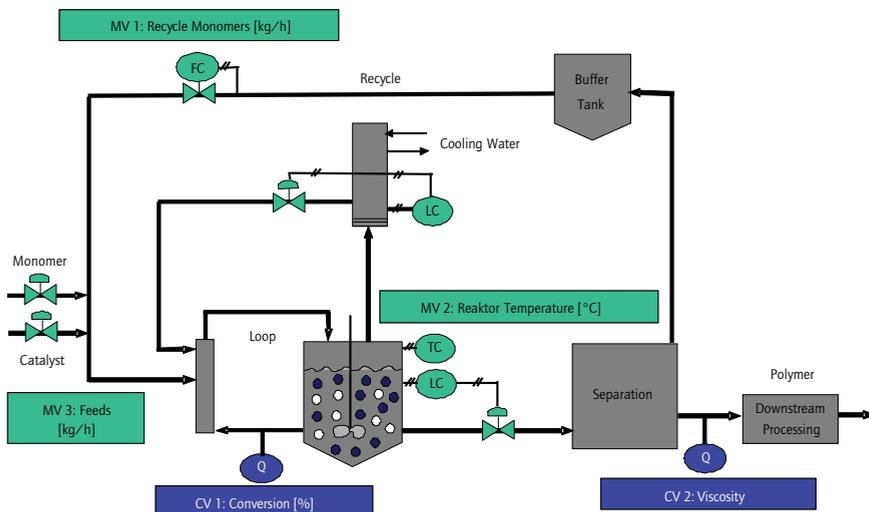
Abb. 2: Beispielproblem zur Lösung von Optimalsteuerproblemen



durchgeführt, welche lediglich die Steuerungen adaptiv diskretisieren und die Zustandsgrößen mit verfügbaren genauigkeitskontrollierten Integrationsverfahren im Sinne des Einzelschießens behandeln [5]. Diese Idee erwies sich als aussichtsreich und wurde in mehreren Entwicklungszyklen erfolgreich ausgearbeitet [36] und auf Probleme mittlerer Größe angewandt. Abb. 3 zeigt eine industrielle Fallstudie, für die eine Lastwechsel- und Produktwechselstrategie optimiert wurde [23], bei der das Optimalsteuerungsproblem ca. 1500 differential-algebraische Nebenbedingungen, mehrere Pfad- und Endpunktbeschränkungen und vier Steuergrößen umfasste. Ein weiterer Durchbruch konnte durch die automatische Erkennung der sich aus den notwendigen Optimalitätsbedingungen ergebenden Schaltstruktur erreicht werden, welche eine Minimalparametrierung der Steuerungen ermöglicht [34]. Die Strukturerkennung kann auch mit der adaptiven Steuergrößendiskretisierung verknüpft werden [33].

Dieses adaptive Verfahren konnte in jüngsten Arbeiten weiter verbessert werden. Nach einer Robustifizierung der Adaptionstrategie, einer effizienteren Behandlung der Gradientenberechnung und einer teilweisen Parallelisierung des Codes konnten Probleme einer Größe gelöst werden, die bisher nicht behandelbar war. In Zusammenarbeit mit einem industriellen Partner wurde beispielsweise ein Lastwechselproblem für eine Chemieanlage gelöst, das durch ein differential-algebraisches Modell von ca. 15 000 Gleichungen mit einer Vielzahl von Pfadbeschränkungen, mit vier Steuerungen und mit

Abb. 3: Prozessschema einer industriellen Fallstudie, die in Zusammenarbeit mit der Bayer AG bearbeitet wurde.



einem ökonomischen Gütefunktional modelliert wurde. Um die notwendige Diskretisierungsgenauigkeit zu erreichen, hätte man für ein gradientenbasiertes Optimierungsverfahren ein differential-algebraisches System von etwa 50 Millionen Gleichungen lösen müssen. Die verwendete Adaptionstrategie hat eine Reduktion auf etwa 4 Millionen Gleichungen und damit die erfolgreiche Lösung des Problems ermöglicht. Jeder optimierte Lastwechsel führt zu einer Kostenersparnis von mehreren zehntausend Euro im Vergleich zur herkömmlichen manuellen Fahrweise.

2.3 AERODYNAMIK UNTER DEM MATHEMATISCHEN MIKROSKOP

Ein scheinbar ganz anderes Anwendungsfeld: Aerodynamik spielt seit jeher eine zentrale Rolle im Flugzeugentwurf. Hierbei ergänzen sich wechselseitig Experiment und numerische Simulation. Ein höherwertiges Modell für aerodynamische Strömungen beruht auf den kompressiblen Navier-Stokes Gleichungen, die alleine bekanntlich höchste Herausforderungen an moderne Algorithmik und Rechnerleistung stellen. Deshalb beschränkt sich die Simulation meist auf „Problemausschnitte“ bzw. vereinfachte Szenarien und muss deshalb durch Experimente ergänzt und in Bezug auf die Modelltreue validiert werden.

Fluid-Struktur-Wechselwirkung. Nun stößt der experimentelle Rahmen gerade bei der Entwicklung von Großraumflugzeugen wie dem A380 zunehmend an seine Grenzen und zwar aus physikalischer und ökonomischer Sicht. Folglich ist es wünschenswert und notwendig, immer weitere Bereiche verlässlich durch Simulation zu erschließen. Gleichzeitig wird bei wachsender Flugzeugdimension der Aspekt der Aeroelastik immer bedeutender. Bedenkt man, dass die Flügelspitze zwischen Ruhelage und Hochauftrieb um mehr als 11 Meter auslenkt, wird klar, welche Kräfte die Strömung auf die Flügelstruktur ausübt. Ferner beeinflusst die Flügelverformung das Auftriebsverhalten und damit die ökonomische Seite des Flugbetriebs. Aus Gründen der Energieersparnis ist man andererseits daran interessiert, die Struktur so leicht wie möglich zu halten. Die numerische Simulation einer solchen Strömungs-Struktur-Wechselwirkung soll unter anderem Antworten auf folgende Fragen liefern. Wie ist die Flügelgeometrie auszulegen, um möglichst ökonomische Flugbedingungen zu garantieren? Wie ist die Tragstruktur auszulegen, damit unter den dynamischen Lasten keine Grenzzustände eingenommen werden, die zu überhöhter Materialermüdung und schließlich zu Bruch oder sogar zu instabilen Flatterzuständen führen?

Um diesen komplizierten Wechselwirkungsprozess numerisch simulieren zu können, sieht man sich folgenden Herausforderungen ausgesetzt: Im mathematischen Modell sind nun zwei Medien gekoppelt, nämlich die Flügelstruktur und die umströmende Luft. Bei der Kopplung des Strömungsproblems mit dem Elastizitätsproblem ist es zum Beispiel wesentlich, die Kopplungsbedingungen an der Mediengrenze so zu formulieren, dass das gesamte Problem korrekt gestellt ist, also eine eindeutige Lösung in Form

von Strömungs- und Strukturgrößen hat, da sonst ein Algorithmus keine sinnvollen Resultate liefern könnte [29–31]. Die Prozesse sind zeitlich inhärent instationär und die Kopplung der beiden Medien führt zu einem insgesamt dynamischen Prozess, der stark unterschiedliche zeitliche Längenskalen beinhaltet, also steif ist. Explizite billige Zeintegratoren sind daher ungeeignet. Stabile implizite Verfahren hingegen erfordern in jedem Zeitschritt die Lösung eines typischerweise riesigen Gleichungssystems mit mehreren Millionen Unbekannten, die unter anderem schon aus der Spannweite relevanter Längenskalen alleine in der Strömungsphase herrühren. Man schreckt im allgemeinen davor zurück, einen Algorithmus für das voll gekoppelte Problem zu entwerfen, der völlig neu zu entwickeln wäre. Stattdessen bevorzugt man, bereits vorhandene Löser für die Strömungs- und Strukturphase zu koppeln, indem wechselseitig in der Strömungs- und Strukturphase gerechnet wird [29–31]. Mathematisch entspricht dies einer Fixpunktiteration, und es ist deshalb sicher zu stellen, dass genügend innere Iterationen durchgeführt werden, die eine korrekte Kopplung gewährleisten. Insgesamt ergibt sich also ein enorm komplexes Problem, das typischerweise unter Verwendung kommerzieller Software nicht handhabbar ist. Bereits in der Anlage des Algorithmus müssen Weichen für unabdingbare Abkürzungen gestellt werden.

Wirbelschleppen. Ein zweiter Problemkreis, der besonders durch Großraumflugzeuge an kommerzieller Aktualität gewinnt, betrifft die abrollenden Wirbelschleppen, die beim Start entstehen. Nachfolgende, insbesondere kleinere Flugzeuge können erst starten, wenn diese Wirbel hinreichend abgeklungen sind. Mögliche ökonomische Vorteile eines Großraumflugzeugs werden somit zumindest teilweise durch eine reduzierte Startfrequenz wieder verspielt. Dieses Problem ist seit Langem präsent und hat zahlreiche Untersuchungen mit dem Ziel ausgelöst, Möglichkeiten zu finden, durch aktive Maßnahmen den Zerfall dieser Wirbelschleppen zu beschleunigen. Der experimentell erreichte Fortschritt ist insgesamt bescheiden. Nähere Aufschlüsse über numerische Simulation zu gewinnen, sind ebenfalls durch die enorme Komplexität des Problems sehr begrenzt. Man müsste das gesamte Wirbelfeld in einem Bereich von etwa fünfzig Flügelängen hinreichend genau darstellen können, während derzeitige Methoden etwa ein bis zwei Flügelbreiten schaffen. Auch hier liegt die zentrale Obstruktion in der enormen Spannweite von Längenskalen, die bei einer Diskretisierung zu berücksichtigen wäre.

Um beide Problemkreise geht es im Sonderforschungsbereich 401 „Strömungsbeeinflussung und Strömungsstrukturwechselwirkungen an Tragflügeln“. In einem Verbund von mehreren Teilprojekten wurde deshalb von Ingenieuren und Mathematikern ein neues Strömungslösungskonzept QUADFLOW entwickelt, in dessen Zentrum ein adaptives Multiskalen-Diskretisierungsverfahren und eine eigens darauf abgestimmte Gittergenerierungsmethode stehen [13]. Hierbei werden maßgeblich Konzepte des CAGD (Computer Aided Geometric Design) verwendet. Diese Methoden spielten insbesondere eine wesentliche Rolle bei der Freiformflächengenerierung, die der Frässteuerung zur Erstellung

Abb. 4: Modell im HiReNASD



des Flügels für das Großexperiment „High Reynolds Number Aerostructural Dynamics (HiReNASD)“ im ETW Windkanal zugrunde liegt. Abb. 4 zeigt das Flügelmodell, welches auf der Grundlage des in Abb. 5 gezeigten CAD Modells erstellt wurde.

Die entwickelten Methoden erlauben uns nun eine erhebliche Verschiebung der Grenzen der Simulation. In Kürze wird QUADFLOW voll parallelisiert sein, so dass dann eine wesentlich effektivere Grundlage für die verlässliche Simulation von Wirbelschleppen in einem stark erweiterten Bereich gegeben sein wird. Ferner sind wir in der Lage, die notwendigen begleitenden aeroelastischen Simulationen mit verlässlicher Strömungs-Struktur-Wechselwirkungsprädiktion durchzuführen. Dies stellt sicherlich einen Höhepunkt der Arbeiten im SFB 401 dar. Allerdings ergab sich bei der Auslegung des Ex-

Abb. 5: CAGD Modell



periments eine weitere Schwierigkeit. Die Modellierung der Flügelkappe erwies sich mit Hilfe herkömmlicher CAGD-Methoden als nicht durchführbar. Insbesondere geht es in diesem Zusammenhang darum, die gewünschte Form mit einer möglichst geringen Anzahl parametrischer Fläschenstücke zu generieren, die hinreichend glatt aneinander gefügt sind, damit eine darauf beruhende NC-Steuerung genügend genaue Fräsergebnisse hervor bringen kann. Bei der Lösung dieses Problems in [11, 12] wurde wesentlicher Nutzen von der engen Verbindung zwischen Gittergenerierung und Oberflächenmodellierung der umströmten Struktur gemacht, die schließlich zu einer qualitativ sauberen Modellierung der Wingletkappe führte (siehe Abb. 16 auf S. 92).

3 STATUS QUO

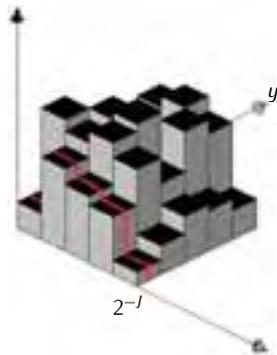
3.1 MULTISKALEN-ZERLEGUNGEN: WAVELETS

In allen erwähnten Anwendungsbereichen liegt eine zentrale Herausforderung darin, den für das jeweilige Problem relevanten Bereich von Längenskalen in der numerischen Behandlung angemessen Rechnung zu tragen, aber eben auch ohne wesentlich mehr zu investieren, als nötig ist. Wir möchten in diesem Abschnitt eine gemeinsame mathematische Konzeptgrundlage etwas herausarbeiten.

Der gemeinsame mathematische Rahmen liegt darin, die betreffenden Objekte wie Messreihen, Steuergrößen, Strömungsgrößen als *Funktionen* aufzufassen. Statt diese Funktionen nun wie üblich auf einem gegebenen Gitter zu approximieren, kann man von einer „geeigneten“ *Darstellung* von Funktionen ausgehen, die zunächst noch die gesamte Information beinhalten. Das Prinzip ist ähnlich wie bei Zahlen. Das arabische Ziffersystem bietet für Arithmetik eine weit geeignetere Zahldarstellung als etwa das römische. So werden von der Zahl π , deren Darstellung unendlich viele Ziffern beinhaltet, nur diejenigen abgerufen, die für die jeweilige Anwendung nötig sind. Nun ist es weit weniger klar, wie Darstellungen für Funktionen aussehen und ein „geeignet“ wird viel stärker vom Anwendungskontext abhängen. Für periodische Funktionen bietet die klassische harmonische Analyse eine Antwort in Form des trigonometrischen Systems und der darauf beruhenden Fourierzerlegung in Überlagerungen von Grundschwingungen, deren Anteil in der Gesamtdarstellung durch den Betrag der Fourierkoeffizienten wider gespiegelt wird. Während diese Darstellung Frequenzanteile perfekt herausfiltert, erlaubt sie keine Lokalisierung örtlicher Phänomene, was etwa für lokal verfeinerte Diskretisierungstiefen notwendig wäre. Die moderne harmonische Analyse stellt daher die Frage allgemeiner, wie man Funktionen in geeignete, gegebenenfalls ganz anders konzipierte Grundbausteine zerlegen kann, deren Koeffizienten in entsprechenden Überlagerungen auch Aufschluß über die Separation örtlich variierender Längenskalen geben können. Eine Antwort bieten *Wavelet-Basen*, die im Verlauf der letzten zwei Jahrzehnte enorme Entwicklungen in Gang gesetzt haben. Wir möchten nun, die Grundideen in stark vereinfachter Form kurz am Beispiel der Bildkompression skizzieren,

um dies dann später in den verschiedenen Anwendungskontexten wieder aufgreifen zu können.

Bilder sind auch nur Funktionen. Ein digitales Bild ist zunächst ein diskretes Objekt, nämlich eine ganzzahlige Matrix, deren Einträge im (einfachen) schwarz-weiß-Fall die Grauwerte der entsprechenden Pixel angeben. Würde man ein Bild in dieser Form speichern, gäbe es keine schnelle Bildübertragung im Internet oder die Speicherkapazität moderner Digitalkameras wäre allzu schnell erschöpft. Auch hier liegt der Schlüssel in einer alternativen Darstellung solcher Bilder, die im Prinzip auf dem bemerkenswerten Schritt beruht, das diskrete Objekt Bild als Funktion aufzufassen und zwar als Histogramm bzw. stückweise Konstante über einem Pixelraster.



Die Verteilung der Grauwerte über einem gegebenen Pixelraster wird als „Funktion“ f über dem Einheitsquadrat aufgefaßt. Definiert man

$$\phi(x) := \chi_{[0,1)}(x) := \begin{cases} 1, & x \in [0, 1); \\ 0, & \text{sonst,} \end{cases}$$

und nehmen wir an, die Schrittweite im Pixelraster sei 2^{-J} , ergibt sich mit den skalierten und verschobenen Box-Funktionen $\phi_{j,k}(x) := \phi(2^j x - k)$ die Darstellung

$$f(x) = \sum_{k=0}^{2^j-1} g_{j,k} \phi_{j,k}(x).$$

Unabhängig davon, wie sehr sich benachbarte Grauwerte unterscheiden und weniger oder mehr kleinskalige Details repräsentieren, sind in dieser Darstellung alle Grauwerte $g_{j,k}$ gleich wichtig. Eine sehr viel günstigere Darstellung, die sehr wohl Bildbereiche starker Detaillausprägung unterscheidet, läßt sich technisch einfacher im Falle einer Variablen erklären. Wir betrachten dazu in dem durch die rote Linie in obiger Abbildung angedeuteten Schnitt zwei nebeneinander liegende Grauwertsäulen (siehe Abb. 6). Man kann die beiden Pixelsäulen durch ihren Mittelwert auf der Vereinigung der beiden kleinen Intervalle ersetzen und den entstehenden Fehler als „Fluktuation“ bzw. Feinskalendetail als Vielfaches des Fluktuationsprofils $\psi_{j-1,k}(x) = \psi(2^j x - k)$ schreiben, wobei

$$\psi(x) = \phi(2x) - \phi(2x - 1) = \begin{cases} 1, & 0 \leq x < 1/2; \\ -1, & 1/2 \leq x < 1 \end{cases} \quad (1)$$

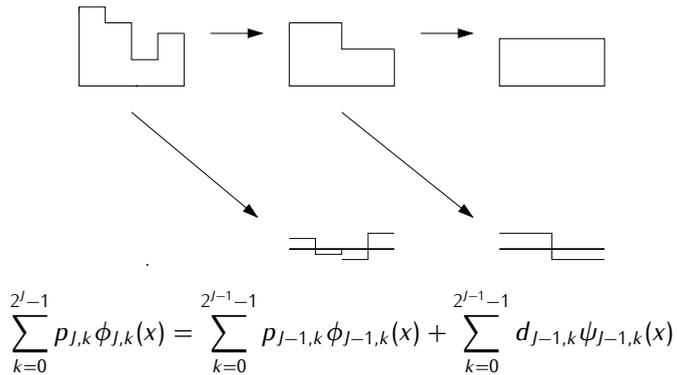
das sogenannte *Haar-Wavelet* ist.

Aus (1) leitet man leicht ab, dass die ursprünglichen Grauwerte $g_{j,k}$, $k = 0, \dots, 2^j - 1$, mit den größeren Mitteln $g_{j-1,k}$, $k = 0, \dots, 2^{j-1} - 1$, und den abgespalte-

nen Fluktuations- oder Detailamplituden $d_{j-1,k}$, $k = 0, \dots, 2^{j-1} - 1$, durch folgende „Zweiskalen-Relation“ miteinander verknüpft sind:

$$g_{j-1,k} = \frac{1}{2} (g_{j,2k} + g_{j,2k+1}), \quad d_{j-1,k} = \frac{1}{2} (g_{j,2k} - g_{j,2k+1}). \quad (2)$$

Die sogenannte *schnelle Wavelet-Transformation* ergibt sich durch Wiederholung dieses Aufspaltschritts

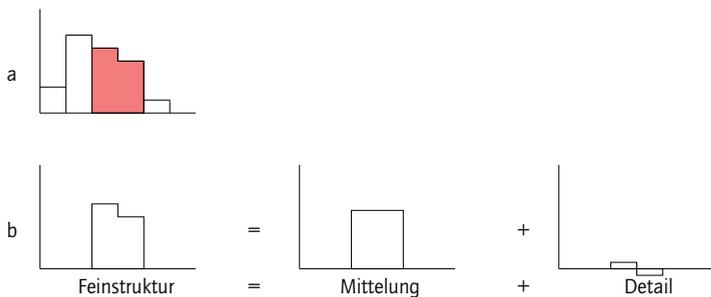


und lässt sich formal in einem Diagramm der Form

$$\begin{array}{ccccccccccc} g_j & \rightarrow & g_{j-1} & \rightarrow & g_{j-2} & \rightarrow & \dots & \rightarrow & g_1 & \rightarrow & g_0 \\ & & \searrow & & \searrow & & & & \searrow & & \searrow \\ & & d_{j-1} & & d_{j-2} & & \dots & & d_1 & & d_0 \end{array} \quad (3)$$

darstellen. Die Information aus dem ursprünglichen Array g_j von Grauwerten ist vollständig in den Arrays der Detailgrößen d_j , $j = 0, \dots, J - 1$ und des Grobmittels g_0

Abb. 6: Zerlegung in gröberes Mittel und Fluktuation



enthalten. Insofern beschreibt obiger Prozess der sukzessiven Detailabspaltung lediglich einen Basiswechsel von der Darstellung über Box-Funktionen in eine Darstellung in den Haar-Wavelets $\psi_{j,k}$. Die neue Darstellung hat aber folgenden wesentlichen Vorteil. Falls nebeneinander liegende Grauwerte fast gleich sind, das Bild also lokal wenig Variation aufweist, werden die Detail-Koeffizienten – Wavelet-Koeffizienten – sehr klein sein. Man könnte erwarten, dass die Vernachlässigung dieser Größen das Bild nur wenig verändert. Darin liegt der Kern der Bildkompression.

Funktionen sind auch nur Folgen. Bevor wir dies etwas quantifizieren, sei ein weiterer prinzipieller Vorteil unterstrichen. Die Waveletdarstellung ist nicht an eine feste feinste Skala gebunden, sondern lässt sich für beliebig hohe Auflösung formulieren, indem immer höhere Skalen hinzugefügt werden. Man kann in der Tat jede quadratintegrale Funktion in Wavelets zerlegen, indem man sallopp gesagt alle Skalen auflöst. Definiert man genauer

$$\|f\|_{L_2} := \left(\int_0^1 |f(x)|^2 dx \right)^{1/2}$$

und skaliert die $\psi_{j,k}$ mit dem Faktor $2^{j/2}$, so dass $\|\psi_{j,k}\|_{L_2} = 1$ gilt, und beachtet man dass

$$\langle \psi_{j,k}, \psi_{l,m} \rangle := \int_0^1 \psi_{j,k}(x) \psi_{l,m}(x) dx = \delta_{(j,k),(l,m)}, \quad (4)$$

erhält man die *Waveletdarstellung*

$$f(x) = \langle f, \phi_{0,0} \rangle + \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} \langle f, \psi_{j,k} \rangle \psi_{j,k} \quad (5)$$

von f , die der Funktion f eine „Ziffernfolge“ $d = \{d_{j,k}\}_{j=-1,k=0}^{\infty,2^j-1}$ mit $d_{j,k} = \langle f, \psi_{j,k} \rangle$ ($d_{-1,0} := \langle f, \phi \rangle$) zuordnet, analog zu Zahldarstellungen. Aufgrund folgender Tatsachen ist diese Darstellung nun für viele Zwecke geeignet: Erstens, aufgrund der Orthonormalität (4) der Basisfunktionen gilt

$$\|f\|_{L_2} = \|\{d_{j,k}\}_{j,k}\|_{\ell_2} := \left(\sum_{j=-1}^{\infty} \sum_{k=0}^{2^j-1} |d_{j,k}|^2 \right)^{1/2}, \quad (6)$$

d. h. die Euklidische Norm der Koeffizientenfolge gleicht der L_2 -Funktionennorm. Für eine beliebige endliche Teilsumme $P_{\Lambda} f := \sum_{(j,k) \in \Lambda} d_{j,k} \psi_{j,k}$ heißt das gerade, dass

$$\|f - P_{\Lambda} f\|_{L_2} = \|\{d_{j,k}\}_{(j,k) \notin \Lambda}\|_{\ell_2} = \left(\sum_{(j,k) \notin \Lambda} |d_{j,k}|^2 \right)^{1/2} \quad (7)$$

gilt. $P_{\wedge} f$ ist also eine umso bessere Approximation an (die unendliche Entwicklung) f , je kleiner die Euklidische Norm $\|\{d_{j,k}\}\|_{(j,k) \in \wedge}$ der in $P_{\wedge} f$ nicht vorkommenden Koeffizienten ist. Kleine Störungen der Koeffizientenfolge bedingen nur kleine Störungen der Funktion – des Bildes.

Aufgrund welcher Eigenschaften von f sind nun manche der Koeffizienten $d_{j,k} = \langle f, \psi_{j,k} \rangle$ klein? Hierzu nutzt man aus, dass die $\psi_{j,k}$ orthogonal zu den konstanten Funktionen sind, d. h. $\langle c, \psi_{j,k} \rangle = c \int_0^1 \psi_{j,k}(x) dx = 0$. Man sagt auch die $\psi_{j,k}$ haben ein *verschwindendes Moment*. Bezeichnet man nun mit $I_{j,k} := [k2^{-j}, (k+1)2^{-j}]$ den Träger von $\psi_{j,k}$, so ergibt dies

$$|d_{j,k}| = |\langle f, \psi_{j,k} \rangle| = \inf_{c \in \mathbb{R}} |\langle f - c, \psi_{j,k} \rangle| \leq \inf_{c \in \mathbb{R}} \|f - c\|_{L_2(I_{j,k})} \leq 2^{-j} \|f'\|_{L_2(I_{j,k})}, \quad (8)$$

d. h. wenn immer f auf $I_{j,k}$ eine kleine Ableitung besitzt, also sehr glatt ist, dann ist $d_{j,k}$ betragsmäßig klein. In diesem Sinne filtern Wavelets Bereiche heraus, in denen die Funktion wenig variiert und zwar auf jeder dyadischen Längenskala. Grob gesagt werden alle Wavelet-Koeffizienten unterhalb eines gewissen Schwellwertes weggelassen. Das genauere Vorgehen einer Kompression/Kodierung, wie sie dem JPEG Standard 2000 entspricht, beruht auf einer baumstrukturierten Bitzuweisung für Wavelet-Koeffizienten oberhalb gewisser Schwellwerte (Quantisierung), was als „hard thresholding“ bezeichnet wird. Für gewisse Modellklassen von Bildern kann rigoros Optimalität solcher Methoden in dem Sinne bewiesen werden, dass bis auf konstante Faktoren die durch die metrischen Entropiezahlen gegebenen unteren Schranken für Aufwand/Genauigkeitsraten erreicht werden [19]. Das Entrauschen von Signalen, d. h. das Herausfiltern von Datenstörungen, folgt ähnlichen Prinzipien in der Form des sogenannten „soft thresholding“ [22].

Nun ist das Haar-Wavelet das einfachste Beispiel. Mit Hilfe von Produktansätzen bekommt man in ähnlicher Weise Wavelets für höherdimensionale Gebiete. In d Raumdimensionen gibt es dann allerdings $2^d - 1$ Fluktuationstypen. Ferner kann man mit mehr technischem Aufwand Wavelets mit höherer Glattheit und vor allem mit mehr verschwindenden Momenten konstruieren. Bei Orthogonalität zu Polynomen vom Grade m würden bei hinreichender Glattheit von f die Wavelet-Koeffizienten $d_{j,k}$ mit der Ordnung $2^{-(m+1)k}$ abklingen, also sehr viel mehr Raum für Kompression bieten. Schließlich zeigt sich, dass die Konstruktion erheblich leichter wird, wenn man auf strikte Orthogonalität verzichtet und sich in (6) mit einer beidseitigen Ungleichung zufrieden gibt, wobei die Konstanten in der unteren und oberen Schranke unabhängig von f sind. Solche Basen nennt man *Riesz-Basen*. Sie treten als Paare biorthogonaler Basen auf [17]. Zusammenfassend sind es in welcher Spielart auch immer meist die folgenden drei Eigenschaften:

- *Lokalität*: die Durchmesser der Wavelets skalieren mit der dyadischen Skala;
- *verschwindende Momente der Ordnung m* : schnelles Abklingen der Koeffizienten in glatten Bereichen;

- *Normäquivalenz*: die Koeffizientennorm lässt sich von oben und unten durch eine Funktionennorm abschätzen, so dass sich Störungen im Funktionen- und Koeffizientenbereich entsprechen und somit Funktionen im Wesentlichen mit Folgen identifizierbar werden, die einer Rechnerbearbeitung naturgemäß viel näher stehen.

Es sind in all den erwähnten Anwendungskontexten (Bild- bzw. Signalverarbeitung, adaptive Optimierungsstrategien oder adaptive Verfahren zur numerischen Lösung partieller Differentialgleichungen) diese Attribute, die es erlauben, Anteile unterschiedlicher Längenskalen quantitativ zu identifizieren.

3.2 MULTISKALENVERFAHREN ZUR DATENANALYSE

Obwohl schon vor einiger Zeit gezeigt wurde, dass die Messdatenkomprimierung durch eine Wavelet-basierte Zerlegung herkömmlichen heuristischen Methoden überlegen ist [39], konnten diese Methoden wegen einer wenig ausgeprägten Kooperationsbereitschaft der Software-Anbieter bisher nicht ihren Weg in kommerzielle Systeme finden. Da die kommerziell eingesetzte stückweise lineare Approximation der Zeitreihen mit dem Swinging-Door- oder dem Boxcar-Backslope-Algorithmus zu sehr guten Ergebnissen führen kann, wenn die Einstellparameter für jede Zeitreihe manuell richtig eingestellt werden, wurde eine Multiskalen-Methode entwickelt, welche die Wahl der Einstellparameter weitgehend automatisiert an die Eigenschaften der Zeitreihe ermöglicht. Die von [3] beschriebene Methode berechnet die Einstellparameter dieser Algorithmen nach einer Multiskalen-Zerlegung des Signals mit biorthogonalen Spline-Wavelets der Ordnung (3,3), vgl. [17]. Die wichtigste Größe ist die Standardabweichung, die nach einer Entrauschung der Zeitreihe mit der Waveshrink-Methode [22] abgeschätzt wird.

Die mit dem neuen Einstellverfahren erzielbaren Verbesserungen werden an zwei Beispielen aus der Chemieindustrie illustriert. Abb. 7 zeigt oben links eine Temperaturmessung, die mit der ursprünglichen Einstellung zu stark komprimiert wird. Die Transiente wird bei einer mittleren absoluten Abweichung von 1,24 K nicht gut wiedergegeben. Mit der korrigierten Kompressionseinstellung sehen die Schwingungen (rechts) bei einer mittleren absoluten Abweichung von 0,28 K bei einem Kompressionsfaktor von 3 natürlicher aus. Abb. 7 zeigt unten links eine irreführende Abwärtsbewegung in einer Temperaturmessung, die sich bei der ursprünglichen Kompression ergibt. Nach der korrigierten Einstellung werden die vorhandenen Schwingungen (rechts) wirklich sichtbar. Die neue Einstellmethode reduziert den Engineering-Aufwand bei einer deutlich verbesserten Qualität der komprimierten Zeitreihen bei vergleichbaren Kompressionsverhältnissen. Damit lassen sich die komprimierten Zeitreihen mit einer hohen Zuverlässigkeit interpretieren.

Eine wichtige Grundlage zur Interpretation von multivariaten Zeitreihen ist die Erkennung von Trends. Ein multivariates Messsignal y_m folgt einem Trend der Ordnung n ,

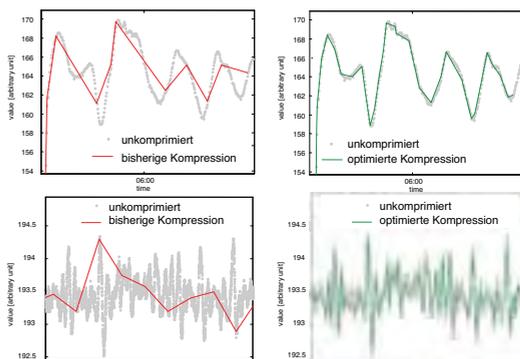
wenn der messbare Trendfehler

$$\epsilon_m(t_0, \Delta t) := \frac{1}{\Delta t} \int_{t_0}^{t_0+\Delta t} \left\| \frac{dy_m(t)}{dt} \right\|_2^2 dt. \quad (9)$$

kleiner als eine vorgegebene Schranke ist, welche die natürlich unvermeidbaren Störungen im Signal quantifiziert. Für $n = 1$ ergibt sich ein stationärer Trend, der für die Analyse der Betriebsbedingungen in kontinuierlichen Prozessen von größter Bedeutung ist. Die nächst höheren Werte von n verweisen auf polynomiale Trends. Lineare ($n = 2$) und quadratische ($n = 3$) Verläufe sind praktisch von Bedeutung, um eine Drift in ausgewählten Prozessgrößen zu identifizieren. Die Schranke ergibt sich dabei aus dem Anwendungskontext oder aber aus einer Analyse der Variation in den Signalen. Die Größe Δt steht für die Länge des Zeitintervalls, in dem der Trend in den multivariaten Messungen beobachtet wird.

Zur Erkennung eines Trends der Ordnung n auf einem gegebenen Zeitintervall (bestimmt durch Anfangswert und Länge) werden die skalaren Signale $y_{m,i}$ nacheinander zunächst mit Hilfe einer Wavelet-Basis zerlegt, dann im Raum der Wavelet-Koeffizienten entrauscht und differenziert [1], um schließlich für jede einzelne Zeitreihe den Trendfehler durch Integration [20] und durch Summation für alle gemessenen Zeitreihen zu berechnen (vgl. Gl. 9). Falls der sich ergebende Trendfehler kleiner als die vorgegebene Schranke ist, ist ein Trend der gegebenen Ordnung auf dem gegebenen Zeitintervall erkannt. Offensichtlich besteht die größte Schwierigkeit darin, bei sehr vielen Messgrößen (z. B. einige hundert) und langen Zeitreihen (z. B. einige hunderttausend Datenpunkte) die Intervalle zu finden, in denen die Messgrößen dem postulierten Trend folgen. Dieses Problem konnte nur durch die Entwicklung eines neuen Multiskalen-Verfahrens effizi-

Abb. 7: Übermäßig komprimierte Daten einer Temperaturmessstelle und automatische Korrektur der Einstellung



ent gelöst werden, das eine rekursive Berechnung des Trendfehlers von der größten zur feinsten Skala voranschreitend berechnet [24]. Mit diesem Algorithmus lassen sich sowohl Anfangs- und Endzeit des Intervalls ggf. bis zur Genauigkeit der Quantisierung der Zeitreihe sehr effizient ermitteln.

Abb. 8 zeigt beispielhaft die Ergebnisse der Suche eines linearen Trends in zwei Zeitreihen eines industriellen Prozesses. Grau hinterlegt sind die Bereiche, in denen der lineare Trend gefunden wurde. Diese Technologie kann auf vielfältige Weise eingesetzt werden. Bei einer Anwendung in Echtzeit wird beispielsweise im Sinne einer Fehlerdiagnose detektiert, zu welchem Zeitpunkt eine nennenswerte Drift einer kritischen Prozessgröße einsetzt. Die Suche nach stationären Trends in einem Datenarchiv wird dann eingesetzt, wenn gemessene Daten in stationären Betriebsphasen verwendet werden sollen, um ein datengetriebenes Modell durch Korrelation von abhängigen und unabhängigen Prozessgrößen beispielsweise für die Vorhersage von Kenngrößen der Produktqualität zu bestimmen. In einer industriellen Fallstudie wurden beispielsweise in einem Datenarchiv Betriebsphasen einer bestimmten Mindestlänge aufgesucht, die als stationär anzusehen sind, um ein solches Modell zur Vorhersage des Schmelzindex eines Polymers aus den die Fahrweise charakterisierenden Prozessgrößen zu bestimmen. Mit geringem Rechenaufwand konnten in einem Datensatz von ca. 700 000 Punkten insgesamt 32 Zeitintervalle mit den gesuchten Eigenschaften gefunden werden. Abb. 9 zeigt eine Zeitreihe mit dem Intervall, das den Anforderungen gerade noch genügt.

Eine Evaluierung dieser Verfahren in der industriellen Praxis hat gezeigt, dass die Methode auch bei großen Datensätzen ausgezeichnet skaliert und zu Ergebnissen hervorragender Qualität führt, die von typischerweise eingesetzten Filtern [26] oder statistischen Verfahren [14] nicht annähernd erreicht werden können.

Abb. 8: Identifikation von linearen Trends in zwei Zeitreihen durch eine rekursive Berechnung des Trendfehlers mit einer Multiresolution

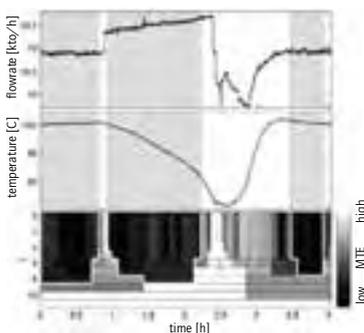
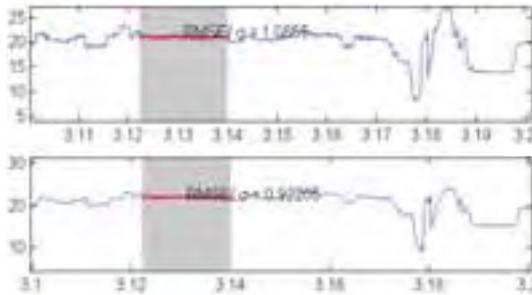


Abb. 9: Interval mit stationärer Zeitreihe von zwei ausgesuchten Prozessgrößen



3.3 LÖSUNG VON OPTIMALSTEUERUNGSPROBLEMEN MIT ADAPTIVER WAVELET-DISKRETISIERUNG

Der gewählte Lösungsansatz für das Optimalsteuerproblem (CDYNOPT)

$$\begin{aligned}
 \min_{\mathbf{u}(t), \mathbf{y}_0} \Phi &= \Phi(\mathbf{y}(t_f)) && \text{(CDYNOPT)} \\
 \text{s.t. } \dot{\mathbf{y}} &= \mathbf{f}(\mathbf{y}(t), \mathbf{x}(t), \mathbf{u}(t)), && t \in [t_0, t_f], \\
 0 &= \mathbf{g}(\mathbf{y}(t), \mathbf{x}(t), \mathbf{u}(t)), && t \in [t_0, t_f], \\
 0 &= \mathbf{y}(t_0) - \mathbf{y}_0, \\
 0 &\geq \mathbf{h}(\mathbf{y}(t), \mathbf{x}(t), \mathbf{u}(t)), && t \in [t_0, t_f], \\
 0 &\geq \mathbf{e}(\mathbf{y}(t_f), \mathbf{x}(t_f), \mathbf{u}(t_f)), \\
 \mathbf{u}(t)^{\min} &\leq \mathbf{u}(t) \leq \mathbf{u}(t)^{\max}, && t \in [t_0, t_f].
 \end{aligned}$$

beruht auf der schon angesprochenen Diskretisierung der einzelnen mathematischen Entscheidungsvariablen $\mathbf{u}(t)$ des Simulationsmodells. $\mathbf{y}(t)$ ist der Vektor der differentiellen Zustände des Modells, $\mathbf{x}(t)$ der Vektor der algebraischen Zustände, die beide auf dem Zeithorizont $t \in [t_0, t_f]$ definiert sind. Das Prozessmodell wird durch die Funktionen $f(\cdot)$ und $\mathbf{g}(\cdot)$ beschrieben. Der Anfangszustand des Systems wird durch \mathbf{y}_0 beschrieben. Für die zeitabhängigen Entscheidungsvariablen gilt $\mathbf{u}(t) \in [\mathbf{u}_{\min}, \mathbf{u}_{\max}]$. Zusätzlich können die Zustände über den gesamten Zeithorizont durch $\mathbf{h}(\cdot)$ oder zur Endzeit $\mathbf{e}(\cdot)$ beschränkt sein.

Meist bestehen die Simulationsmodelle aus vielen Tausend Variablen, jedoch nur einige wenige sind die Entscheidungsvariablen, mit denen man den Prozess steuern kann. Die Entscheidungsvariablen werden durch

$$u_i(t) = \sum_{k \in \Sigma_i} p_{i,k} \phi_{i,k}(t), \quad i = 1 \dots n_u, \quad (10)$$

diskretisiert, wobei n_u die Zahl der Entscheidungsvariablen ist. Hingegen werden alle anderen Variablen durch eine numerische Integration gelöst, deren Ergebnis natürlich von den Entscheidungsvariablen $\mathbf{z} := [p_{u,i,k}]$ abhängt. Ein Optimierungsalgorithmus muss nun für sämtliche Parameter, welche die Entscheidungsvariablen diskretisieren, den optimalen Wert finden. So wird das Optimalsteuerungsproblem in das Nichtlineare Optimierungsproblem

$$\min_{\mathbf{z}} \Phi = \Phi_M(\mathbf{y}(\mathbf{z}, t_f)) \quad (\text{NLP})$$

$$\begin{aligned} \text{s.t. } & 0 \geq \mathbf{h}^y(\mathbf{y}(\mathbf{z})), \\ & 0 \geq \mathbf{e}(\mathbf{y}(\mathbf{z}, t_f)), \\ & \mathbf{z}^{\min} \leq \mathbf{z} \leq \mathbf{z}^{\max} \end{aligned}$$

umformuliert. Um (NLP) lösen zu können, muss wiederholt die numerische Integration der Zustände des Modells für sich stetig verbessernde Trajektorien der Entscheidungsvariablen durchgeführt werden. Da der Optimierungsalgorithmus Gradienten benötigt, werden zusätzlich zu den Differentialgleichungen des Simulationsmodells die Sensitivitätsgleichungen gelöst [36], welche die Ableitungen des Modells nach den Parametern des diskretisierten Problems darstellen:

$$\dot{\mathbf{s}}_i^y = \frac{\partial \mathbf{f}}{\partial \mathbf{y}} \mathbf{s}_i^y + \frac{\partial \mathbf{f}}{\partial \mathbf{x}} \mathbf{s}_i^x + \frac{\partial \mathbf{f}}{\partial \mathbf{z}_i}, \quad (11a)$$

$$\mathbf{0} = \frac{\partial \mathbf{g}}{\partial \mathbf{y}} \mathbf{s}_i^y + \frac{\partial \mathbf{g}}{\partial \mathbf{x}} \mathbf{s}_i^x + \frac{\partial \mathbf{g}}{\partial \mathbf{z}_i}. \quad (11b)$$

Die Qualität der Lösung ist nun stark abhängig von der Feinheit der gewählten Diskretisierung. Unser Ansatz betrachtet eine ganze Sequenz von Optimierungsproblemen, die sich aus unterschiedlichen Diskretisierungseinheiten ergibt [33–35]. Beginnend auf der größten Diskretisierungsebene werden zunächst grobskalige Anteile erfasst, wofür kurze Rechenzeiten ausreichen. Bei der Lösung des Optimierungsproblems der nächst feineren Diskretisierungsebene können dann die vorher berechneten groben Lösungen als Startwerte verwendet werden. Dies wird so oft wiederholt, bis die Lösung sich nur noch marginal verändert. Zudem werden die Verfeinerungen der Diskretisierungen nicht global durchgeführt sondern adaptiv, also nur dort, wo eine höhere Auflösung am dringendsten benötigt wird, um mit möglichst geringer Zahl von Freiheitsgraden eine möglichst hohe Lösungsqualität zu sichern. Die Grundlage der adaptiven Verfeinerungsstrategie ist die geschickte Ausnutzung der Eigenschaften von biorthogonalen Spline-Wavelets. Um ein Gitter bereitstellen zu können, das exakt auf die Lösung des aktuell betrachteten Systems zugeschnitten ist, müssen zwei Elemente in diesem Algorithmus vorhanden sein, nämlich die

- Verfeinerung der Funktionsdarstellung, wo eine Verbesserung der Lösung zu erwarten ist, und die
- Eliminierung von Detailbeiträgen wenn keine Verschlechterung der Lösung bewirkt wird.

Zur adaptiven Diskretisierung mit Wavelets werden die Steuergrößen entsprechend

$$u_i(t) = c_{0,0}\phi_{0,0}(t) + \sum d_{i,k}\psi_{i,k}(t) \quad (12)$$

zu einer Wavelet-Basis reformuliert.

Eliminierung. Die optimale Lösung $u_i^{*,l}$ des Verfeinerungsschritts l kann in eine äquivalente Multiskalenformulierung überführt werden, so dass

$$u_i^{*,l} = \mathbf{d}_{\Lambda_i}^{*,l,T} \Psi_{\Lambda_i} \quad (13)$$

gilt, wobei Λ_i die Menge der Wavelet-Indizes ist. Um nun Kandidaten für eine Eliminierung zu finden, wird eine geeignete Approximation $\bar{u}_i^{*,l}$ gesucht, für die bei einer gegebenen Toleranz ϵ'

$$\frac{\|\bar{u}_i^{*,l} - u_i^{*,l}\|_{L_2}}{\|u_i^{*,l}\|_{L_2}} \leq \epsilon' \quad (14)$$

gilt. Die Äquivalenz der Normen erlaubt, diese Anforderung in Bedingungen für diese Koeffizienten zu überführen, die eliminiert werden können, wenn $d_{\Lambda_i}^{*,l} < \epsilon$ gilt. Die übrigen Koeffizienten bilden die signifikanten Koeffizienten und werden für die nächste Optimierung der Sequenz gebraucht.

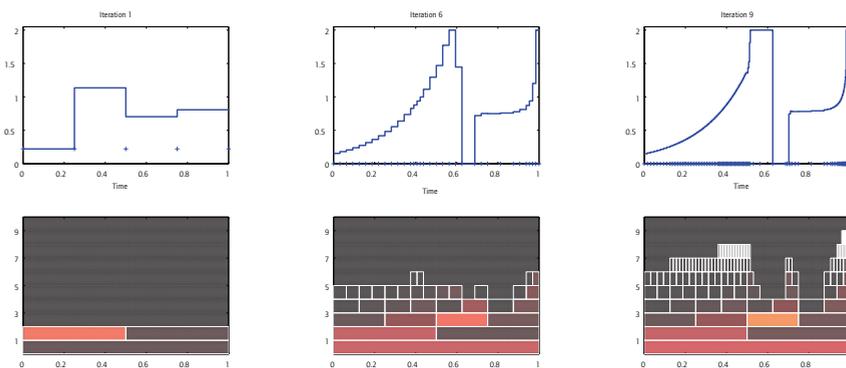
Verfeinerung. Neben der Eliminierung von Diskretisierungsstellen bildet die gezielte Verfeinerung das zweite Standbein der adaptiven Strategie. Die Herangehensweise ähnelt der Eliminierung. Während kleine Wavelet-Koeffizienten aus $\mathbf{d}_{\Lambda_i}^{*,l}$ eliminiert werden, indizieren große Koeffizienten die Notwendigkeit der Verfeinerung, um die Lösung der nächsten Optimierung verbessern zu können. Hier können nun die Nachbarn $\psi_{i,k-1}$ und $\psi_{i,k+1}$ oder die Basisfunktionen der nächst höheren Ordnung $\psi_{i+1,2k}$ und $\psi_{i+1,2k+1}$ hinzugezogen werden. Zur Verfeinerung werden nun diejenigen Wavelets hinzugezogen, die einen definierten Prozentsatz κ der Norm der Wavelet-Koeffizienten der letzten Optimierung ausmachen.

Ein Beispiel. Eine entsprechende Verfeinerungssequenz ist schematisch in Abb. 10 dargestellt. Die oberen Diagramme der Abbildung zeigen die Verfeinerung der Diskretisierung eines Steuerprofils in einem Bioprozess.

Eine Bakterienkultur in einem Bioreaktor produziert ein spezielles Protein und benötigt dafür Glukose. Um eine optimale Ausbeute an Protein zu erzielen, muss die Glu-

kosedosierung optimiert werden. Die oberen Diagramme zeigen die optimalen Glukosezulaufraten. Für jeden Iteration der Optimierungssequenz werden feinere Wavelets mit zu den Basisfunktionen hinzugenommen (untere Diagramme). Der Algorithmus detektiert, in welchen Teilintervallen die Lösung noch schlecht ist (angegeben durch die Farbe der Kästchen), und erhöht dort die Genauigkeit. Die optimale Trajektorie der neunten und damit letzten Iteration (rechtes Bild) liefert eine deutlich höhere Ausbeute als die ersten Optimierungen der Sequenz. Es stellt sich heraus, dass im größten Teil des Zeitintervalls nur eine mäßige Auflösung erforderlich ist. Der steile Anstieg der Zufuhr bei 0,5 und zum Schluss sowie das Ende der „Hungerzeit“ bei 0,7 müssen allerdings sehr präzise getroffen werden. Die unteren Diagramme der Abb. 10 zeigen die Ergebnisse der Wavelet-Analyse des Steuerprofils, wobei die Farbe die Notwendigkeit einer Verfeinerung indiziert. Durch die adaptive Diskretisierung wird das Problem signifikant verkleinert und benötigt deutlich geringere Rechenzeiten. Die Rechenzeiten konnten sogar soweit reduziert werden, dass Optimalsteuerungsprobleme nun in Echtzeit gelöst werden können, damit der Prozess im laufenden Betrieb permanent am optimalen Arbeitspunkt arbeiten kann [27].

Abb. 10: Darstellung einer adaptiven Optimierungssequenz



3.4 ADAPTIVE VERFAHREN FÜR PARTIELLE DIFFERENTIALGLEICHUNGEN DER STRÖMUNGSMCHANIK

Wie gesagt spielt die Simulation von Strömungsprozessen unterschiedlicher Ausprägungen in unzähligen Anwendungsfeldern eine zentrale Rolle. Über die Kette

Anwendung → mathematisches Modell → Analyse → Diskretisierung → Algorithmus
→ Abbildung auf Rechnerarchitekturen

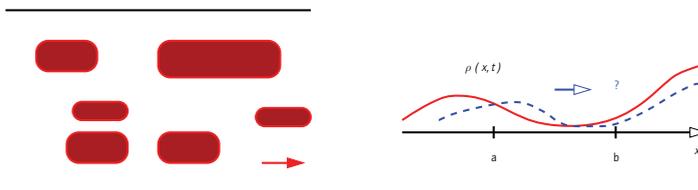
wird man meist mit riesigen (algebraischen) Gleichungssystemen konfrontiert, die durch Black-Box-Methoden nicht mehr sinnvoll handhabbar sind. Die typische Herangehensweise für Strömungen in kompressiblen Medien sei zunächst an einem einfachen Beispiel illustriert, an dem auch erkennbar wird, wo die konzeptionellen Verbindungen zur Bild- bzw. Signalanalyse liegen.

Modell. Um die im Alltag erlebten Wellen- und Verdichtungsphänomene im Straßenverkehr zu verstehen, mag man zunächst die einzelnen Fahrzeuge als Partikel in einem fließenden System auffassen (siehe linke Seite in Abb. 11), das im Grenzfall als Kontinuum betrachtet wird, wobei die Straße selbst auf eine Linie reduziert wird und die über Anzahl der Autos pro Autolänge definierte Dichte $\rho(x, t)$ als Funktion von einer Ortsvariablen x und einer Zeitvariablen t betrachtet wird, siehe linker Teil der Abb. 11. Insbesondere ist die Dichte dann eine nichtnegative Funktion mit der Eigenschaft $0 \leq \rho(x, t) \leq 1$. Man möchte nun die zeitliche Entwicklung dieser Funktion verstehen. Dazu ist klar, dass die Geschwindigkeit der einzelnen Partikel mit der Dichte verknüpft sein sollte, also $u(x, t) = U(\rho(x, t))$ gelten sollte. Ein sinnvoller Ansatz wäre

$$U(\rho(x, t)) := u_{\max}(1 - \rho(x, t)),$$

was ja besagt, dass bei geringerer Dichte zunehmend höhere Geschwindigkeiten zugelassen sind. So etwas würde man allgemein „konstituierende Gleichung“ nennen. Um

Abb. 11: Verkehrsfluss



nun die Dynamik des Systems zu erfassen, wird eine *Erhaltungseigenschaft* verwendet. Dazu beachte man, dass von der Dimension her, die Größe

$$f(\rho) := U(\rho)\rho$$

eine Massenflussrate darstellt. Greift man nun irgendein (zeitunabhängiges) Intervall $[a, b]$ der Straße heraus (siehe Abb. 11), dann entspricht die zeitliche Änderungsrate der in diesem Intervall enthaltenen Masse gerade der Differenz der Massenflussraten an den Intervallgrenzen, d. h.

$$\frac{d}{dt} \int_a^b \rho(x, t) dx = f(\rho(a, t)) - f(\rho(b, t)) = - \int_a^b \frac{d}{dx} \underbrace{f(\rho(x, t))}_{:=U(\rho(x, t))\rho(x, t)} dx. \quad (15)$$

Daraus schließt man formal, dass für *jedes* Intervall (a, b)

$$\int_a^b \partial_t \rho(x, t) + \partial_x f(\rho(x, t)) dx = 0 \quad (16)$$

gelten muss, woraus folgt (falls der Integrand beispielsweise stetig ist), dass der Integrand verschwinden muss, also die Differentialgleichung

$$\partial_t \rho + \partial_x f(\rho) = 0 \quad (17)$$

gegebenenfalls mit einer Anfangsbedingung $\rho(x, 0) = \rho_0(x)$ erweitert, gelten muss.

Dies ist das einfachste Beispiel einer skalaren Erhaltungsgleichung, die dieses Modell für den Verkehrsfluss begründet. Bilanzgesetze für Masse, Impuls und Energie, ergänzt durch konstituierende Relationen, die Materialeigenschaften enthalten, führen auch für sehr viel komplexere Verhältnisse bei kompressiblen Strömungen wie in der Aerodynamik auf partielle Differentialgleichungen ähnlichen Typs

$$\partial_t u + \partial_x (f(u) + g(\partial_x u)) = S, \quad (18)$$

wobei nun allerdings u ein Vektorfeld von Erhaltungsgrößen wie Dichte, Impuls und Energie ist, (18) also keine skalare Gleichung, sondern ein System von Funktionen in mehreren Ortsvariablen x ist. Der zusätzliche Term $g(\partial_x u)$ bezieht dann bei höherwertigen Modellen Wärme- und Reibungseffekte mit in die Bilanzen ein, wobei Ortsableitungen zweiter Ordnung ins Spiel kommen. Schließlich werden durch den Term S mögliche Quellen oder Senken modelliert.

Es ist im allgemeinen unmöglich, Lösungen solcher Gleichungen in geschlossener Form anzugeben. Während bei skalaren Gleichungen eine weit ausgebauten Theorie vorliegt, sind im Falle von Systemen selbst fundamentale Fragen nach der Existenz und Eindeutigkeit von Lösungen im räumlich 3-dimensionalen Fall offen. Im Vertrauen auf die Natur hindert das natürlich nicht daran, solche Gleichungen *numerisch* wenigstens annähernd zu lösen.

Diskretisierung. Hier kommt das Prinzip der *Diskretisierung* ins Spiel, das Gleichungen von Funktionen in Gleichungssysteme überführt, deren Unbekannte nun diskrete Größen wie Arrays von Zahlen sind. Je genauer allerdings die Lösung des diskreten Problems die Lösung des zugrunde liegenden kontinuierlichen Problems in einem zu präzisierenden Sinne approximieren soll, umso mehr Freiheitsgrade wird man ins diskrete System investieren müssen, d. h. umso größer werden die resultierenden Gleichungssysteme.

Um ein einfaches Diskretisierungsprinzip anzudeuten, betrachten wir wieder den einfachen Fall des Verkehrsflusses. Statt ein einzelnes Intervall (a, b) zu betrachten, zerlegen wir das gesamte Rechengebiet in Teilintervalle (x_i, x_{i+1}) mit Intervallgrenzen x_i . Für jedes dieser Teilintervalle muss die Bilanz (15) gelten. Multipliziert man diese Bilanz noch mit $1/\Delta x$, $\Delta x := x_{i+1} - x_i$, ergibt sich

$$\frac{d}{dt} \underbrace{\frac{1}{\Delta x} \int_{x_i}^{x_{i+1}} \rho(x, t) dx}_{:=\rho_i(t)} + \frac{1}{\Delta x} \left(f(\rho(x_{i+1}, t)) - f(\rho(x_i, t)) \right), \quad i = 0, \dots, N. \quad (19)$$

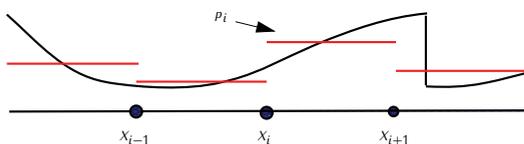
Oftensichtlich ergibt sich eine Evolution für die *Zellmittelwerte* (siehe Abb. 12)

$$\rho_i(t) := \frac{1}{\Delta x} \int_{x_i}^{x_{i+1}} \rho(x, t) dx, \quad i = 0, \dots, N, \quad (20)$$

was die Grundlage sogenannter *Finiter-Volumen-Verfahren* liefert. Es sei betont, dass vom Ansatz her hierbei nicht die Dichte selbst, sondern *Funktionale* der Dichte – stückweise Konstante – in der Zeit propagiert werden. Die Dichte als Funktion selbst ist daraus gegebenenfalls mit höherer Ordnung zu rekonstruieren. Diskretisiert man nun die Zeitvariable im einfachsten Fall durch $\frac{d}{dt} \rho_i(t) \approx (\rho_i(t + \Delta t) - \rho_i(t)) / \Delta t$, setzt $\rho_i(t_n) := \rho_i^n$, und wählt man geeignete „numerische Flüsse“ $F(\rho_{i-1}^n, \rho_i^n)$ als Approximation für die tatsächlichen Flüsse, die ja nicht von den Zellmittelwerten, sondern von Punktwerten abhängen, ergibt sich als diskretes Analogon zu (17)

$$\rho_i^{n+1} - \rho_i^n + \frac{\Delta t}{\Delta x} \left(F(\rho_i^n, \rho_{i+1}^n) - F(\rho_{i-1}^n, \rho_i^n) \right) = 0, \quad i = 0, \dots, N. \quad (21)$$

Abb. 12: Zellmittelwerte



Hierbei ist \tilde{n} entweder gleich n oder $n + 1$. Im ersteren Fall ist das Verfahren explizit. Die Werte im $(n + 1)$ sten Zeitschritt ergeben sich durch Einsetzen von Werten aus dem n ten Zeitschritt. Im Falle der Strukturkopplung hat man es mit Modellen der Form (18) zu tun, die im allgemeinen aufgrund steifer Komponenten in der Zeit *implizit* zu diskretisieren sind. Dann ist $\tilde{n} = n + 1$ und somit kommen die Unbekannten ρ_i^{n+1} auf beiden Seiten von (21) vor. Es ergibt sich in jedem Zeitschritt also ein *nichtlineares Gleichungssystem* in $N + 1$ Unbekannten ρ_i^{n+1} , $i = 0, \dots, N$.

Je größer man N wählt, je kleiner sind die Intervalle und eine umso bessere Übereinstimmung zwischen lokaler Dichteverteilung und Zellmittelwerten ist zu erwarten. Nun wird die Dichteverteilung lokal recht unterschiedliches Verhalten aufweisen. In manchen Bereichen weicht die Dichte nur sehr wenig vom Zellmittelwert ab, und dies wäre auch noch der Fall, wenn die zugrundeliegenden Intervallabschnitte größer gewählt würden. In anderen Bereichen, etwa in der Umgebung eines „Schocks“ – einer Unstetigkeit, die einem Stoppen des Verkehrsflusses entspricht – bräuchte man sehr viel kleinere Intervalle, um eine gute Übereinstimmung zwischen Mittelwerten und exaktem Verlauf zu garantieren, siehe Abb. 12.

Das Ziel einer adaptiven Lösungsmethode ist es nun, im Verlauf des Lösungsprozesses *automatisch* die Zellgrößen so zu wählen, dass eine gewünschte Gesamtgenauigkeit der numerischen Lösung unter Benutzung von möglichst wenigen Zellen, also möglichst wenigen Unbekannten realisiert wird. Derartige Aufgaben verlangen eine sehr enge Verknüpfung der Analysis der betreffenden Differentialgleichungen mit dem Entwurf solcher Algorithmen und ist daher eine inhärent mathematische Fragestellung, [15, 16]. Eine effiziente Implementierung erfordert neue Datenstrukturen, die die naturgemäß komplexere dynamische Datengenerierung handhaben können, womit im Zusammenhang mit Parallelisierung ausgeprägt informatische Aspekte ins Spiel kommen.

Ein neues Adaptionskonzept – eine Analogie zur Signalkompression. Bisher bekannte Adaptionismethoden für kompressible Strömungen sind entweder vorwiegend heuristischer Natur oder auf sehr eingeschränkte Fälle begrenzt. Der im SFB 401 gewählte Ansatz ist völlig anders und verwendet Konzepte der „Computational Harmonic Analysis“.

Hier einige grundlegende Gesichtspunkte dieser Entwicklungen: Das adaptive Diskretisierungsverfahren zielt wie gesagt darauf ab, eine hinreichend genaue numerische Lösung der Strömungsgleichungen mit einer möglichst geringen Anzahl von Unbekannten, sprich Freiheitsgraden zu gewährleisten. Die Entwicklung, theoretische Fundierung und Realisierung adaptiver Lösungskonzepte in konkreten Anwendungskontexten ist ein Kernthema am Lehrstuhl für Mathematik des Instituts für Geometrie und Praktische Mathematik [15, 16]. Es zeigt sich, dass dadurch die Problemgröße im vorliegenden Anwendungsbereich bisweilen um einen zwei- bis dreistelligen Faktor reduziert werden kann. Die Grundkonzeption unterscheidet sich von herkömmlichen adaptiven Methoden durch die zugrunde liegende Multiskalenzerlegung des Strömungsfeldes, die auf modernen Wavelet-Methoden beruht, ganz ähnlich wie sie bei der Messdatenanalyse und auch

bei der Diskretisierung des Optimalsteuerungsproblems zum Zuge kommen, [21, 32]. Das zugehörige Gittergenerierungskonzept beruht ebenfalls auf einer neuen Idee, die einerseits Adaptivität unterstützt und andererseits der Notwendigkeit Rechnung trägt, effizient zeitlich veränderliche Geometrie handhaben zu können. Das Rechengebiet wird wie bei herkömmlichen Blockkonzepten in hexagonale Blöcke zerlegt, die allerdings anders als sonst üblich als Bild einer B-Spline-basierten Abbildung generiert werden. Es werden also niemals die aufgrund der Adaptivität sich dauernd verändernden immer noch großen Gitterpunktmengen gespeichert, sondern lediglich die sehr viel kleineren Felder von Kontrollkoeffizienten der zugrunde liegenden B-Spline-Darstellungen, deren Größe konstant bleibt.

Um noch beim einfachen Beispiel des Verkehrsflusses zu bleiben, wird ja jedem Zeitpunkt t das Strömungsfeld durch einen Array von Zellmittelwerten repräsentiert, ganz analog zu einem Pixel-Array eines digitalen Bildes. Genau wie in der Bildkompression die Wavelet-Zerlegung Aufschluss über die Relevanz unterschiedlicher Skalenanteile gab, kann man die Strömungsdaten in ihre Multiskalenanteile zerlegen, indem man eine geeignete stückweise konstante Waveletbasis wählt. „Geeignet“ heißt hier insbesondere, dass sie eine möglichst hohe Ordnung verschwindender Momente hat, um glatte Strömungsbereiche effektiv herauszufiltern und aus den resultierenden inhomogen verteilten Zellgrößen immer noch hochgenaue Rekonstruktionen des eigentlichen Strömungsfeldes zu gestatten, vgl. [21].

Die Analogie zur Bildkompression ist natürlich durch den Umstand begrenzt, dass ja ein Bild mit seiner Gesamtinformation vorliegt, während das Strömungsfeld unbekannt ist und nur implizit durch die Differentialgleichung gegeben ist. Die Detektion der relevanten Skalenanteile muss also durch die Einbeziehung der Differentialgleichung geeignet ergänzt werden [13, 21, 32]. [18] und [32] stellen ausführlich die wesentlichen Konzepte und ihre analytische Fundierung dar.

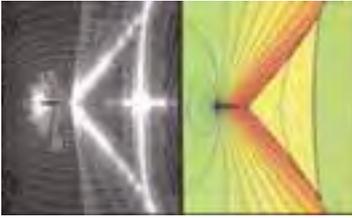
Vom Prinzip her hat dann das Verfahren folgende Form:

0. Waveletzerlegung der Anfangswerte und Bestimmung eines Startgitters im Prinzip wie bei der Bildkompression;
1. Prädiktion eines Gitters für den nächsten Zeitschritt, wobei der dominant hyperbolische Charakter der Differentialgleichung mit entsprechenden Informationsausbreitungsschranken benutzt wird;
2. Zeitintegration des Strömungsfeldes in lokal verfeinerter Zellmitteldarstellung;
3. Wavelet-Zerlegung des propagierten Feldes, Kompression und Übergabe an (1).

Dieses Verfahren ist in gleicher Weise auf die Bilanzgleichungen der Gasdynamik anwendbar und wird im adaptiven Strömungslöser QUADFLOW verwendet [13].

Folgende Beispiele sollen die Wirkungsweise illustrieren. Bei Abb. 13 handelt es sich um eine klassische Benchmark-Rechnung für ein BAC-12 Profil. Bei einem Anstellwinkel von 0° wird eine reibungsfrei modellierte Strömung von 0.95 Mach simuliert. Links ist

Abb. 13: Reibungsfreie Transschallströmung



das adaptiv verfeinerte Gitter zu sehen. Es deutet die Spannweite der relevanten Längenskalen in der Strömung an. Die rechte Seite zeigt die entsprechende Druckverteilung. Hierbei kam es darauf an, Stoßinteraktionen auch weit ab vom Profil scharf auflösen zu können. Ein herkömmliches Verfahren auf der Grundlage uniformer Gitterverfeinerung würde für die hier erzielte Zielgenauigkeit etwa 77 Millionen Gitterzellen erfordern, während die adaptive Methode mit etwa 55 Tausend Zellen auskommt. Daraus ergibt sich ein Eindruck der erreichbaren Komplexitätsreduktion, die durch Adaptivität bewirkt werden kann.

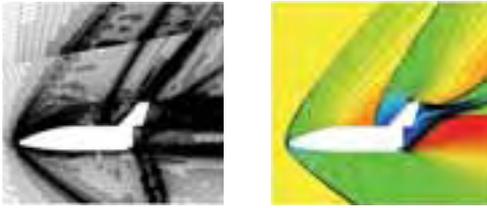
Abb. 14 zeigt eine Hochauftriebskonfiguration in einem sehr viel niedrigeren Machzahlbereich. In diesem Fall werden sinnvollerweise Reibungseffekte berücksichtigt. Der rechte Teil der Abbildung zeigt die exzellente Übereinstimmung der experimentell und numerisch ermittelten Widerstandsbeiwerte, während der linke Teil das adaptive Gitter und die berechnete Machverteilung zeigt.

Eine gleichermaßen verlässliche Auflösung der relevanten Strömungsskalen in einem anderen physikalischen Regime, nämlich im Hyperschallbereich von 8.5 Mach, wird in Abb. 15 illustriert.

Abb. 14: Reibungsbehaftete Hochauftriebskonfiguration



Abb. 15: Reibungsfreie Hyperschallströmung



Ein adaptiver Gittergenerierer. Eine wesentliche Grundlage der obigen Wavelet-basierten Adaptionskonzepte bilden die damit eng verwobene Gittergenerierung, die anders als gewohnt auf der Auswertung B-Spline-basierter Abbildungen eines Rechengebiets auf die unterschiedlichen Gitterblöcke beruht [28]. Dies unterstützt sowohl eine effiziente Gitternachführung bei instationären Gebietsgeometrien als auch Gitterverfeinerung und -vergrößerung. Insbesondere konnten auf dieser Grundlage der Geometrieverarbeitung wesentliche Probleme bei der Erstellung des Windkanalmodells bewältigt werden. Im weiteren Verlauf des Projekts soll die Flügelkappe durch Varianten von Winglets ersetzt werden, die unterschiedliche Krümmungsradien und Neigungswinkel aufweisen. Inzwischen konnten Methoden entwickelt werden, derartige Winglets automatisch zu generieren (siehe Abb. 16). Aufgrund der speziellen Designvorgaben als auch der Anforderungen des Strömungslösers für die Navier-Stokes Gleichungen konnte dies mit Hilfe kommerzieller CAD-Systeme nicht geleistet werden. Näheres über die verwendeten Modellierungs- und Oberflächenglättungsmethoden findet man in [11, 12].

Abb. 16: Winglet-Kappe



4 STÄRKEN/SCHWÄCHEN-ANALYSE

Interdisziplinär ausgerichtete Arbeitsgruppen, wie sie sich an der RWTH Aachen insbesondere im Rahmen des „Centers for Computational Engineering Science“ herausgebildet haben, bieten im Prinzip hervorragende Voraussetzungen, langfristig tragende, methodologische Fortschritte im Bereich Modellierung und Simulation zu erarbeiten, die durchaus Durchbruchcharakter haben und, ganz im Sinne des Schlagens vieler Fliegen mit einer Klappe, potenzierte Früchte tragen können. Eine wesentliche Voraussetzung ist eine Wissenskulturlage aller beteiligten Parteien, die auf Komplementarität und damit erhöhte Synergien statt auf einseitige Zulieferarbeiten abzielt. Die skizzierten Arbeiten zeigen unserer Meinung nach, wie erfolgreich dies sein kann, wenn man vor allem nicht durch die kurzen Produktionstaktzeiten der industriellen Entwicklungsprozesse gebunden ist. Gerade weil, sicher als Antwort auf ökonomischen Druck, ein nicht zu übersehender Trend zum Outsourcing existiert, muss die Universität komplementäre Forschungsschienen entwickeln, die dazu beitragen, nachhaltige Innovation zu sichern. Dabei ergeben sich unserer Erfahrung nach allerdings folgende Hindernisse. In Grundlagenbereichen eilt das konzeptionelle Potential oft der Anwendungspraxis sogar erheblich voraus. Die chronisch dünne Personaldecke und die minimalisierte Infrastruktur an universitären Forschungsinstituten nimmt allerdings oft den Charakter der Mangelverwaltung an. Folglich reicht die Kapazität bei weitem nicht aus, innovative Ansätze auf eine Ebene der Gebrauchsfertigkeit zu heben, die einem Nichtspezialisten einen routinemäßigen Umgang mit solchen Tools gestatten. Selbst unterhalb der Benutzerebene haben die entwickelten Verfahren, die bereits in diesem Stadium oft Software mit mehreren hunderttausend Code-Zeilen benötigen, lediglich Pilot-Charakter, die weder hinreichend mächtige Schnittstellen zu realen Anwendungssituationen aufweisen, noch in ihrer Struktur hinreichend optimiert sind, das tatsächliche Potenzial auch auszuschöpfen. Da viele Entwickler zudem auf zeitlich begrenzten Stellen sitzen, ist die Kontinuität des Erarbeiteten oft stark gefährdet. Dieser wohl auf absehbare Zeit bestehenden strukturellen Mangelsituation steht zudem eine gerade durch die schneller werdenden Industrietaktzeiten noch verschärfte Schwierigkeit gegenüber, innovative Ansätze in die Industriepraxis zu tragen, was wiederum ein notwendiges, insbesondere korrigierendes Feedback vermindert.

5 VISIONEN UND HANDLUNGSEMPFEHLUNGEN

Substantielle Verbesserungen des Status quo sehen wir deshalb in der Schaffung besserer Schnittstellen zwischen universitärer Forschung speziell im Grundlagenbereich und potenziellen Abnehmern in der Industrie. Hier wäre es sicherlich enorm hilfreich, wenn auch im industriellen Umfeld der Entscheidungsträger der Blick dafür erheblich geschärft werden könnte und insbesondere die Einsicht wachsen würde, gerade in den zeitlichen Entwicklungsperspektiven ausgesprochene Mehrskaligkeit zu akzeptieren und

in die jeweilige Planung mit einzubeziehen. Die bestehenden Defizite sind der DFG im Prinzip bewusst und die Einrichtung von Transferprojekten kann als eine Reaktion auf diese Erkenntnis gewertet werden. Hier muss sich unserer Meinung nach allerdings noch eine erheblich verbesserte Symmetrie zwischen den Partnern Universität und Industrie einstellen.

DANKSAGUNG

Der DFG sind wir aufgrund ihrer steten Förderung u. a. im Rahmen des Leibniz-Programms verbunden. Ferner möchten wir uns ganz besonders bei Dr.-Ing. F. Alsmeyer (AIXCAPE e. V.), und Dipl.-Ing. A. Hartwich (AVT-PT), sowie bei Dr. K.-H. Brakhage und Dr. S. Müller (IGPM) für wertvolle Unterstützung bei der Zusammenstellung dieser Arbeit bedanken.

6 LITERATUR

- [1] Abramovich, F. und B.W. Silvermann (1998). Wavelet decomposition approaches to statistical inverse problems. *Biometrika* **85**, 115–129.
- [2] Alsmeyer, F. (2005). Trend-based treatment of process data: application to practical problems. In: *Proc. of the World Congress of Chemical Engineering 2005, Glasgow, UK, July 10–14*.
- [3] Alsmeyer, F. (2006). Automatic adjustment of data compression in process information management systems. In: *Proc. of the 16th European Symposium on Computer Aided Process Engineering and 9th International Symposium on Process Systems Engineering, Garmisch-Partenkirchen*. pp. 1533–1538.
- [4] Biegler, L.T. (1984). Solution of dynamic optimization problems by successive quadratic programming and orthogonal collocation. *Comp. Chem. Eng.* **8**, 243–248.
- [5] Binder, T., A. Cruse, C. Villar und W. Marquardt (2000). Dynamic optimization using a wavelet based adaptive control vector parameterization strategy. *Comput. Chem. Eng.* **24**, 1201–1207.
- [6] Binder, T., L. Blank, H.G. Bock, R. Burlisch, W. Dahmen, M. Diehl, T. Kronseder, W. Marquardt, J.P. Schlöder und O. v. Stryk (2001a). *Introduction to Model Based Optimization of Chemical Processes on Moving Horizons*. pp. 295–339. Online Optimization of Large Scale Systems. Springer-Verlag Berlin Heidelberg.

- [7] Binder, T., L. Blank, W. Dahmen und W. Marquardt (1998). Towards Multiscale Dynamic Data Reconciliation. In: *NATO-ASI on "Nonlinear Model Based Process Control", Antalya, Turkey, 10-20. 8. 1997* (C. Kravaris R. Berber, Ed.). Kluwer Academic Publishers. pp. 623-665.
- [8] Binder, T., L. Blank, W. Dahmen und W. Marquardt (2001b). Iterative algorithms for multiscale state estimation, Part 1: Concepts. *J. Optim. Theor. Appl.* **111**(3), 501-527.
- [9] Binder, T., L. Blank, W. Dahmen und W. Marquardt (2001c). Iterative algorithms for multiscale state estimation, Part 2: Numerical investigations. *J. Optim. Theor. Appl.* **111**(3), 529-551.
- [10] Binder, T., L. Blank, W. Dahmen und W. Marquardt (2002). On the Regularization of Dynamic Data Reconciliation Problems. *J. Process Cont.* **12**(4), 557-567.
- [11] Brakhage, K.-H. und P. Lamby (2007). Application of B-spline techniques to the modeling of airplane wings. Submitted to CAGD.
- [12] Brakhage, K.-H. und P. Lamby (2007). Modeling of airplane wings with winglets. In: *10th International Conference on Numerical Grid Generation in Computational Field Simulations*, Sep. 16-20, 2007, FORTH, Greece.
- [13] Bramkamp, F., B. Gottschlich-Müller, Ph. Lamby, M. Hesse, S. Müller, Ballmann, J., K.-H. Brakhage, und W. Dahmen (2003). *H*-adaptive multiscale schemes for the compressible Navier-Stokes equations - Polyhedral discretization, data compression and mesh generation. In: *Notes on Numerical Fluid Mechanics, Flow Modulation and Fluid-Structure-Interaction at Airplane Wings* (ed. by J. Ballmann), Springer-Verlag, **84**, 125-204.
- [14] Cao, S. und R.R. Rhinehart (1995). An efficient method for on-line identification of steady state. *J. Proc. Contr.* **5**, 363-374.
- [15] Cohen, A., W. Dahmen, und R. DeVore (2003). Adaptive Wavelet Schemes for Nonlinear Variational Problems, *SIAM J. Numer. Anal.*, **41**, 1785-1823.
- [16] Cohen, A., W. Dahmen, und R. DeVore (2004). Adaptive wavelet techniques in Numerical Simulation. In: *Encyclopedia of Computational Mechanics*, (R. De Borste, T. Hughes, E. Stein, eds.), Wiley-Interscience, 157-197.
- [17] Cohen, A., I. Daubechies, und J.-C. Feauveau (1992). Biorthogonal bases of compactly supported wavelets. *Comm. Pure and Appl. Math.*, **45**, 485-560.

- [18] Cohen, A., S. Mahmoud Kaber, S.Müller, M. Postel (2003). Fully Adaptive Multiresolution Schemes for Conservation Laws. *Mathematics of Computation*, **72**, 183–225.
- [19] Cohen, A., W. Dahmen, I. Daubechies, und R. DeVore (2001). Tree approximation and optimal encoding, *Applied and Computational Harmonic Analysis*, **11**, 192–226.
- [20] Dahmen, W. und C.A. Micchelli (1993). Using the refinement equation for evaluating integrals of wavelets, *SIAM Journal on Numerical Analysis*, **30**, 507-537.
- [21] Dahmen, W., B. Gottschlich-Müller, und S. Müller (2001). Multiresolution schemes for conservation laws, *Numerische Mathematik*, **88**, 399-443.
- [22] Donoho, D.L. und I.M. Johnstone (1995). Adapting to unknown smoothness with wavelet shrinkage. *J. Am. Stat. Assoc.* **90**, 1200–1224.
- [23] Dünnebier, G., D. van Hessem, J. Kadam, K.-U. Klatt und M. Schlegel (2005). Optimization and control of polymerization processes. *Chemical Engineering Technology* **28**, 575–580.
- [24] Flehmig, F. und W. Marquardt (2006). Detection of multivariable trends in measured process quantities. *J. Process Cont.* **16**, 947–957.
- [25] Foss, A. S. (1973). Critique of chemical process control theory. *AIChE J.* **19**, 209–214.
- [26] Gertler, J. und H.S. Chang (1986). An instability indicator for expert control. *IEEE Control Systems Magazine* **6**, 14–17.
- [27] Hartwich, A., M. Schlegel, L. Würth und W. Marquardt (2008). Adaptive control vector parameterization for nonlinear model-predictive control. *International Journal of Robust and Nonlinear Control* **18**(8), 845–861.
- [28] Lamby, P. (2007). Parametric Multi-Block Grid Generation and Application to Adaptive Flow Simulation. *PhD Dissertation*, RWTH Aachen.
- [29] Massjung, R., J. Hurka, J. Ballmann, und W.,Dahmen (2003). On well-posedness and modeling for nonlinear aeroelasticity. In *Notes on Numerical Fluid Mechanics*, Springer Verlag, **84**.

- [30] Massjung, R. (2002), Numerical Schemes and Well-Posedness in Nonlinear Aeroelasticity. *PhD Thesis, RWTH Aachen*.
- [31] Massjung, R. (2006). Discrete conservation and coupling strategies in nonlinear aeroelasticity. *Comp. Meth. Applied Mech. Eng.*, **196**, 91–102.
- [32] Müller, S. (2002), Adaptive Multiscale Schemes for Conservation Laws (2002). *Lecture Notes on Computational Science and Engineering*, **27**, Springer.
- [33] Schlegel, M. und W. Marquardt (2006a). Adaptive switching structure detection for the solution of dynamic optimization problems. *Ind. Eng. Chem. Res.*, 2006, **45**(24), 8083–8094
- [34] Schlegel, M. und W. Marquardt (2006b). Detection and exploitation of the control switching structure in the solution of dynamic optimization problems. *J. Proc. Control* **16**(3), 1731–1751.
- [35] Schlegel, M., K. Stockmann, T. Binder und W. Marquardt (2005). Dynamic optimization using adaptive control vector parameterization. *Comp. Chem. Eng.* **29**, 1731–1751.
- [36] Schlegel, M., W. Marquardt, R. Ehrig und U. Nowak (2004). Sensitivity analysis of linearly-implicit differential-algebraic systems by one-step extrapolation. *Appl. Num. Math.* **48**, 83–102.
- [37] Strang, G. und T. Nguyen (1996). *Wavelets and Filter Banks*. Wellesley-Cambridge Press, Wellesley, MA, USA.
- [38] Tsang, T.H., D.M. Himmelblau und T.F. Edgar (1975). Optimal control via collocation and nonlinear-programming. *Int. J. Control* **21**, 763–768.
- [39] Watson, M.J.A., A. Liakopoulos, D. Brzakovic und C. Georgakis (1998). A practical assessment of process data compression techniques. *Ind. Eng. Chem. Res.* **37**, 267–274.

> AKTIVE STRÖMUNGSBEEINFLUSSUNG – EINE MATHEMATISCHE HERAUSFORDERUNG

RUDIBERT KING, VOLKER MEHRMANN UND WOLFGANG NITSCHKE

1 EXECUTIVE SUMMARY

Die nachhaltige Entwicklung von Straßen- und Luftverkehrsträgern erfordert einen deutlich vermindertem Energieverbrauch und signifikant niedrigere Schadstoff- und Lärmemissionen. Um diesen Herausforderungen gerecht zu werden, sind nichtgraduelle Verbesserungen notwendig. Eine zentrale Rolle bei der Bewältigung dieser Zukunftsaufgaben spielt die aktive Beeinflussung der auftretenden Strömungsprozesse. Während in der Vergangenheit Verbesserungen im Wesentlichen durch passive Beeinflussungen erzielt wurden, lassen sich die für die Zukunft geforderten sprunghaften Leistungssteigerungen nur noch mit aktiven Maßnahmen erzielen. Charakteristisch bei dieser Entwicklung ist die intensive interdisziplinäre Zusammenarbeit von zahlreichen Wissenschaftsgebieten.

Dieser Artikel skizziert den Stand der Forschung in der aktiven Strömungsbeeinflussung, stellt die auftretenden Herausforderungen dar und beschreibt die notwendigen Maßnahmen, um sich diesen Herausforderungen zu stellen.

Der Beitrag der Mathematik ist dabei sehr vielschichtig, er reicht von der mathematischen Modellierung der Strömungsprozesse, deren numerischer Simulation auf modernen Höchstleistungsrechnern, der Modellreduktion, der Entwicklung von Verfahren zur modellbasierten Regelung oder Optimierung der Prozesse bis hin zur merkmalsbasierten Visualisierung immenser experimenteller oder numerischer Datensätze.

Während die mathematische Modellierung von Strömungsprozessen schon seit über 100 Jahren ein wichtiges Forschungsgebiet ist, hat das Thema der numerischen Strömungssimulation erst in der letzten 30 Jahren durch die Rechnerentwicklung einen großen Schub erfahren. Trotz aller Fortschritte sind die Rechenzeiten für die Simulation von Strömungen um reale Fahrzeuggeometrien noch immer exorbitant hoch. Daher ist eine weitere Beschleunigung der Simulationsmethoden von zentraler Bedeutung.

Für eine aktive Strömungsbeeinflussung im Realbetrieb sind zudem Reaktionszeiten notwendig, die auf Einfachprozessoren im Millisekundenbereich liegen. Daher ist es einerseits notwendig, die mathematischen Modelle durch vereinfachte (reduzierte) Modelle zu ersetzen, und andererseits Algorithmen aufzustellen, wie die Informationen, die

mittels Sensormessungen gewonnen werden, in die Regelung/Optimierung einbezogen werden können.

Für die ökonomische und ökologische Weiterentwicklung des Hochtechnologiestandorts Deutschland ist eine Intensivierung der Forschungsförderung im Gebiet der Strömungsbeeinflussung von zentraler Bedeutung.

2 ERFOLGSGESCHICHTEN

Die Untersuchung von turbulenten Strömungsprozessen ist eine zentrale Fragestellung bei der Entwicklung und beim Betrieb von modernen Verkehrsflugzeugen, aber auch bei Schienen- und Straßenfahrzeugen. Neben der experimentellen Untersuchung durch Windkanalexperimente und der theoretischen physikalischen Analyse von Modellgleichungen, hat sich die numerische Simulation von Strömungen in den letzten 30 Jahren als dritte Säule zur Analyse instationärer Strömungsprozesse fest etabliert. Auf der Basis der Modellgleichungen bzw. sehr starken Vereinfachungen dieser Gleichungen, die durch physikalisch oder mathematisch basierte Modellreduktionsmethoden gewonnen werden, können dann mathematische Regelungs- bzw. Optimierungsmethoden entwickelt und angewendet werden.

Mit modellgestützten Regelungsmethoden der Strömungsbeeinflussung kann man heute den Lärm von Turbinen deutlich dämpfen, den Auftrieb von Tragflügeln erhöhen, den Widerstand von Fahrzeugen verringern und die Vermischung von Luft und Kraftstoff in Verbrennungssystemen so gestalten, dass weniger Schadstoffe entstehen.

Bläst man beispielsweise aus der Vorderkante einer Landeklappen eines Flugzeuges, siehe das Schema in Abb. 1, periodisch geringe Mengen Luft in die Strömung ein, so lässt sich der Auftrieb deutlich erhöhen. In Abb. 2 ist exemplarisch die erreichte Auftriebssteigerung dargestellt. Hierdurch sind zukünftig steilere Start- und Landevorgänge möglich, wodurch die Lärmbelastung am Boden sinkt.

Die große Herausforderung, die sich heute stellt, ist die Übertragung solcher Konzepte auf reale, geometrisch komplex geformte Flügel mit Anbauten, wie den Triebwerks gondeln. Erste Schritte in dieser Richtung werden zur Zeit im DFG-Sonderforschungsbereich 557 *Beeinflussung komplexer turbulenter Scherströmungen* an der TU Berlin in Kooperation mit Airbus Deutschland unternommen.

Ein anderes erfolgreiches Beispiel für die aktive Strömungsbeeinflussung ist die Reduktion von schädigenden Druckschwankungen in einem Brenner, z. B. einer Turbine, siehe Abb. 3, die bei schadstoffarmem Betrieb auftreten können. [8]. Durch sekundäre Einspritzung können die Druckschwankungen deutlich reduziert werden, siehe Abb. 4.

3 AKTIVE BEEINFLUSSUNG VON STRÖMUNGEN, STATUS QUO

Die aktive Strömungsbeeinflussung spielt eine zentrale Rolle für die nachhaltige Entwicklung von Straßen- und Luftverkehrsträgern. Die in der „Vision 2020“ der europäischen

Abb. 1: Versuchsaufbau

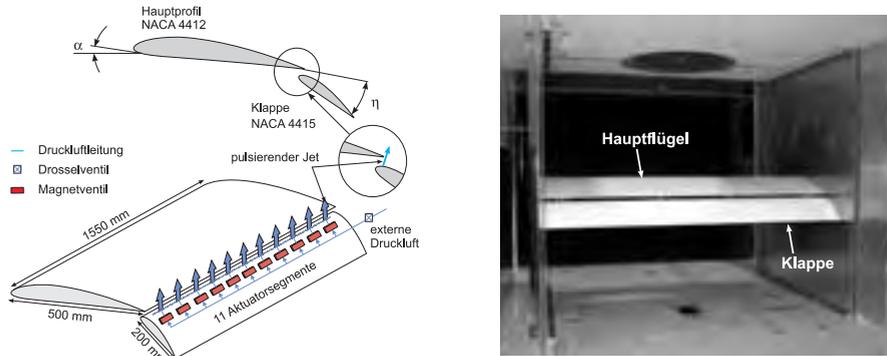


Abb. 2: Normierter Auftrieb eines Flügels bei Veränderung des Anstellwinkels

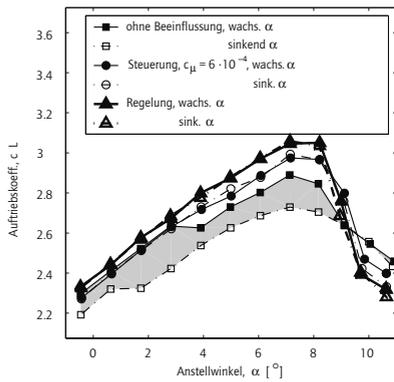


Abb. 3: Druckschwankungen in einer Turbinen-Brennkammer

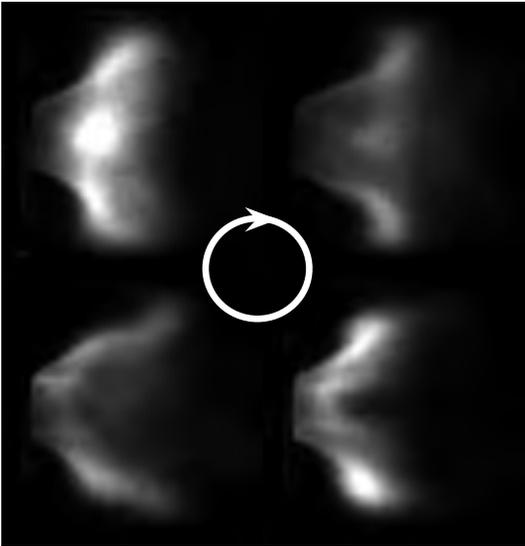
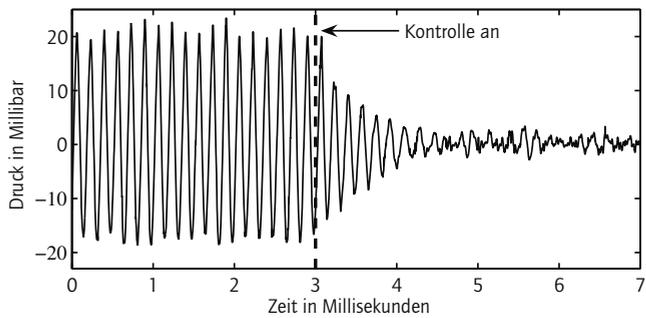


Abb. 4: Beeinflussung von Druckschwankungen in einer Brennkammer



Luftfahrtinitiative ACARE [1] geforderten Verringerungen von Emissionen, Treibstoffverbrauch und Lärm trotz erwarteter Verdreifachung des Lufttransportes in den nächsten 15 Jahren lassen sich nicht mehr durch graduelle Verbesserungen, beispielsweise mit Turbulatoren auf Flügeln, erreichen. Vielmehr sind nichtgraduelle Leistungssteigerungen gefordert, wie sie die aktive Manipulation von Scherschichten ermöglicht.

Hierzu ist eine intensive interdisziplinäre Zusammenarbeit von verschiedenen wissenschaftlichen Teildisziplinen notwendig (Experiment, Regelungstechnik, Sensorik, Aktuatorik, Strömungssimulation, Analysis, Numerische Mathematik, Mathematische Visualisierung, Wissenschaftliches Rechnen).

Die Beiträge der Mathematik, z. B. im Entwicklungszyklus von modernen Flugzeugen, werden dabei durch die folgende tabellarische Aufstellung verdeutlicht.

- Entwicklung und Untersuchung der mathematischen Modellgleichungen zur Beschreibung der turbulenten Strömung in Verbindung mit Experimenten.
- Darstellung (idealerweise 3D) eines virtuellen mathematischen Modells des physikalischen Modells im Rechner.
- Entwicklung und Analyse von numerischen Methoden zur Simulation bzw. Visualisierung der Strömung um das im Rechner erzeugt virtuelle Modell.
- Simulationen an realen Testkonfigurationen, Vergleich mit Experimenten, Verbesserung der mathematischen Modelle und der numerischen Simulationsmethoden.
- Systematische Veränderung/Optimierung von Designparametern im mathematischen Modell, zur passiven Beeinflussung anhand physikalischer Experimente bzw. numerischer Simulationen.
- Entwicklung von approximativen, reduzierten mathematischen Modellen für die aktive Beeinflussung der Strömung.
- Entwicklung von (echzeitfähigen) mathematischen Methoden zur Regelung der Strömung.

Dieser komplexe Entwicklungsprozess ist natürlich nicht statisch, sondern es werden durch Rückkopplung und Vergleich mit Experimenten kontinuierlich Veränderungen am verwendeten mathematischen Modell, sowie an den numerischen Methoden zur Simulation bzw. Regelung und Optimierung vorgenommen. Diese sehr grundlagenorientierte Forschung erfordert eine sehr intensive interdisziplinäre Zusammenarbeit, wie sie in verschiedenen von der DFG geförderten Sonderforschungsbereichen stattfindet (so zum Beispiel im SFB 401 *Strömungsbeeinflussung und Strömungs-Struktur-Wechselwirkung an Tragflügeln* an der RWTH Aachen oder im SFB 557 *Beeinflussung von komplexen turbulenten Scherströmungen* an der TU Berlin).

3.1 MODELLIERUNG

Eine typische Klasse von mathematischen Modellgleichungen zur Beschreibung von turbulenten Strömungen um eine Tragfläche sind die instationären Navier-Stokes Gleichungen

$$\left. \begin{aligned} \frac{\partial}{\partial t} v + \nabla \cdot (v \otimes v) + \frac{1}{\rho_f} \nabla p - \nu \Delta v &= 0 \\ \nabla \cdot v &= 0 \end{aligned} \right\} \text{ in } \Omega, \quad (1)$$

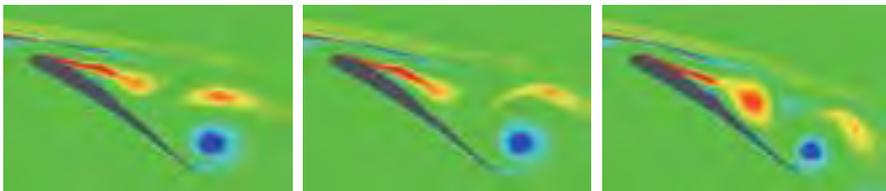
bei denen v die Geschwindigkeit der Strömung und p den Druck in einem Berechnungsgebiet Ω des dreidimensionalen Raums bezeichnet. Dabei ist ρ_f die Dichte und ν die kinematische Viskosität der Luft. Die Größe $1/\nu$ bezeichnet man allgemein als Reynoldszahl Re . Dazu kommen noch Rand- und Anfangsbedingungen.

Will man nun die Strömung beeinflussen, in dem man an einer Stelle der Tragfläche Luft ausbläst bzw. angesaugt wie in Abb. 1 dargestellt, so führt dies zu zeitveränderlichen Randbedingungen. Auf der Basis dieses Modells kann nun in der Simulation die Intensität bzw. die Frequenz, mit der Luft ausgeblasen wird, verwendet werden, um den Auftrieb zu erhöhen, siehe Abb.2, oder um abgerissene Strömung wieder anzulegen, siehe Abb. 5 aus [11].

Bei sehr schnellen Strömungen mit viel Turbulenz (dies entspricht sehr großen Reynoldszahlen von 100 000 oder mehr), ist es auch auf heutigen Hochleistungsrechnern nicht möglich, das Strömungsverhalten mit hochauflösenden Verfahren, wie der direkten numerischen Simulation (DNS) oder der Large Eddy Simulation (LES) zu simulieren, geschweige denn die Optimierung der Strömung auf der Basis dieser Methoden zu beschreiben. Man würde für eine 3D-Simulation ein Berechnungsgitter mit ca. 10 Milliarden Knoten (siehe z. B. [6]) benötigen und müsste sehr viele nichtlineare Gleichungssysteme dieser Größe lösen.

Als Alternative zu den hochauflösenden Methoden werden daher heute vereinfachte Modelle wie die Reynolds-gemittelten Navier-Stokes-Gleichungen, typischerweise gekoppelt mit vereinfachten statistischen Turbulenzmodellen, verwendet. Durch die Mittelung werden nicht mehr alle Einzelheiten aufgelöst, aber das qualitative Verhalten der

Abb. 5: Strömungsfeld für verschiedene Anregungsintensitäten



Strömung kann erfasst werden. Solche Techniken sind zur Zeit in fast allen universitären oder kommerziellen Strömungscodes realisiert.

Allerdings sind auch diese Methoden wegen der langen Rechenzeiten heutzutage nur schwer in der Optimierung bzw. Regelung einsetzbar. Realistische Methoden (siehe z. B. [11]) benutzen daher hybride Ansätze, bei denen nur in kleinen physikalisch relevanten Teilbereichen des Rechengebiets (wie zum Beispiel in der Nähe der Tragfläche eines Flugzeugs) hochauflösenden Methoden wie die direkte numerische Simulation verwendet werden, während im restlichen Gebiet gemittelte Modelle und vereinfachte Verfahren zum Einsatz kommen.

Ein Beispiel für die Umsetzung dieser Methoden bei der Strömungsbeeinflussung zeigt die Abb. 6 aus [11].

Ein anderer mathematisch motivierter Forschungsansatz, der ein großes Potential in Richtung Regelung/Steuerung hat, ist die adaptive Lösung der instationären Strömungsgleichungen. In diesem Prozess wird das Gitter, welches für die numerische Simulation verwendet wird, während der Simulation kontinuierlich an das Verhalten der Lösung angepasst. Durch Schätzung des bei der Simulation gemachten Fehlers wird nur dort, wo eine hohe Auflösung benötigt wird, diese auch erzielt, siehe z. B. Abb. 7 aus [5].

Ein dritter, sehr aktueller Ansatz der Forschung ist die Approximation des hochdimensionalen mathematischen Modells (d. h. des semidiskretisierten Modells mit 10 Milliarden Gleichungen) durch sehr vereinfachte niederdimensionale Modelle mit nur einigen wenigen, z. B. zehn Gleichungen, die die für die Optimierung/Regelung essentiellen Größen erfassen. Diese Aufgabe der Modellreduktion ist heute nicht nur in der Strömungsbeeinflussung, sondern in fast allen Bereichen von Naturwissenschaft und Technik ein zentrales Forschungsthema [2, 4]. Ein wesentlicher Beitrag, den die mathematische Forschung hier leistet bzw. zu leisten hat, ist die Entwicklung und Analyse neuer Modellreduktionsmethoden, insbesondere die Konstruktion beweisbarer Fehlerschranken.

Abb. 6: Geregelte Strömung

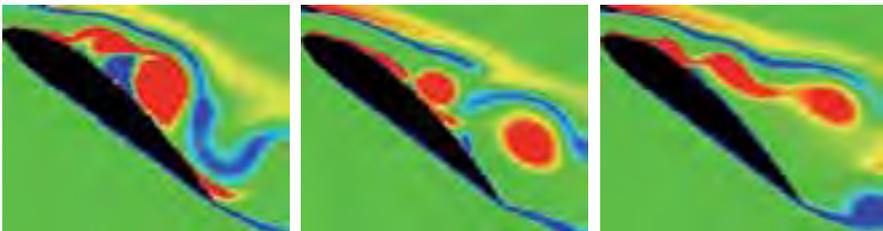


Abb. 7: Adaptive Gitterverfeinerung

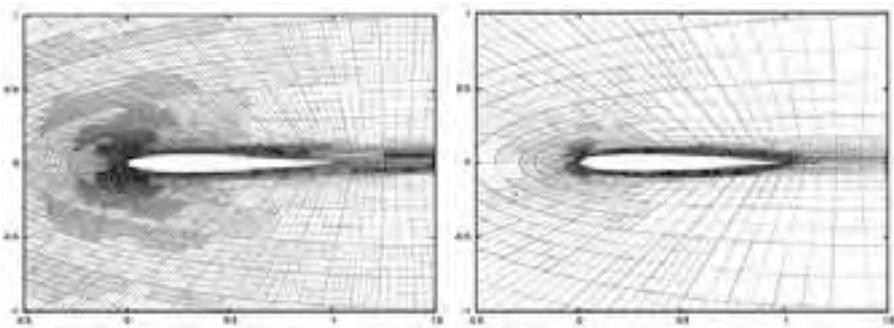
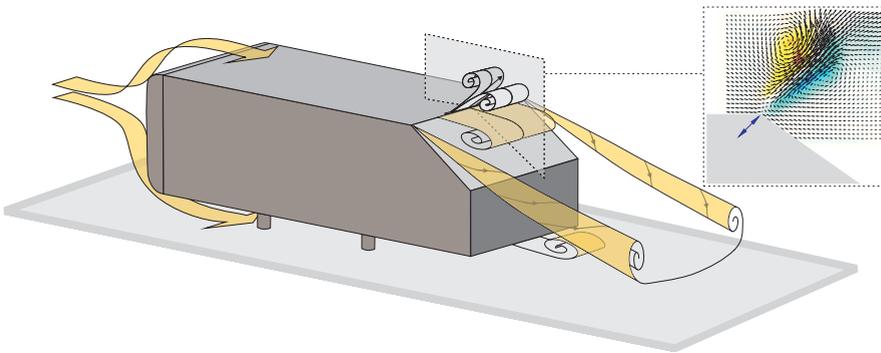


Abb. 8: Vereinfachtes Fahrzeugmodell

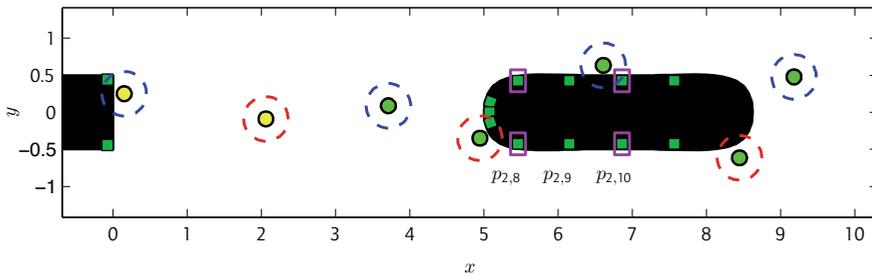


Mit niederdimensionalen mathematischen Modellen ist es heute z. B. möglich, bei hintereinanderfahrenden, sehr einfachen Fahrzeugmodellen (siehe Abb. 8) aus Druckmessungen an der Gehäusewand in Echtzeit auf große Wirbel in der Strömung zurückzurechnen, die für eine unregelmäßige Fahrt und damit für einen verminderten Fahrkomfort verantwortlich sind (siehe Abb. 9). Mit dieser Kenntnis kann im nächsten Schritt gezielt in die Strömung eingegriffen werden.

4 STÄRKEN/SCHWÄCHEN-ANALYSE, HERAUSFORDERUNGEN

Die Modellierung, Simulation, Optimierung und Regelung von komplexen Strömungsprozessen ist für die ökonomische und ökologische Entwicklung von zentraler Bedeutung.

Abb. 9: Schätzung von Wirbelstrukturen



Die für die Zukunft notwendigen Leistungssteigerungen lassen sich nur noch mit aktiven Maßnahmen erzielen, die zentral auf modellbasierten Regelungen/Optimierungen beruhen. Die gegenwärtigen Modelle und sowohl die darauf beruhenden direkten Simulationsmethoden als auch die vereinfachten hybriden Methoden sind jedoch zur Zeit nicht echtzeitfähig und daher nicht direkt zur Strömungsbeeinflussung geeignet.

Von zentraler Bedeutung ist daher die Entwicklung von schnelleren und effizienteren Methoden zur Strömungssimulation, die jedoch gleichzeitig das Potential besitzen, mit geeigneten Modellreduktionsmethoden kombiniert zu werden. Dies ist bei vielen kommerziellen Strömungslösern heute nicht möglich. Es müssen weiterhin Methoden entwickelt werden, die eine automatische Adaptierung der Simulationsalgorithmen auf das Ziel der Regelung/Optimierung hin erlauben. Ansätze in diese Richtung sind z. B. in [12] beschrieben.

Ein anderer Ansatz ist die mathematische Weiterentwicklung von effizienten und mit Fehlerschranken versehenen Modellreduktionsmethoden bzw. die direkte Entwicklung von neuen niederdimensionalen Strömungsmodellen, die für die echtzeitfähige Regelung/Optimierung geeignet sind.

Daneben sind bessere echtzeitfähige numerische Regelungs- bzw. Optimierungsmethoden zu entwickeln, die auf der Basis niederdimensionaler Modelle in Echtzeit für reale Tragflügel mit Anbauten geeignet sind und die flexibel mit den Simulations- bzw. Sensordaten arbeiten können.

5 VISIONEN UND HANDLUNGSEMPFEHLUNGEN

Die aktive Strömungsbeeinflussung ist ein Paradebeispiel für die zukünftige wissenschaftlich technische Entwicklung, bei der interdisziplinäre Zusammenarbeit zur Lösung der vorliegenden Herausforderungen eine direkte gesamtgesellschaftliche Bedeutung hat. Die mathematische Forschung stellt hier eine wesentliche Kernkomponente dar.

Eine Intensivierung der interdisziplinären Grundlagenforschung und des Transfers von Grundlagenergebnissen in die industrielle Praxis ist in diesem Gebiet unbedingt notwendig.

Um die großen Herausforderungen bewältigen zu können, werden zudem dringend NachwuchswissenschaftlerInnen benötigt, die in der Lage sind, in interdisziplinären Teams zu arbeiten. Dazu müssen interdisziplinäre Studienprogramme (Master/Promotion) initiiert werden, die diesen Bedürfnissen Rechnung tragen.

6 LITERATUR

- [1] European aeronautics: A vision for 2020. Meeting societies needs and winning global leadership. Report der europäischen Luftfahrtinitiative ACARE. Europäische Kommission 2001.
- [2] A. Antoulas. *Approximation of large-scale dynamical systems*. SIAM, Philadelphia, 2005.
- [3] R. Becker, R. King, R. Petz, and W. Nitsche. *Adaptive closed loop separation control on a high-lift configuration using extremum seeking*, AIAA-Journal, 45(6):1382–1392, 2007.
- [4] P. Benner, V. Mehrmann, and D. C. Sorensen, (Hrsg.). *Dimension reduction of large-scale systems*. Lecture Notes in Computational Science and Engineering Vol. 45, Springer Verlag, Heidelberg, 2005.
- [5] W. Bangerth and R. Hartmann und G. Kanschat. *Deal.II – a general-purpose object-oriented finite element library* ACM Trans. Math. Softw., 33(4), 2007.
- [6] M. Griebel, T. Dornseifer und T. Neunhoffer. *Numerical simulation in fluid dynamics*. SIAM, Philadelphia, 1998.
- [7] L. Henning, D. Kuzmin, V. Mehrmann, M. Schmidt, A. Sokolov und S. Turek. *Flow Control on the basis of FEATFLOW–MATLAB coupling*. R. King (Hrsg), Active Flow Control Notes on Numerical Fluid Mechanics and Multidisciplinary Design, 95, Springer Verlag, Berlin, pp. 325–338, 2007.
- [8] J. P. Moeck, M. R. Bothien, D. Guyot und C. O. Paschereit. *Phase-shift control of combustion instability using (combined) secondary fuel injection and acoustic forcing*. R. King (Hrsg), Active Flow Control Notes on Numerical Fluid Mechanics and Multidisciplinary Design, 95, Springer Verlag, Berlin, pp. 408–421, 2007.

- [9] R. King (Hrsg.), *Active flow control* Notes on Numerical Fluid Mechanics and Multidisciplinary Design, 95, Springer Verlag, Berlin, 2007.
- [10] R. Petz, W. Nitsche. *Active separation control on the flap of a two-dimensional generic high-lift configuration*. Journ. of Aircraft 44(3): 865-874, 2007.
- [11] M. Schatz, B. Günther und F. Thiele. *Computational investigation of separation for high lift airfoil flows*. R. King (Hrsg), Active Flow Control, Notes on Numerical Fluid Mechanics and Multidisciplinary Design, 95, Springer Verlag, Berlin, pp. 173-189, 2007.
- [12] M. Schmidt. *Systematic discretization of input/output maps and other contributions to the control of distributed parameter systems with a focus on fluid control*. Vdm Verlag Dr. Müller, Saarbrücken, 2008.

> DATA-MINING FÜR DIE ANGEBOTSOPTIMIERUNG IM HANDEL

JOCHEN GARCKE, MICHAEL GRIEBEL UND MICHAEL THESS

1 EXECUTIVE SUMMARY

Der Handel ist weltweit durch gravierenden Wettbewerbsdruck gekennzeichnet. Deutschland gilt im Einzelhandel sogar als der wettbewerbsintensivste Markt der Welt. Um in diesem Wettbewerb zu bestehen, vertrauen die meisten Händler auf extensive Formen *ungezielten Massenmarketings*. Hierbei werden alle potentielle Kunden mit den gleichen riesigen Katalogen, ungezählten Werbebroschüren, aufdringlichen Lautsprecherdurchsagen und schrillen Bannerwerbungen bearbeitet. Im Ergebnis sind nicht nur die Kunden genervt, sondern auch die Rücklaufquoten von Marketingkampagnen seit Jahren fallend. Um dies zu vermeiden, ist eine *Individualisierung des Massenmarketings* empfehlenswert, in der Kunden auf sie zugeschnittene, individuelle Angebote erhalten. Die Aufgabe besteht darin, *dem richtigen Kunden zum richtigen Zeitpunkt zum richtigen Preis das richtige Angebot* zu unterbreiten. Dies stellt sich primär als eine mathematische Aufgabenstellung heraus, die die Bereiche Statistik, Optimierung, Analysis und Numerik betrifft. Die hierbei auftretenden Aufgabenstellungen der Regressionsanalyse, des Clusterings und der Optimalen Steuerung sind durch hohe Dimensionen und riesige Datenmengen gekennzeichnet und erfordern neue mathematische Konzepte und Verfahren.

Die solchen Ansätzen zu Grunde liegende (semi)-automatisierte Wissensentdeckung durch Analyse großer Datenmengen hat sich in den letzten Jahren zum Anwendungs- und Forschungsgebiet des Data-Mining entwickelt. Mit dem Begriff des Maschinellen Lernens wird dabei dessen algorithmischer Kern bezeichnet. Dieser Bereich war ursprünglich eher in der Informatik angesiedelt. Darüber hinaus hat die Verbindung zur Statistik von Anfang an eine wesentliche Rolle gespielt. In den letzten Jahren sind weitere mathematische Aspekte gerade in der Forschung verstärkt behandelt worden. Hier sei beispielsweise der Bereich der statistischen Lerntheorie erwähnt. Verfahren mit einem solchen Hintergrund werden mittlerweile nicht zuletzt auf Grund ihrer reichhaltigen mathematischen Unterfütterung in vielen Anwendungsbereichen erfolgreich eingesetzt.

Unter der Annahme, dass ähnliche Kundendaten für ein ähnliches Kundenverhalten stehen, ist ein Einschätzen von neuen Kunden auf Basis des Verhaltens von bisherigen

Kunden möglich. Grundlegend ist hierbei, dass viele moderne maschinelle Lernansätze die Darstellung von Funktionen über hochdimensionalen Merkmalsräumen beinhalten. Die nun mögliche gekoppelte nichtlineare Behandlung von verschiedenen Merkmalen wie beispielsweise Einkommen, Schuldenstand, Zahl der Kinder, oder Art des Autos erlauben deutlich verbesserte Schätzungen des wahrscheinlichen Kundenverhaltens. Die Approximationstheorie und die Numerik leisten deswegen bereits jetzt bei Neu- und Weiterentwicklung von maschinellen Lernverfahren einen substantiellen Beitrag. Dies wird sich in der Zukunft weiter verstärken.

Mittlerweile ist die Numerik für moderat höherdimensionale Probleme in einer ersten Stufe so weit entwickelt, dass sie auch für erste praxisrelevante Aufgaben des Data-Mining bei der Angebotsoptimierung im Handel Verwendung findet. Darüber hinaus ist jedoch insbesondere für wirklich hochdimensionale Probleme weitergehende Forschung und Entwicklung notwendig um noch mehr Merkmale und damit Dimensionen effizienter darstellen zu können. Hier erscheint es dringlich geboten speziell zugeschnittene Forschungsprogramme aufzulegen. Hierdurch könnte es mittelfristig gelingen die Methodenentwicklung so weit voranzutreiben, dass insbesondere vollständig automatisierte interaktive Systeme zur Angebotsoptimierung zur Serienreife gebracht werden. Dies würde auch helfen, die internationale Wettbewerbsfähigkeit Deutschlands im Handel zu stärken.

Dieser Artikel beschreibt die Rolle der Mathematik für die Angebotsoptimierung im Handel, insbesondere hinsichtlich der Approximationstheorie und Numerik. Dabei werden sowohl der aktuelle Stand und bisherige Erfolge, als auch neue Entwicklungen und Herausforderungen dargelegt.

2 ANGEBOTSOPTIMIERUNG IM HANDEL: ÜBERSICHT UND STATUS QUO

Im Handel ist die Dynamik des Marktes besonders groß. Das liegt zum einen daran, dass die Produkte weitgehend vergleichbar sind, da alle Händler insbesondere über den Großhandel Zugang zu einem weitgehend gleichen Sortiment haben. Zum anderen hat das Internet zahlreiche neue Geschäftsmodelle hervorgebracht und damit die Internationalisierung und den Wettbewerbsdruck im Handel drastisch verschärft. Insgesamt liegen umfassende – wenngleich auch zumeist anonyme – Daten über das Verhalten von Kunden vor, die bei der Optimierung von Angeboten genutzt werden können.

Die Aufgabenstellungen in der Angebotsoptimierung im Handel lassen sich im Wesentlichen in folgende Teilgebiete untergliedern:

- Kampagnenoptimierung (d. h. Auswahl der Zielgruppen und -personen),
- Cross- und Up-Selling (d. h. Zusatzverkäufe an Kunden),
- Sortimentsgestaltung (d. h. Produktangebot und -kategorien),
- Preisoptimierung (d. h. Optimierung der Produktpreise und Promotions).

Seit einigen Jahren setzen Händler zur Untersuchung des Kundenverhaltens zunehmend mathematische Verfahren in verschiedenen Bereichen des Marketings ein, doch ist deren Nutzung noch zaghaft und sowohl von der Branche als auch der Marketingaktivität abhängig. Während im *E-Business* eine beachtliche Durchdringung von – allerdings oft noch unausgereiften – Verfahren zu verzeichnen ist, wird im *Versandhandel* Mathematik hauptsächlich zur Optimierung von Mailing-Kampagnen eingesetzt, also der Auswahl von Kunden mit hoher Reaktionswahrscheinlichkeit auf spezielle Angebote. Schlusslicht ist schließlich der *stationäre Handel*; doch eröffnet gerade die hier stattfindende technologische Revolution der interaktiven digitalen Einkaufshilfen neue interessante Möglichkeiten für die Entwicklung und den Einsatz neuer mathematischer Verfahren für die Angebotsoptimierung.

Günstig für die Verwendung mathematischer Verfahren ist die Tatsache, dass der Handel durch ein hohes Maß an Kundeninteraktion geprägt ist. Somit liegen umfangreiche Daten über das Kundenverhalten vor. Mathematik hat sich hier zunächst in einigen klassischen Data-Mining-Feldern bewährt. So werden für die Mailingoptimierung Klassifikationsverfahren auf breiter Front eingesetzt. Zunehmend erfolgreich ist auch die Nutzung von Clusterverfahren für die Kundensegmentierung bei der Erstellung von Themenkatalogen. Andere Bereiche wie Echtzeitanalyse und -angebote setzen bisher zumeist nur einfachste Techniken ein. Schließlich ist im strategischen Bereich des Warengruppenmanagements, also für Aufgaben der Sortiments- und Preisoptimierung, der Einsatz moderner mathematischer Instrumente noch die Ausnahme. Gerade hier aber liegt – in Kombination mit Echtzeitanalysen der Optimalen Steuerung – ein wichtiges zukünftiges Betätigungsfeld für die interdisziplinäre Zusammenarbeit von Betriebswirten, Informatikern und Mathematikern. Weitere Informationen zu Data-Mining Methoden im Handel, Marketing und Customer Relationship Management findet man etwa in [2] oder [10].

Im Folgenden wollen wir exemplarisch auf die ersten beiden Aufgabenstellungen, d. h. die Kampagnenoptimierung und das Cross- und Up-Selling, und dort vorhandene Erfolgsgeschichten bei der Nutzung mathematischer Verfahren näher eingehen.

2.1 KAMPAGNEOPTIMIERUNG

Die Kampagnenoptimierung ist hinsichtlich der Nutzung der Mathematik am weitesten fortgeschritten. Dabei geht es im wesentlichen darum, Werbekampagnen zielgruppengenau umzusetzen. Dies umfasst die Definition der Ziele und des Ablaufs der Kampagne sowie deren Auswertung. Dabei wird zwischen Zielgruppen (Segmentierung) und Zielpersonen (Individualisierung) unterschieden. Während also Zielgruppen streng anhand eines oder mehrerer Merkmale definiert sind (beispielsweise weiblich), basieren Zielpersonen auf individuellen Einschätzungen in Form einer numerischen Bewertung, dem sogenannten Score. Ein Beispiel für die Segmentierung ist die Einschränkung des Versands eines Sport-Katalogs auf sportaffine Kunden, die bereits Sportartikel gekauft

Tab. 1: Gewinnrechnung für eine klassische Mailingkampagne

Fixkosten	50.000 EUR	=	50.000 EUR
Kosten für M	$100.000 \cdot 1,50$ EUR	=	150.000 EUR
Kosten für NF	$98.500 \cdot 1,50$ EUR	=	147.750 EUR
Kosten für IP	$2.500 \cdot 5$ EUR	=	12.500 EUR
Gesamtkosten		=	360.250 EUR
Einnahmen	$1.000 \cdot 500$ EUR	=	500.000 EUR
Gewinn		=	139.750 EUR

Break-Even-Point: 721 Reagierer

haben. Im Fall der Individualisierung wird hingegen für jeden einzelnen Kunden geprüft, ob er für *diesen* Sport-Katalog affin ist, unabhängig davon, ob er zum Segment der Sport-affinen Kunden gehört. Für die Segmentierung werden zumeist Clusteranalysen verwendet, während für die Individualisierung meist Klassifikations- und Regressionsverfahren eingesetzt werden. Dies wollen wir im Folgenden am Beispiel der Mailingoptimierung ausführlicher diskutieren.

Erfolgsgeschichte: Mailingoptimierung. Beim Versand von Katalogen werden derzeit oft noch alle Kunden einer Adressliste angeschrieben, unabhängig davon wie hoch ihre Reagierwahrscheinlichkeit ist. Man spricht davon, dass die Kataloge unpersonalisiert versendet werden. Die Kosten für eine solche Mailingwerbemaßnahme setzen sich vereinfacht dargestellt aus den Fixkosten (primär Katalogerstellung und -druck), sowie den Kosten für den Mailingversand (M), das Nachfassen (NF) und die telefonische Bestellannahme (IP) zusammen. In Tab. 1 ist die Gewinnrechnung für eine exemplarische Mailingaktion mit 100 000 Empfängern, 1 % Reagierwahrscheinlichkeit und 500 Euro Einnahmen je Reagierer dargestellt. Der Einfachheit halber wird hierbei der Rohertrag mit dem Gewinn gleichgesetzt.

Um nun die Profitabilität des Mailings zu erhöhen, sollen durch den Einsatz von Data Mining nur noch die Kunden mit der höchsten Reagierwahrscheinlichkeit angeschrieben werden. Der Ablauf einer solchen optimierten Mailingaktion ist in Abb. 1 schematisch dargestellt. Dazu werden historische Mailings für den Katalog basierend auf dem vorhandenen Kundenstammdaten analysiert, daraus Modelle zur Erstellung der Kundenprofile für die Reagierwahrscheinlichkeit maschinell gelernt und diese dann auf die potentiell anzuschreibenden 100 000 Kunden angewendet. Im Ergebnis erhält man eine Scoring-Liste, aus der schließlich die 40 000 Kunden mit den höchsten Scores (d. h. der potentiell größten Reagierwahrscheinlichkeit) ausgewählt und angeschrieben

Tab. 2: Gewinnrechnung für eine optimierte Mailingkampagne

Fixkosten	50.000 EUR	=	50.000 EUR
Kosten für M	$40.000 \cdot 1,50$ EUR	=	60.000 EUR
Kosten für NF	$38.600 \cdot 1,50$ EUR	=	57.900 EUR
Kosten für IP	$2.100 \cdot 5$ EUR	=	10.500 EUR
Gesamtkosten		=	178.400 EUR
Einnahmen	$950 \cdot 500$ EUR	=	475.000 EUR
Gewinn		=	296.600 EUR

Break-Even-Point: bereits nach 357 Reagierern!

werden. Bei fachgerechter Umsetzung eines solchen Data-Mining-Ansatzes kann dabei mehr als eine Verdopplung der Antwortrate erreicht werden, hier seien es 950 Reagierer (ca. 2,4 %).

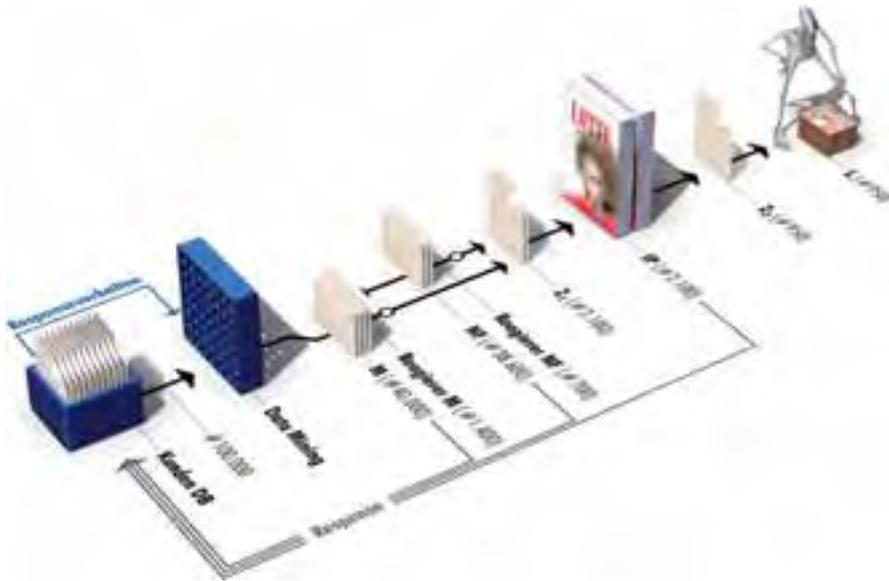
Die Gewinnrechnung in Tab. 2 zeigt, dass im Vergleich zur klassischen Mailingkampagne der Gewinn trotz weniger Einnahmen im Ergebnis von ca. 140 000 EUR auf ca. 297 000 EUR steigt. Darüber hinaus werden Kunden nicht mehr so stark mit Katalogen überflutet was ein weiteres Plus ist. Solche personalisierten Mailingkampagnen werden mittlerweile von etlichen Versandhändlern erfolgreich eingesetzt.

Die Optimierung einer Mailingkampagne basiert dabei auf der Klassifikation von Kunden mittels eines Modells, das aus vorhandenen Kundendaten gelernt wurde. Ein Beispiel für ein solches Klassifikationsverfahren ist in Abb. 2 genauer beschrieben. Mathematisch basiert es auf der approximativen Rekonstruktion und Auswertung von hochdimensionalen Funktionen über einem Zustandsraum von Kundenattributen. Dabei kann die Methode der dünnen Gitter [5, 12] zur Approximation hochdimensionaler Funktionen eingesetzt werden. Alternativen sind hier kernbasierte Ansätze mit beispielsweise radialen Basisfunktionen oder neuronalen Netzen. Darüber hinaus existieren viele weitere Techniken zur Klassifikation [8].

2.2 CROSS- UND UP-SELLING

Nun betrachten wir ein zweites Beispiel für die Angebotsoptimierung im Handel, das Cross- und Up-Selling. Jeder Verkäufer weiß, dass es leichter ist einem Bestandskunden zusätzliche Produkte zu verkaufen als neue Kunden zu gewinnen. Genau diese zentrale Thematik der Kundenwertsteigerung adressiert das so genannte Cross- und Up-Selling. Die Aufgabe besteht darin, Kunden, die durch Interesse oder Kauf Produktpräferenzen signalisiert haben, ergänzende Produkte anzubieten (Cross-Selling) bzw. zum Kauf hö-

Abb. 1: Ablauf einer optimierten Mailingkampagne



herwertiger Substitutionsprodukte anzuregen (Up-Selling). Neben der Umsatzsteigerung auf der Seite des Händlers führt gutes Cross-Selling auf der Seite des Kunden zu größerer Zufriedenheit. Indem der Kunde gezielt die ihn interessierenden Angebote erhält, spart er Zeit und kann lästiges Suchen vermeiden.

Cross-Selling beginnt bereits mit der Disposition der Produkte im Markt. Dies ist traditionell Aufgabe des Disponenten. Seit einigen Jahren werden hier aber auch mathematische Verfahren verwendet. Dabei werden vorrangig Methoden der *Clusteranalyse* eingesetzt; besonders die *Warenkorbanalyse* erfreut sich steigender Beliebtheit. Diese Verfahren arbeiten *transaktions-basiert* und analysieren z.B. die Kassendaten auf Verbundkäufe. Häufig zusammen gekaufte Produkte werden dann auch auf der Fläche zusammen plziert. Alternativ dazu werden *content-basierte* Methoden eingesetzt, die Produkte und Kategorien über deren Attribute analysieren (Farbe, Beschreibung, *Klang*, etc.) und daraus Produktcluster finden.

Neben der Disposition von Produkten hat das E-Business eine völlig neue Art des aktiven und automatisierten Cross-Sellings eingeführt: *Empfehlungsmaschinen* (*Recommendation Engines*) und *Berater* (*Avatare*) weisen den Kunden auf ergänzende Produkte und Services hin. Vorreiter war hier der Onlineshop *Amazon.com*. Kunden werden gezielt während des Einkaufs und des damit verbundenen Warenkorbprozesses auf Ansichtssei-

Abb. 2: Dünn-Gitter-Klassifikation für die Optimierung von Mailingkampagnen

Klassifikation und Regression mit Regularisierungsnetzwerken.

Das Problem der Optimierung einer Mailingkampagne stellt ein *Klassifikationsproblem* dar, das mathematisch als ein *hochdimensionales Approximationsproblem* formuliert werden kann.

Dabei beschreibt $T = \{x_i \in \mathfrak{R}^d\}_{i=1}^M$ eine Menge von M Kunden, welche bereits an einer vorhergehenden Mailingkampagne teilgenommen haben und durch d Merkmale wie Geschlecht, Alter und Beruf gekennzeichnet sind. Weiterhin sei für jeden Kunden der Wert des Zielmerkmals $y_i \in \{-1, +1\}$ bekannt, d. h. ob der Kunde auf die Kampagne reagiert hat oder nicht. Ziel ist es aus den vorliegenden Daten eine Funktion zu rekonstruieren, die den wahrscheinlichen Zusammenhang zwischen den Merkmalen und dem Kundenverhalten wiedergibt und somit Vorhersagen für neue Kunden ermöglicht. Dies ist exemplarisch in zwei Dimensionen im nebenstehenden Bild dargestellt.

Der Lösungsprozess gliedert sich in zwei Phasen: Zunächst wird in der *Trainingsphase* aus den historischen Daten T die Klassifikationsfunktion f berechnet, welche den Zusammenhang zwischen Kundenmerkmalen und Reaktionswahrscheinlichkeit beschreibt.

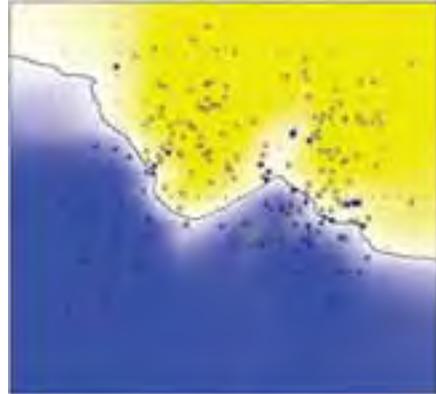
Um die Klassifikationsfunktion f zu berechnen, wird ein Minimierungsproblem

$$\min_{f \in V} R(f)$$

in einem geeigneten Funktionenraum V gelöst, wobei R ein Operator ist, der $V \rightarrow \mathfrak{R}$ abbildet. Für die anspruchvollsten derzeitigen Verfahren kann dieser mit Hilfe des sogenannten Regularisierungsnetzwerkansatzes [7] aufgeschrieben werden:

$$R(f) = \frac{1}{M} \sum_{i=1}^M C(f(x_i, y_i)) + \lambda \phi(f)$$

mit $C(x, y)$, dem Fehlerfunktional zwischen Daten und der Klassifikationsfunktion, z. B. $C(x, y) = (x - y)^2$, $\phi(f)$ dem Regularisierungsoperator, der Priorannahmen über f ausdrückt, z. B. $\phi(f) = \|\nabla f\|^2$, und dem Regularisierungsparameter λ .

Klassifikation mittels einer Dünn-Gitterfunktion

Bei der eigentlichen numerischen Berechnung wird nun eine diskrete Approximation der Funktion f bestimmt. Auf Grund der hohen Dimensionen lässt sich hier beispielsweise die *Methode der dünnen Gitter* [5, 12] verwenden. Dieser Ansatz weist zudem nur eine lineare Skalierung mit der Zahl der Daten auf, für Details siehe [6]. Die Methode der dünnen Gitter bricht den Fluch der Dimensionen konventioneller gitterbasierter Diskretisierungsansätze durch die Verwendung einer abgeschnittenen Tensorprodukt-Multiskalenbasis. Die Anzahl der benötigten Gitterpunkte wächst zwar weiterhin noch exponentiell mit der Zahl der Dimensionen, allerdings mit einer im Vergleich sehr viel kleineren Basis. In der *Applikationsphase* wird nun die Klassifikationsfunktion f auf die Daten neuer Kunden angewendet. Dabei liefert f den *Score* – je höher dieser ist, desto wahrscheinlicher gehört der Kunde der Reagiererkategorie an. Daher werden nach der Anwendung der Klassifikators auf neue Kundendaten die zugehörigen Kunden nach ihrem Scorewert absteigend sortiert und danach die n ersten Kunden für die Kampagne ausgewählt. Für die Wahl der richtigen *Ausstattungsichte* n spielen Kosten für die Kampagne und erwarteter Kundenumsatz eine Rolle.

Heutzutage werden für Mailingkampagnen oft bis zu einige Millionen Kunden analysiert um eine möglichst gute Auswahl zu treffen. Der Einsatz von Klassifikationsverfahren zur Mailingoptimierung bringt dabei bis zu 100 % höhere Erträge, wobei hier zu beachten ist, dass die Antwortraten beim herkömmlichen Mailing typischerweise unter 1 % liegen.

ten mit passenden Produkten hingewiesen („Kunden, die dieses Produkt gekauft haben, kauften auch ...“). Wenn auch die frühen Algorithmen von *Amazon.com* auf einfacher Korrelationsanalyse basierten, so haben sie doch den modernen Empfehlungsmaschinen und adaptiven Analysesystemen den Weg gewiesen. Im E-Business haben sich Empfehlungsmaschinen mittlerweile etabliert und werden in verallgemeinerter Form für vielfältigste Anwendungen wie Suchen, Interessenabgleich, persönliche Seiten und dynamische Navigationsgestaltung benutzt. Zugleich ist dies zum Thema umfangreicher universitärer Forschung geworden und es gibt derzeit eine Vielzahl von Veröffentlichungen hierzu. Die aktuellen Methoden reichen von Clustering und Textanalyse, über Bayesische Netze und neuronale Netze bis hin zu komplexen hybriden Lösungen. Wenn auch das mathematische Niveau vieler Ansätze noch unbefriedigend ist, so kann es dennoch keinen Zweifel daran geben, dass hier momentan ein spannendes Forschungsthema für die Angewandte Mathematik entsteht. Dies wollen wir im Folgenden am Beispiel der dynamischen Programmierung bei Produktvorschlägen ausführlicher diskutieren.

Erfolgsgeschichte: Produktvorschläge. Empfehlungsmaschinen nehmen heute einen wichtigen Platz in der automatisierten Kundenkommunikation ein. Eine Empfehlungsmaschine bietet basierend auf dem Klick- und Kaufverhalten der Kunden automatisch passende Produktvorschläge an. Hierbei lernt die Empfehlungsmaschine online direkt aus der Kundeninteraktion. Das klassische Beispiel sind Online-Shops. Empfehlungsmaschinen erhöhen den Umsatz um bis zu 20 % und führen zu größerer Kundenzufriedenheit. Doch ist damit ihr Potential noch lange nicht erschöpft: Moderne Empfehlungsmaschinen variieren mittlerweile Design, Sortiment und Preise abhängig von den Nutzern und erschließen völlig neue Möglichkeiten der Personalisierung.

Im stationären Handel schien der Einsatz automatischer Empfehlungssysteme bislang technisch unmöglich, obwohl das Interesse dafür durchaus vorhanden ist, da die meisten Kaufentscheidungen erst vor Ort fallen. Doch bahnt sich gerade hier ein grundlegender Wandel an. Mittlerweile stehen in ersten Shopping Malls elektronische Geräte wie der Personal Shopping Assistant (PSA) zur Verfügung, ein persönlicher Einkaufsberater, der auf dem Einkaufswagen platziert wird. Kunden können nun Produkte aus dem Regal an den Scanner des PSA halten und bekommen neben detaillierten Informationen zum Produkt passende Produktempfehlungen (Abb. 4). Diese Systeme ermöglichen erstmals eine vollständig automatisierte Interaktion mit dem Kunden im Geschäft, etwa in Form von Echtzeit-Couponsing auf dem Kassenbon abhängig vom Einkauf oder in Form von Preisänderungen in Echtzeit mittels elektronischer Preisschilder. Damit ergibt sich ein interessantes und weites neues Feld für automatische Analysesysteme sowie Empfehlungsmaschinen zur Bereitstellung von Produktvorschlägen und Preisoptimierung.

Eine moderne mathematische Formulierung solcher adaptiver Produktempfehlungen erfolgt mittels Ansätzen aus dem Gebiet der stochastischen optimalen Steuerung und wird in Abb. 3 genauer diskutiert. Dabei wird eine Reinforcement Learning Formu-

Abb. 3: Kundeninteraktion mit Hilfe von stochastischer optimaler Kontrolle

Produktempfehlung als Reinforcement Learning Problem

Das Problem der adaptiven Produktempfehlungen kann als *Reinforcement Learning* Problem dargestellt werden, das mathematisch auf der Theorie der Optimalen Steuerung basiert. Dazu gibt es neben theoretischen Ansätzen bereits auch eine erfolgreiche kommerziell Umsetzung.

Im Reinforcement Learning wird eine Menge von Zuständen s betrachtet. In jedem Zustand kann eine Aktion a aus einer Menge von Aktionen ausgewählt werden, die zu einem neuen Zustand s' mit der skalaren *Belohnung* r führt. Die Summe aller Belohnungen soll hierbei über eine *Episode* maximiert werden. Im einfachsten Fall einer Empfehlungsmaschine entsprechen die Zustände den angeschauten Produkten, die Aktionen den empfohlenen Produkten und die Belohnungen den Produktpreisen im Falle des Kaufs.

Sofern nun die sogenannte Markov-Eigenschaft erfüllt ist, kann zu einem *Markov Decision Prozeß* übergegangen werden, der durch die diskrete *Bellman-Gleichung* beschrieben wird:

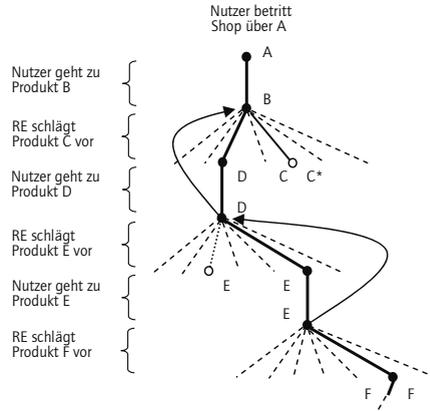
$$V^\pi(s) = \sum_a \pi(s, a) \sum_{s'} P_{ss'}^a [R_{ss'}^a + \gamma V^\pi(s')]$$

mit

- $P_{ss'}^a$ Übergangswahrscheinlichkeiten
- $R_{ss'}^a$ Übergangsbelohnungen
- γ Discountrate

Hierbei ist $V^\pi(s)$ eine *Zustandswertfunktion*, die jedem Zustand s die erwartete kumulierte Belohnung über den Rest der Episode zuordnet. Es wird nun die optimale *Policy* $\pi(s, a)^*$ gesucht, welche die stochastische Wahl der richtigen Aktion a im Zustand s beschreibt. Somit liefert $\pi(s, a)^*$ die optimalen Empfehlungen.

Interpretation der Kundeninteraktion als Wechselspiel zwischen Empfehlungsmaschine und Kunde



Zur Berechnung der optimalen Policy werden verallgemeinerte Verfahren der Optimalen Steuerung benutzt, insbesondere *Policy Iteration* und *Value Iteration* Verfahren.

Da in vielen Applikationen die Zustandswertfunktion $V^\pi(s)$ auf Grund der Vielzahl an Zuständen nicht mehr tabular erfasst werden kann, wird sie in diesen Fällen über Regressionsansätze approximiert. Hier spielt wiederum die effiziente und approximative Darstellung von hochdimensionalen Funktionen eine entscheidende Rolle. Es ist weiterhin zu unterscheiden zwischen RL-Verfahren, die ein Modell kennen und nutzen, d. h. die Übergangswahrscheinlichkeiten $P_{ss'}^a$ kennen (z. B. über Regression); hier können die klassischen Methoden der Optimalen Steuerung benutzt werden, welche die Berechnung für Dutzende von Millionen Zuständen in wenigen Stunden erlauben. Auf der anderen Seite stehen modellfreie Verfahren – insbesondere Monte Carlo und TD-Verfahren – die zum reinen Online-Lernen genutzt werden.

Abb. 4: Produktempfehlungen im Personal Shopping Assistant



lierung verwendet, die auf eine diskrete Hamilton-Jacobi-Bellmann-Gleichung führt. Wiederum treten hochdimensionale Zustandsräume auf für deren Behandlung bisher neuronale Netze, kernbasierte Ansätze oder Entscheidungsbäume verwendet werden [11]. Hier gibt es zudem wieder die Möglichkeit die Methode der dünnen Gitter einzusetzen [9].

3 AUSBLICK

Ein wichtiger Aspekt von erfolgreichen Data-Mining-Verfahren, wie sie im Handel auftreten, ist die approximative und effiziente Darstellung von hochdimensionalen Funktionen. Hier sind gerade im letzten Jahrzehnt Fortschritte erzielt worden um den so genannten Fluch der Dimension [1] zu brechen, d. h. Verfahren zu entwickeln und zu analysieren, deren Komplexität nicht exponentiell mit der Dimension des zu Grunde liegenden Raums skaliert. Neben der bereits erwähnten Methode der dünnen Gitter [5] seien hierzu Niedrig-Rang Tensorproduktansätze [3, 4] sowie diverse nichtlineare Approximationsverfahren wie etwa neuronale Netze, Kernmethoden und LASSO erwähnt. Jeder Ansatz hat dabei seine eigene spezifische Annahme an die zu approximierende Funktion und die Daten damit der Fluch der Dimension auch wirklich umgangen werden kann. Hier ist weitere Forschungsarbeit im Bereich der Numerik hochdimensionaler Probleme erforderlich um die Effizienz der Verfahren voranzubringen und ihren Einsatz

bereich, insbesondere für Probleme der Angebotsoptimierung im Handel, besser zu verstehen.

Eine grundlegende Weiterentwicklung der hier dargestellten Angebotsoptimierung im Handel besteht darin in Zukunft den kompletten *Kundenlebenszyklus* zu managen. Bisher werden Marketingaktionen wie Kampagnen, Cross- und Upselling sowie Sortiments- und Preisoptimierung zumeist isoliert betrachtet. Diese Maßnahmen zur Ertragssteigerung erweisen sich dabei oft als nicht nachhaltig. Im Idealfall soll über alle fünf Dimensionen (Kunde, Inhalt, Zeit, Werbeform, Preis) des Eingangsproblems eine Optimierung stattfinden mit dem Ziel den Kundenwert über den gesamten Kundenlebenszyklus zu maximieren. Auf mathematischer Seite kann dabei ein Beitrag bei der Quantifizierung und Optimierung des Wertes eines jeden Kunden erfolgen. So kann eine Ertragsrechnung auf Kundenbasis stattfinden, die mit Methoden der stochastischen optimalen Kontrolle beziehungsweise hier in der Form des Reinforcement Learning im Hinblick auf den Ertrag pro Kunden maximiert wird. Solche erweiterten Optimierungsprobleme erfordern wiederum die Darstellung von Funktionen (dem Kundenwert) über einem hochdimensionalen Raum (dem Zustandsraum). Der Zustandsraum gibt hier den Kunden, die Produkte sowie deren Preise und den Zeitpunkt und die Form der Werbeaktionen wieder. Das Unternehmen kann Aktionen durchführen, wie zu bestimmten Zeiten geeignete Werbung zu passenden Produkten mit attraktiven Preisen zu schalten. Der Erfolg misst sich im erzielten Kundenwert. Interessant für die Zukunft sind dabei die Entwicklung neuer adaptiver Regressionsverfahren, etwa für die Echtzeitoptimierung von Kampagnen, sowie die optimale Steuerung, etwa für die Umsatzoptimierung über dem gesamten Kundenlebenszyklus. Beide Fragestellungen müssen dabei in hochdimensionalen Approximationsräumen betrachtet und gelöst werden.

Abschließend seien kritische Bemerkungen zur kommerziellen Nutzung von personenbezogenen Daten und Wohnumfeldinformationen erlaubt. Banken, Versicherungen, und, wie dargestellt, zunehmend auch der Einzelhandel, arbeiten in immer stärkerem Maße mit Kundenprofilen und Risikogruppen. Gute Konditionen gibt es dabei nur für gute Kunden. Verstärkt wird dieser Trend durch immer umfassendere Datenbanken, durch Kundenkarten und einen schwunghaften Handel mit Verbraucherdaten. Es gibt Untersuchungen, die vor einer zunehmend pauschalen Beurteilung von Verbrauchergruppen warnen. Es besteht die Gefahr, dass viele Personengruppen gewisse Dienstleistungen oder Produkte nicht in Anspruch nehmen können, weil sie von Unternehmen, die ein Data Mining-Modell in ihren Geschäftsprozess integriert haben, auf Grund der Datenlage automatisch gesperrt werden. Kunden, die abhängig von ihrem Wohnort erhöhte Kreditzinsen zahlen müssen, Versicherte, die pauschal auf Grund von Erkrankungen keine Berufsunfähigkeitsversicherung erhalten oder Personen, die wegen verspäteter Bezahlung der Handyrechnung und resultierendem negativen SCHUFA-Eintrag keine Wohnung bekommen, sind aktuelle Beispiele. Um solchen Fehlentwicklungen zu begegnen, ist auf Dauer eine hohe Transparenz der Data-Mining-Lösungen möglichst mit Erklärungskomponenten notwendig („warum bekomme ich das empfohlen, wie kam die Emp-

fehlung konkret zu Stande?“). Auch sind existierende Datenschutzgesetze unter diesen Aspekten vom Gesetzgeber kritisch zu hinterfragen und gegebenenfalls anzupassen.

4 VISIONEN UND HANDLUNGSEMPFEHLUNGEN

Es ist in der Zukunft durchaus vorstellbar, dass sich Data-Mining-Verfahren in Form von Analysesystemen sowie Empfehlungsmaschinen zur Bereitstellung von Produktvorschlägen und Preisoptimierung für eine vollständig automatisierte Interaktion mit dem Kunden im Geschäft durchsetzen werden. Der Schwerpunkt wird hierbei auf Interaktivität und Adaptivität liegen, wobei Online-Learning eine wichtige Rolle spielen wird. Ein weiterer Trend ist sicherlich, den kompletten Kundenlebenszyklus (Customer Life Cycle) zu managen und dabei den Kundenwert für ein Unternehmen über den gesamten Kundenlebenszyklus zu maximieren (Lifetime value). Technisch wäre dies schon heute weitgehend machbar, alle Informationen und Daten des Kundenlebenszyklus sind prinzipiell in den Unternehmen vorhanden, werden jedoch aktuell oft noch nicht genutzt, da sie meist IT-technisch noch nicht direkt aufbereitet zur Verfügung stehen. Weiterhin sollte auch der oben erwähnte Echtzeitansatz nicht mehr allzu schwierig umsetzbar sein, da die Unternehmen aktuell in Richtung serviceorientierter Architektur (SOA) gehen. Insgesamt sind somit die technischen Voraussetzungen für den Einsatz von Data-Mining-Methoden für die Angebotsoptimierung im Handel gegeben.

Notwendig ist jedoch noch die mathematische Weiterentwicklung von interaktiven, adaptiven Online-Learning-Methoden. Bisher existieren hier meist nur heuristische ad hoc Ansätze. Eine moderne mathematische Formulierung solcher adaptiven Produktempfehlungen kann mittels Ansätzen aus dem Gebiet der stochastischen optimalen Steuerung über eine Reinforcement Learning Formulierung gelingen, die auf eine diskrete Hamilton-Jacobi-Bellmann-Gleichung führt. Für deren effiziente und schnelle Lösung sind jedoch neue Algorithmen und Verfahren insbesondere für den hochdimensionalen Fall erst zu entwickeln beziehungsweise bestehende Methoden substantiell zu verbessern. Dies kann nur in interdisziplinärer Zusammenarbeit zwischen Mathematik, Informatik und Anwendern geschehen. Dazu ist jedoch ein spezifisches Förderungsinstrument zu schaffen, das es interessierten Gruppen ermöglicht in diesem Bereich aktiver zu werden. Mit dem Schwerpunktprogramm 1324 „Mathematische Methoden zur Extraktion quantifizierbarer Informationen aus komplexen Systemen“ der Deutschen Forschungsgemeinschaft sowie dem Programm „Mathematik für Innovationen in Industrie und Dienstleistungen“ des BMBF wurden aktuell sowohl auf praktischer als auch theoretischer Seite wichtige erste Schritte gemacht, die es in der Zukunft zu verstetigen und auszubauen gilt. Hierzu ist aber neben der institutionellen Förderung auch der finanzielle Einsatz der Industrie und des Handels gefragt, der leider noch sehr zögerlich und wenig risikofreudig ist.

5 LITERATUR

- [1] R. Bellman, *Dynamic Programming*, Princeton Univ. Press, 1957.
- [2] *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*, Wiley and Sons, 2004.
- [3] G. Beylkin und M. J. Mohlenkamp, *Algorithms for Numerical Analysis in High Dimensions*, *SIAM J. Sci. Comput.*, 26:2133–2159, 2005.
- [4] S. Börm, L. Grasedyck und W. Hackbusch, *Hierarchical Matrices*, Lecture Note 21, Max Planck Institute for Mathematics in the Sciences, Leipzig, 2003.
- [5] H.J. Bungartz und M. Griebel, *Sparse Grids*, *Acta Numerica*, 13:147–269, 2004.
- [6] J. Garcke, M. Griebel und M. Thess, *Data Mining with Sparse Grids*, *Computing*, 67:225–253, 2001.
- [7] F. Girosi, M. Jones und T. Poggio, *Regularization Theory and Neural Network Architectures*, *Neural Computation*, 7:219–265, 1995.
- [8] T. Hastie, R. Tibshirani und J. Friedman, *The Elements of Statistical Learning*, Springer, 2001.
- [9] R. Munos, *A Study of Reinforcement Learning in the Continuous Case by the Means of Viscosity Solutions*, *Machine Learning*, 40(3):265–299, 2000.
- [10] B. Schoenrade, *Data Mining im Marketing: Status quo und Entwicklungspotential*, Vdm Verlag, Dr. Müller, 2007.
- [11] R.S. Sutton and A.G. Barto, *Reinforcement Learning: An Introduction*, MIT Press, 1998.
- [12] C. Zenger, *Sparse Grids*, In *Parallel Algorithms for Partial Differential Equations*, *Proceedings of the Sixth GAMM-Seminar*, Kiel, 1990, volume 31, *Notes on Num. Fluid Mech.*, Vieweg-Verlag, 241–251, 1991.

NETZE

> PLANUNGSPROBLEME IM ÖFFENTLICHEN VERKEHR

RALF BORNDÖRFER, MARTIN GRÖTSCHEL UND ULRICH JAEGER

1 EXECUTIVE SUMMARY

Millionen von Menschen werden allein in Deutschland täglich von Bussen, Bahnen und Flugzeugen transportiert. Der *öffentliche Personenverkehr* (ÖV) ist von großer Bedeutung für die Lebensqualität einzelner, aber auch für die Leistungsfähigkeit ganzer Regionen. Qualität und Effizienz von ÖV-Systemen hängen ab von politischen *Rahmenbedingungen* (staatlich geplant, wettbewerblich organisiert) und der Eignung der *Infrastruktur* (Schienensysteme, Flughafenstandorte), vom vorhandenen *Verkehrsangebot* (Fahr- und Flugplan), von der Verwendung angemessener *Technologien* (Informations-, Kontroll- und Buchungssysteme) und dem bestmöglichen Einsatz der *Betriebsmittel* (Energie, Fahrzeuge und Personal). Die hierbei auftretenden *Entscheidungs-, Planungs- und Optimierungsprobleme* sind z. T. gigantisch und „schreien“ aufgrund ihrer hohen Komplexität nach Unterstützung durch Mathematik.

Dieser Artikel skizziert den Stand und die Bedeutung des Einsatzes von *Mathematik bei der Planung und Durchführung von öffentlichem Personenverkehr*, beschreibt die bestehenden Herausforderungen und regt zukunftsweisende Maßnahmen an.

Der gegenwärtige Beitrag der Mathematik zum ÖV ist – je nach Verkehrsträger – unterschiedlich tiefgehend. Die Unterstützung des *Luftverkehrs* durch Mathematik ist schon weit voran geschritten. Im *Busverkehr* sind in den letzten Jahren signifikante Fortschritte erkennbar, während im *Bahnverkehr* noch deutliche Verbesserungsmöglichkeiten zu verzeichnen sind. In allen Bereichen sind die vorhandenen *Potentiale* bei weitem noch nicht ausgeschöpft.

Für einige ÖV-Fragestellungen, z. B. Fahrzeug- und Personaleinsatzplanung im Bus- und Flugverkehr, sind ausgezeichnete mathematische Werkzeuge verfügbar und an vielen Stellen im Einsatz. In anderen Bereichen ist die Leistungsfähigkeit der gegenwärtigen Algorithmen noch nicht ausreichend, um die im ÖV auftretenden Problemgrößen zu beherrschen; dies gilt z. B. für die Umlaufplanung im Bahnbereich. Verschiedene Themen sind aus mathematischer Sicht weitgehend unberührt; es fehlen (außer im Flugverkehr) u. a. praxistaugliche Modelle zur Angebots- und Preisplanung. ÖV-Infrastrukturplanung erfolgt nahezu mathematikfrei, obwohl in diesem Bereich gewaltige Investitionssummen bewegt werden. Hier tun sich Fragestellungen auf, die von Ingenieuren, Ökono-

men, Politikern und Mathematikern nur in gemeinsamer Arbeit angegangen werden können.

Die Autoren schlagen unter anderem vor, zwei konkrete Themen aufzugreifen, die kurzfristig in Angriff genommen werden können, von grundlegender Bedeutung auch über den Verkehr hinaus sind, zu einer deutlichen Verbesserung der Zusammenarbeit aller am Verkehr beteiligten Partner führen sollten und, falls erfolgreich, von echtem Nutzen für Betreiber und Kunden sind:

- Diskrete Optimalsteuerung: Echtzeit-Umplanung von Verkehrssystemen bei Störungen
- Modellintegration: Angebotsplanung im Bus- und Bahnbereich.

Diese Themen könnten von BMBF, BMWi oder DFG gefördert und im Rahmen von interdisziplinären Forschungsprojekten bearbeitet werden.

2 ERFOLGSGESCHICHTEN

Was bringt der Einsatz von Mathematik im öffentlichen Personenverkehr? Drei Beispiele beleuchten den Nutzen der Mathematik für den Kunden, den Planer und den Eigentümer.

Fahrplanauskunft. Das Wälzen von dicken Kursbüchern und Flugplänen zur Bestimmung einer besten Verbindung in einem Bus-, Bahn- oder Flugnetz ist Vergangenheit. Heute bieten die Nahverkehrsbetriebe, Bahnen und Fluggesellschaften *elektronische Fahrplanauskunftssysteme* an, die diese Verbindungsinformationen per Internet oder Mobiltelefon jederzeit bequem, schnell, aktuell und kostenlos liefern. Damit dieser Service funktioniert, benötigt man zunächst einmal korrekte und vollständige Daten. Die „Intelligenz“ zur Nutzung dieser Daten liefert die Mathematik: gute Methoden zur Be-

Abb. 1: Die Fahrinfo der BVG empfiehlt eine Route.

Details	Datum	Abfahrt	Ankunft	Dauer	Umst.	Verkehrsmittel
		früher				
	30.07.08	15:24	16:25	1:01	2	   
	30.07.08	15:31	16:35	1:04	4	    
	30.07.08	15:44	16:45	1:01	2	   

rechnung von *kürzesten Wegen in Netzwerken*.¹ Geeignete Algorithmen hierfür sind seit den fünfziger Jahren bekannt. Ihr kundenfreundlicher Einsatz wurde durch die rasante Entwicklung der Informationstechnologie in den letzten Jahren Realität.

Beispiele für elektronische Fahrplanauskunftssysteme sind die Fahrinfo der Berliner Verkehrsbetriebe (<http://www.fahrinfo-berlin.de>), siehe Abb. 1, das von der Deutschen Bahn verwendete Hafas (<http://reiseauskunft.bahn.de>) und die Flugplanauskunft der Lufthansa (<http://www.lufthansa.de>). Die Basismethode zur Berechnung von kürzesten Wegen ist der *Algorithmus von Dijkstra* [17]. Dieser hat im Laufe der Jahre viele Verfeinerungen und Verbesserungen erfahren, um große Netzwerke und komplexe Nebenbedingungen bearbeiten zu können, siehe hierzu [12].

Abb. 2: Sonderangebote für Flüge Anfang und Mitte 2008 (Quellen: <http://www.lufthansa.com/online/portal/lh/de/specials> am 3. 1. 2008; <http://www.airberlin.com>, 3. 6. 2008)

Lufthansa betterFly	
Weggen Sie mit Lufthansa günstig in die schönsten Städte Europas	
+ ab Berlin	ab 99 €*
+ ab Düsseldorf	ab 99 €*
+ ab Frankfurt	ab 99 €*
+ ab Hamburg	ab 99 €*
+ ab München	ab 99 €*
+ ab Stuttgart	ab 99 €*
+ Alle Ziele	ab 99 €*
* inkl. 10 € Lufthansa Ticket Service Charge	



Revenue Management. Lufthansa bietet zur Jahresmitte 2008 Flüge zu verschiedenen Zielen innerhalb Europas für 99 Euro an, siehe Abb. 2, Air Berlin Flüge nach Paris ab 29 Euro und manchmal findet man sogar Flugtickets für unter 10 Euro. Wie kommen solche Preise zustande? Sie sind das Ergebnis ausgefeilter Ticketverkaufsstrategien, die man als *Revenue* oder *Yield Management* bezeichnet. Die Idee ist die folgende. Hat eine Fluggesellschaft erst einmal einen Flug in ihren Flugplan aufgenommen, stehen die Kosten im wesentlichen fest. Das Ziel ist dann die Maximierung der Einnahmen. Ursprünglich gab es dazu zwei Strategien: die klassischen Carrier verlangten hohe Preise (und hatten zu gewissen Zeiten leere Flugzeuge), die Billigflieger füllten ihre Flugzeuge ausschließlich mit preiswerten Tickets. Heute verfolgen alle Fluggesellschaften (je nach Firma unterschiedliche) Mischungen dieser Strategien, wobei der Clou darin besteht, die Sitzplatzkontingente und Preise kontinuierlich der bereits realisierten und noch erwarteten Nachfrage anzupassen. Diese Anpassungen werden von vielen Fluggesellschaften mit Hilfe von mathematischen Verfahren der *stochastischen Optimierung* getroffen. Auf

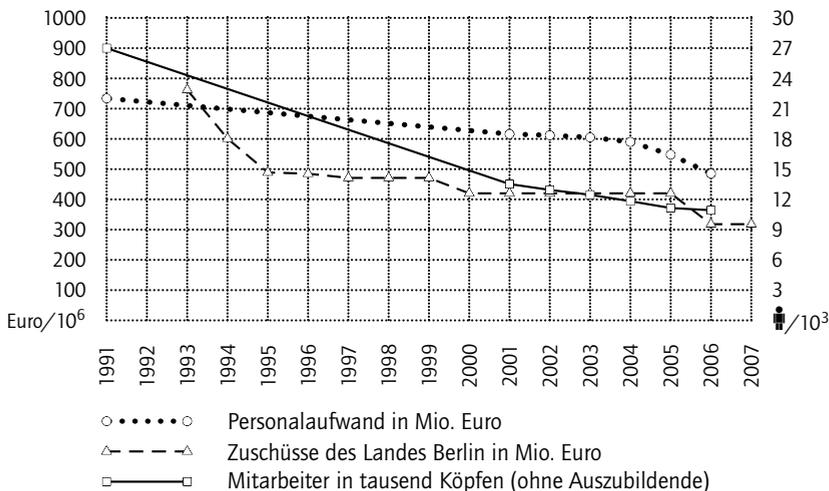
¹ Kunden bemängeln gelegentlich, dass sie bessere oder billigere Routen finden können als die Auskunftssysteme. Das liegt jedoch nicht an der Mathematik, sondern in der Regel an „Parametereinstellungen“ wie Mindestumsteigezeiten.

Basis solcher Prognosen kann es sinnvoll sein, zu bestimmten Zeiten Restkapazitäten zu sehr niedrigen Preisen zu verkaufen, um wenigstens noch geringfügige Einnahmen zu erzielen, anstatt leere Plätze durch die Gegend zu fliegen.

Die oben beschriebene und heute allgemein praktizierte Form des Revenue Managements wurde um 1980 entwickelt. Berühmt ist in diesem Zusammenhang der Konkurrenzkampf von American Airlines mit dem Billigflieger PeopleExpress, den AA mit Hilfe von „Super Saver“ und „Ultimate Super Saver“-Tickets, die mit Yield-Management-Methoden verkauft wurden, für sich entscheiden konnte. Anlässlich der Verleihung des Edelman-Preises der INFORMS im Jahr 1991 hat AA nachgewiesen, durch Revenue Management von 1988 bis 1991 zusätzliche Einnahmen in Höhe von 1,4 Milliarden USD erzielt zu haben [32]. Nach weiteren Verbesserungen wurde der Betrag später sogar auf 1 Milliarde USD pro Jahr beziffert [9]. Das bekannteste Revenue-Management-Verfahren ist die *EMSR-Regel* (Expected Marginal Seat Revenue), die besagt, dass man für einen Flug solange Tickets einer Buchungsklasse verkauft, wie der erwartete Gewinn positiv ist [26]. Ausgehend von dieser Grundform sind im Laufe der Zeit eine Vielzahl von Verfahren zur Buchungskontrolle entwickelt worden, die von der Betrachtung einzelner Flüge („leg control“) über die Berücksichtigung einfacher Netzeffekte („segment control“) bis zur Behandlung ganzer Reiseketten gehen („origin-destination control“); siehe [34] für einen aktuellen Überblick über den Stand der Forschung in diesem Bereich der stochastischen Optimierung.

Fahrzeug- und Personaleinsatzplanung. Abb. 3 zeigt die Entwicklung der Berliner Verkehrsbetriebe (BVG) in den letzten Jahren anhand dreier Kennzahlen: von 1991 bis 2006 sank die Zahl der Mitarbeiter von 27 002 auf 10 982, die Personalkosten sanken im gleichen Zeitraum von 734 Mio. Euro auf 485 Mio. Euro, die Zuschüsse der öffentlichen

Abb. 3: Die Berliner Verkehrsbetriebe in Zahlen (Quelle: [33])



Hand sanken von 762 Mio. Euro im Jahr 1993 auf 318 Mio. Euro im Jahr 2006 [33]. Diese bemerkenswerten Reduktionen wurden bei gleichbleibendem Verkehrsangebot erreicht. Sie sind natürlich nicht nur, aber auch auf mathematische Optimierung zurückzuführen. In den Worten von Andreas Sturmowski, dem Vorstandsvorsitzenden der BVG [33]: „Der Einsatz von IT-gestützten Planungssystemen ermöglicht signifikante Verbesserungen in den Planungsbereichen der BVG! Der Ressourceneinsatz wird optimiert. Produktivitätssteigerungen [werden] beim Einsatz von Personal und Fahrzeugen [erreicht durch die] Minimierung von Betriebsfahrten, [. . .], bessere Nutzung von Betriebshofkapazitäten [. . . und] Verringerung des Personalbedarfs durch optimierte Dienstpläne [. . .].“ Möglich gemacht wird dies durch Methoden der *kombinatorischen und gemischt-ganzzahligen Optimierung*, die die in diesem Bereich typischerweise auftretenden extrem großen Probleme bewältigen können. Diese „Rechenpower“ führt zu einer erheblichen Beschleunigung und Qualitätssteigerung im Planungsprozess. Als Kunde bemerkt man von diesen Fortschritten direkt nichts; indirekt ist jedoch die Wirkung auf Ticketpreise und staatliche Subventionen erheblich.

Mathematische Methoden zur *Umlauf- und Dienstplanoptimierung* werden heute als Module in den marktführenden Planungssystemen angeboten. Dazu gehören im Bereich ÖPNV die Systeme MICROBUS 2 der Berliner IVU Traffic Technologies AG (<http://www.ivu.de>), HASTUS der kanadischen Firma Giro Inc. (<http://www.giro.ca>) und Turni der italienischen Firma Double-Click (<http://www.turni.it/page001.htm>), im Luftverkehr die Systeme NetLine der Lufthansa Systems GmbH (<http://www.lhsystems.com>) und Carmen der amerikanischen Jeppesen-Gruppe (<http://www.carmen.se>), im Bahnverkehr das System railRMS von Jeppesen oder die Optimierungsmodule der amerikanischen Firma Innovative Scheduling, Inc. (<http://www.innovativescheduling.com>). Der aktuelle Stand der mathematischen Forschung ist in den Proceedingsreihen zu den jeweils alle drei Jahre stattfindenden internationalen CASPT- und nationalen Heureka-Konferenzen dargestellt [10, 11, 14, 18, 35, 36].

3 ÖV-PLANUNGSPROBLEME: ÜBERSICHT UND STATUS QUO

Der öffentliche Personenverkehr (ÖV) ist von hoher gesellschaftlicher Relevanz. Tab. 1 zeigt exemplarisch anhand einiger Daten der Berliner Verkehrsbetriebe (BVG), der Deutschen Bahn (DB) und der Lufthansa (LH) für das Geschäftsjahr 2006, dass der ÖV auch ein wichtiger Wirtschaftsfaktor ist.

Tab. 1: Öffentlicher Verkehr 2006 in Deutschland; Quellen: [4], [15], [16]. Die jeweiligen Zahlen sind interpretationsbedürftig, da sie z. T. nach unterschiedlichen Methoden ermittelt wurden. Die Mitarbeiterzahl der BVG wird z. B. angegeben als „Mitarbeiterzahl im Jahresdurchschnitt“ inkl. Auszubildende. Die Zahlen der DB und der LH enthalten unter anderem auch Güterverkehre.

Firma	Pkm/Jahr	Mitarbeiter	Umsatz/Euro	Gewinn/Euro
BVG	4,074 Mrd.	12.685	0,636 Mrd.	23 Mio.
DB	74,788 Mrd.	229.200	30,053 Mrd.	1.680 Mio.
LH	110,330 Mrd.	93.541	19,849 Mrd.	845 Mio.

Durch die anhaltende Verteuerung des Individualverkehrs wird die Bedeutung des ÖV weltweit weiter zunehmen. In Deutschland kostet eine einzige Tankfüllung im Sommer 2008 mit ca. 75 Euro ungefähr so viel wie eine Monatskarte der BVG für die Tarifzone AB, die das gesamte Stadtgebiet umfasst.

Öffentlicher Personenverkehr gliedert sich in die folgenden drei Bereiche:

- öffentlicher Personennahverkehr mit Bus, Tram, U-, S-Bahn (ÖPNV),
- Schienenpersonennah- und -fernverkehr (SPNV, SPFV),
- ziviler Passagierluftverkehr (ZPLV).

Wir werden diese im Folgenden verkürzt als *Bus-, Bahn- und Luftverkehr* bezeichnen. Die dort auftretenden Planungsprobleme kann man aus betriebswirtschaftlicher Sicht wie folgt klassifizieren:

- Betriebsplanung (d. h. Fahrzeug- und Personaleinsatz)
- Betriebssteuerung (z. B. Verspätungs- und Störungsmanagement)
- Angebotsplanung (z. B. Infrastruktur, Fahr-/Flugplan)
- Regulierung von Wettbewerbsbedingungen (z. B. Ausschreibung).

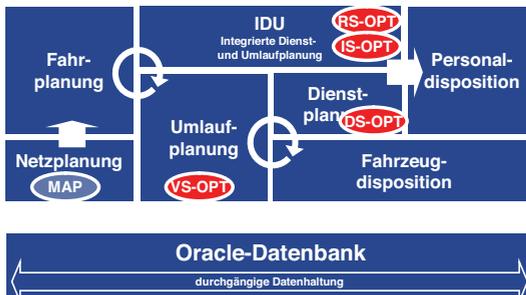
In all diesen Anwendungsfeldern werden bereits mathematische Methoden eingesetzt. Der Stand der Technik und die Durchdringung der Praxis sind jedoch unterschiedlich. In der *Betriebsplanung* haben sich mathematische Optimierungsmethoden als Industriestandard etabliert, bei der *Betriebsteuerung* und *Angebotsplanung* befindet man sich (mit Ausnahmen im Luftverkehr) im Forschungsstadium, während die Mathematik der *ÖV-Regulierung* noch in den Kinderschuhen steckt. Unterschiede gibt es auch bei den Verkehrsträgern. Die Planungsprobleme im Bus-, Bahn- und Flugverkehr ähneln sich zwar, sind aber doch nicht gleich. Technische und organisatorische Gegebenheiten sowie das Marktumfeld beeinflussen die Gliederung der Planungsprozesse und die Ausgestaltung der einzelnen Aufgaben, was wiederum Rückwirkungen auf den mathematischen Zugang hat. Aktuell ist der Einsatz von Mathematik im Luftverkehr am stärksten und bei der Bahn am schwächsten ausgeprägt.

Wir geben im Folgenden eine Übersicht über den Stand der wichtigsten Entwicklungen.

3.1 BETRIEBSPLANUNG

Die Betriebsplanung bestimmt – bei vorher festgelegtem Angebot – den Einsatz von Fahrzeugen und Personal. Der Betrieb eines Busses (allgemeine Verwaltung eingerechnet) kostet in Deutschland rund 250 000 Euro im Jahr. Von diesen Kosten entfallen in Deutschland etwa 70 % auf das Personal und knapp 25 % auf das Fahrzeug. Im Bahn- und Luftverkehr ist der Anteil der Fahrzeugkosten und hier insbesondere der Treibstoffkosten am Gesamtaufwand höher.

Abb. 4: Operative Planung im ÖPNV mit MICROBUS 2 (Quelle [28])



Die Probleme der Planung von Fahrzeugen und Personal lassen sich in ihren Grundversionen als Standardprobleme der kombinatorischen Optimierung auffassen. Man kann sie als *Mehrgüterflussprobleme* oder *Pfadüberdeckungsprobleme* formulieren. Diese Modelle sind gut untersucht und können heute in Dimensionen gelöst werden, die eine Behandlung realistischer und relevanter Szenarien mit einem hohen Detaillierungsgrad erlauben. Technische Nebenbedingungen, gesetzliche Vorgaben und Betriebsvereinbarungen erhöhen jedoch die Komplexität der Fragestellungen so stark, dass die Entwicklung von Spezialmethoden erforderlich ist. Diese weisen einen hohen Reifegrad auf und werden heute (wie bereits erwähnt) von Softwarefirmen weltweit angeboten.

Die genaue Organisation des Planungsprozesses und die Zerlegung in Einzelprobleme hängt von der Sichtweise auf den Betrieb ab und unterscheidet sich nach Ländern, Verkehrsträgern und Softwareanbietern. Abb. 4 zeigt die Sicht der IVU Traffic Technologies AG auf den ÖPNV. In dem Planungssystem MICROBUS 2 werden auf Basis der Netz- und Fahrplanung für einen Betriebstag Fahrzeugumläufe und Fahrdienste geplant, die anschließend zu Dienstreihenfolgen verkettet werden. Diese Planungsschritte werden durch mathematische Optimierungstools unterstützt, die im Bild rot gekennzeichnet sind. Wir geben im Folgenden einen zusammenfassenden Überblick und erläutern die Probleme aus Anwendungs- und aus mathematischer Sicht.

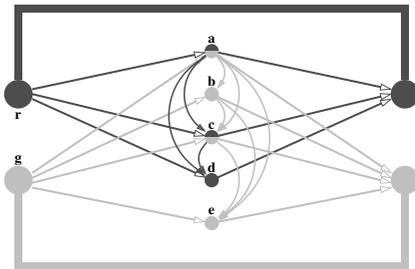
Fahrzeugeinsatzplanung. Mathematische Verfahren wurden in der Verkehrsoptimierung erstmals erfolgreich im Bereich der *Busumlaufplanung* eingesetzt. In diesem Planungsschritt werden für die Fahrzeuge Abfolgen von Fahrten und Leerfahrten, die sog. *Umläufe*, konstruiert, so dass jeder „Fahrgastfahrt“ (eine im Fahrplan ausgewiesene Fahrt) ein geeignetes Fahrzeug zugewiesen wird. Man plant hier im Wesentlichen für einen Betriebstag. Ein Bus verlässt morgens seinen Betriebshof und kehrt abends wieder zu-

rück. Das Ziel ist, alle Fahrgastfahrten abzudecken und die dabei entstehenden Kosten (durch Leerfahr- und Wartezeiten) zu minimieren. Diese Aufgabe kann mathematisch als *Mehrgüterflussproblem* formuliert werden, siehe Abb. 5. Im Fall der BVG führt das zu einer Optimierungsaufgabe mit rund 100 Millionen Variablen, die man mit modernen Algorithmen in weniger als einer Stunde auf einem handelsüblichen PC flottenoptimal lösen kann [27]. Die BVG hat damit im praktischen Einsatz nach eigenen Angaben im Jahr 2003 allein im Betriebshof Spandau 38 Busse, das ist etwa jeder fünfte, und 377 Betriebsstunden an unproduktiver Zeit eingespart ([30]).

Im Luftverkehr wird die Flugzeugumlaufplanung gewöhnlich in die Teilschritte *Fleet Assignment* und *Tail Assignment* unterteilt. Beim Fleet Assignment wird entschieden, mit welchem Flugzeugtyp ein bestimmter Flug eines Flugplans durchgeführt wird. Für die einzelnen Flugzeugflotten werden anschließend beim Tail Assignment konkrete *Rotationen* einzelner Flugzeuge inklusive notwendiger Wartungsaufenthalte geplant. Es gibt zwei Gründe für diese Aufteilung. Einmal unterscheiden sich die Kosten für den

Abb. 5: Umlaufplanung im ÖPNV (siehe [27])

Umlaufplanungsgraph:



Das Problem der Fahrzeugumlaufplanung im ÖPNV kann mathematisch als ein *Mehrgüterflussproblem* modelliert werden. Basis des Modells ist ein Planungsgraph, in dem die Fahrgastfahrten durch alle möglichen Leerfahrten verbunden werden. Die Fahrzeuge der verschiedenen Typen „fließen“ durch diesen Graphen, so dass jede Fahrgastfahrt von einem Umlauf mit passendem Typ überdeckt wird. Formal kann dieses Modell wie folgt beschrieben werden. Wir bezeichnen mit \mathcal{D} eine Menge von Fahrzeugtypen, mit \mathcal{T} eine Menge von Fahrgastfahrten, mit A^d eine Menge von Bögen für die möglichen Leerfahrten für jeden

Fahrzeugtyp d und mit κ^d die verfügbare Anzahl an Fahrzeugen von Typ d . Wir führen 0/1 Variablen x_{ij}^d für jeden Bogen (i, j) und jeden Fahrzeugtyp d ein. x_{ij}^d besagt, ob die Leerfahrt (i, j) von einem Fahrzeug des Typs d durchgeführt wird (in diesem Fall ist $x_{ij}^d = 1$) oder nicht ($x_{ij}^d = 0$).

Die zugehörige Formulierung als ganzzahliges lineares Programm ist:

$$\begin{aligned} \min \quad & \sum_{d, (i, j)} c_{ij}^d x_{ij}^d \\ & \sum_{d, (t, j)} x_{ij}^d = 1, \quad \forall t \in \mathcal{T}, \\ \sum_{d, (t, j)} x_{ij}^d - \sum_{d, (t, t)} x_{it}^d = 0, \quad & \forall t \in \mathcal{T}, \quad d \in \mathcal{D}, \\ \sum_{(d, j)} x_{dj}^d \leq \kappa^d, \quad & \forall d \in \mathcal{D}, \\ x \geq 0 \text{ und ganzzahlig.} \end{aligned}$$

Bei der Berliner BVG unterscheidet man rund 50 Fahrzeugtypen, rund 28.000 Fahrgastfahrten und rund 100 Millionen Leerfahrten. Dies übersetzt sich in dem obigen Modell in ein Mehrgüterflussproblem mit rund 100 Millionen Variablen und mehreren Hunderttausend Gleichungen und Ungleichungen.

Einsatz von größeren Flugzeugen sehr viel stärker als im Busverkehr, so dass Fleet Assignment enger mit der Angebotsplanung verzahnt ist als im Busverkehr. Zweitens ist der Planungshorizont im Wochen- bis Monatsbereich. Die mathematische Formulierung der beiden Aufgaben unterscheidet sich. Fleet Assignment kann wie bei Bussen als Mehrgüterflussproblem modelliert werden, während das Tail Assignment auf Pfadüberdeckungsprobleme führt und mit *Spaltengenerierung* gelöst wird, siehe Abb. 6. Man kann heute optimale Umlaufpläne für einen kompletten Nahverkehrsbetrieb oder eine ganze Fluggesellschaft berechnen.

Im Bahnbereich treten zusätzliche technische Nebenbedingungen auf, die unter anderem auch von den Gegebenheiten auf Bahnhöfen und den verwendeten Zugtypen abhängen können und die die mathematischen Modelle erheblich komplizieren. Hier ist die mathematische Methodik noch nicht weit genug, um ganze Bahnbetriebe behandeln zu können. Derzeit basiert Umlaufplanung bei der Bahn noch in großem Maße auf manueller Planung und heuristischen Methoden.

Personaleinsatz. In der Regel unterscheidet man bei der Personaleinsatzplanung zwei Teilschritte: *Crew/Duty Scheduling* (im Bus- und Bahnbereich auch *Dienstplanung* genannt) und *Crew/Duty Rostering* (*Dienstreihenfolgeplanung*). Beim Scheduling wird eine Anzahl von Diensten erzeugt, die noch nicht personenbezogen sind; hierbei werden arbeitsrechtliche Regeln eingehalten und Kosten minimiert. Im anschließenden *Rostering* werden diese (anonymen) Dienste (bei Bussen *Tagesdienste*, bei Flugzeugen mehrtägige sog. *Pairings*) zu Dienstreihenfolgen über einen längeren Zeitraum verkettet und konkreten Personen zugeordnet; dabei werden unter anderem Fairness-Bedingungen und Wünsche des Personals berücksichtigt.

Die ersten großen Erfolge beim Einsatz mathematischer Methoden zur Planung des Personaleinsatzes gab es im Bereich des Luftverkehrs. Die Personalkosten großer Airlines bewegen sich im Milliardenbereich und stellen hinter dem Treibstoff den zweitgrößten Kostenblock dar ([1]). Anfang der neunziger Jahre gelang es erstmals, große Crew-Scheduling-Probleme mit tausenden von Flügen optimal zu lösen ([21]). Die Optimierungstechnik hat sich seither rasant weiterentwickelt und als Industriestandard etabliert. Alle großen Airlines verwenden heute mathematische „Crew Optimizer“ [37]. Methodisch basieren diese Verfahren auf *Set Partitioning Modellen*, mit denen die sehr komplexen Regelwerke zu Pausen-, Ruhezeiten usw. abgebildet werden können. Diese Modelle werden mit *Spaltenerzeugungsverfahren* gelöst, siehe Abb. 6.

Diese Methoden sind mit zeitlicher Verzögerung auf den Bus- und teilweise auch auf den Bahnverkehr übertragen worden. Die hier auftretenden Probleme haben aber eine eigene Kombinatorik, die unter anderem dadurch entsteht, dass es im Bus- und Bahnverkehr andere Freiheitsgrade in Bezug auf die Ablösevorgänge gibt (im Busbereich werden z. B. „fliegende Wechsel“ durchgeführt); dafür wird im Busverkehr die Einsatzplanung auf den Betriebstag bezogen, während man im Luftverkehr typischerweise für mehrere Wochen oder einen Monat plant.

Abb. 6: Dienstplanung im ÖPNV [6]

Dienstplanungsgraph im ÖPNV mit Morgen- und Nachmittagspitze:



Das Problem der Dienstplanung im ÖPNV kann mathematisch als ein *Pfadüberdeckungsproblem* aufgefasst werden. Ähnlich wie in der Umlaufplanung betrachtet man einen Planungsgraphen, in dem Arbeitseinheiten, die sog. *Dienstelemente*, durch alle möglichen Verknüpfungen miteinander verbunden werden. Die Fahrdienste entsprechen dann „Pfadern“ in diesem Planungsgraphen, so dass jedes Dienstelement von einem Dienst überdeckt wird. Der Unterschied zur Fahrzeugumlaufplanung besteht in komplexen Regeln für die Zulässigkeit von Diensten wie z. B. Pausenregeln, die von der Form des Dienstes als Ganzes abhängen und sich nicht lokal entscheiden lassen. Dies führt dazu, dass man bei der ma-

thematischen Behandlung nicht mehr mit Variablen für Verknüpfungen von Dienstelementen auskommt, sondern Variablen für ganze Dienste einführen muss, was die Modelle größer und komplizierter macht.

Diese Betrachtung führt auf ein formal sehr einfaches sog. *Set-Partitioning-Problem*. Es betrachtet für eine Menge von Dienstelementen I alle zulässigen Dienste J . Für jeden solchen Dienst j wird eine 0/1-Variable x_j eingeführt, die besagt, ob Dienst j in einem Dienstplan durchgeführt wird (in diesem Fall ist $x_j = 1$) oder nicht ($x_j = 0$). Bei großen, industriellen Problemen kann diese Formulierung nicht mehr explizit behandelt werden, da es schon bei kleinen Problemen mit 200 Fahrten oder Flügen Milliarden möglicher Dienste gibt [22]. Die Modelle können aber implizit gelöst werden, indem mit Hilfe sog. *Spaltengenerierungsverfahren* die (in einem mathematisch präzisen Sinne) jeweils interessanten Dienste dynamisch generiert werden, siehe [3, 13]

Die zugehörige Formulierung als ganzzahliges lineares Programm ist

$$\begin{aligned} \min \quad & \sum_{j \in J} c_j x_j \\ & \sum_{j \ni i} x_j = 1, \quad \forall i \in I \\ & x \geq 0 \text{ und ganzzahlig.} \end{aligned}$$

Heutzutage werden Dienstplanungsprobleme im ÖPNV und im Luftverkehr mit mehreren hundert oder tausend Personen routinemäßig gelöst [2, 6].

Insgesamt ist der Personaleinsatz der Bereich, in dem der Einsatz mathematischer Methoden am weitesten fortgeschritten ist. Gelegentlich gibt es Kritik, wie z. B. der Schlagzeile der Atlanta Constitution vom 13. 8. 1994 zu entnehmen ist: „Delta to furlough 101 more pilots in bid to cut costs“. Dies hat aber nichts mit der mathematischen Methodik zu tun. Man kann auch auf hohe Sozialverträglichkeit hin optimieren. Manager stellen jedoch eher Kostensenkungsmaßnahmen in den Vordergrund.

3.2 STEUERUNG DES BETRIEBS

Wie man aus Abb. 7 ersieht, tritt neben die Planung gleichberechtigt die Steuerung (Control) des Betriebs. In der Praxis entstehen durch schlechtes Wetter, Unfälle, technische Schäden und Streiks Störungen. Die Chicago Tribune berichtet am 27. 12. 2007 hierzu: "According to its pilots union . . . , United has run low on crews to fly its planes. That's a result of lean staffing, scheduling practices and freakishly bad weather that have caused large numbers of pilots to hit the maximum number of monthly duty hours allowed by federal regulators well before December's end." Es gibt verschiedene Möglichkeiten, dieser Problematik zu begegnen.

Stochastische und robuste Optimierung. Eine theoretisch befriedigende Möglichkeit ist der Einsatz von Modellen der stochastischen Optimierung, die vorausschauend Wahrscheinlichkeiten von Störungsereignissen berücksichtigt. Ein Hindernis beim Einsatz stochastischer Methoden ist die geringe Verfügbarkeit von Erwartungswerten und Verteilungsdaten in der Praxis. Aus diesem Grunde wird seit einigen Jahren eine vereinfachte Variante untersucht, die sog. *robuste Optimierung*, bei der nur gewisse Unsicherheitsintervalle im Sinne einer Worst-Case-Analyse Berücksichtigung finden. Sie hat den Weg in die Anwendungen noch nicht gefunden. In den im praktischen Einsatz befindlichen Planungswerkzeugen wird eine gewisse Vorsorge dadurch getroffen, dass man an einigen Stellen Puffer einbaut.

Abb. 7: Airline Planning und Control mit dem System NetLine (Quelle [8])



Echtzeit-Optimierung. Durch stochastische oder robuste Optimierung und Puffer schafft man Voraussetzungen dafür, dass Störungen durch momentane Umplanungen beseitigt werden können. Umplanung ist ein Anwendungsfeld der *Online-* bzw. *Echtzeit-Optimierung*, welche bei Störungen Anpassungen ermittelt, die so schnell wie möglich wieder auf einen stabilen Betriebszustand führen und die durch Störungen auftretenden Schäden so stark wie möglich reduziert. Methoden der Echtzeit-Optimierung sind z. B. bei der Disposition der Gelben Engel des ADAC erfolgreich im betrieblichen Einsatz [20]. Im Bus- und Bahnverkehr ist im Prinzip die notwendige IT-Infrastruktur für Echtzeit-Optimierung vorhanden. Hier wird die Störungsbeseitigung jedoch weiterhin durch erfahrene Dispatcher besorgt. Im Luftverkehr gibt es erste Erfolge beim Einsatz von mathematischen Methoden, z. B. bei der amerikanischen Fluggesellschaft Continental Airlines [38]. Die Softwarehersteller haben begonnen, sich des Themas anzunehmen.

Mangelhafte Betriebssteuerung führt nicht nur zu Kundenverärgerung, sondern auch zu enormen Kosten. So müssen Fluggesellschaften Übernachtungen für Passagiere bezahlen und Leer- und Zusatzflüge durchführen, Bus- und Bahnunternehmen erhalten bei Qualitätsmängeln nicht die vereinbarten Betriebszuschüsse.

Robuste und Echtzeit-Optimierung sind neue und aktuelle Forschungszweige, die im Verkehrssektor nur dann zum Tragen kommen können, wenn eine enge Kooperation mit der Praxis erfolgt.

3.3 ANGEBOTSPLANUNG

Bei der Angebotsplanung unterscheidet man drei wichtige Bereiche: Baumaßnahmen, Fahrzeugbeschaffung und die Linien-, Fahr- und Preisplanung.

Infrastrukturplanung. Der eigentlich bedeutendste Bereich der ÖV-Planung sind die Entscheidungen über Investitionen in Infrastruktur, wie z. B. den Bau von Flughäfen, Eisenbahnlinien und Bahn- oder Betriebshöfen. Infrastrukturkosten sind enorm. So hat die ICE-Neubaustrecke von Köln nach Frankfurt über 5 Mrd. Euro gekostet, für den Bau des neuen Berliner Flughafens BBI werden 2–5 Mrd. Euro veranschlagt. Einmal getroffen, können Infrastrukturentscheidungen nur in sehr langen Zeiträumen geändert werden. Entsprechend ist eine extrem sorgfältige Analyse aller Entscheidungsmöglichkeiten erforderlich.

Die heute verwendete Entscheidungsmethodik bei Baumaßnahmen basiert auf der Simulation einiger weniger Szenarien im Sinne von Fallstudien (begleitet von in der Regel hochaufgeladenen politischen Diskussionen). Bei Einzelmaßnahmen, die in komplexem Zusammenhang mit der gesamten Verkehrsinfrastruktur stehen, greift diese Methodik zu kurz. Im Bahnbereich, um ein Beispiel zu nennen, wird diskutiert, ob der teure ICE-Strecken den höchsten Nutzenzuwachs in Bezug auf das Gesamtnetz erbringt oder ob nicht viele andere, kleinere Maßnahmen sinnvoller wären. Baumaßnahmen im Busbereich konzentrieren sich auf die Standortplanung für Betriebshöfe, ein Thema, das

derzeit speziell aufgrund von Zusammenschlüssen eine große Rolle spielt. Hierzu sind Netzoptimierungsrechnungen erforderlich. Uns ist nicht bekannt, dass dies bei der Infrastrukturplanung im ÖV heute in nennenswertem Umfang durchgeführt wird. Aus mathematischer Sicht ähnlich gelagerte Probleme treten in der Telekommunikation auf. Hier führt z. B. der DFN-Verein eine gleichzeitige Standort- und Netzoptimierung für das deutsche Wissenschaftsnetz (derzeit X-WIN) seit vielen Jahren regelmäßig durch [5]. Die hierfür entwickelte Methodik könnte als Startpunkt für eine mathematische Infrastrukturplanung im ÖV dienen.

Fahrzeugbeschaffung. Die Beschaffung von Flugzeugen, Zügen, Bussen etc. stellt einen hohen Kostenfaktor dar. Ein A380 kostet über 200 Mio. Euro, ein ICE3 über 20 Mio. Euro, ein 25 m Gelenkbus 500 000 Euro. Die Bahn betreibt 236 ICE-Triebzüge [15], die BVG über 1.300 Omnibusse [4], die Lufthansa 411 Passagierflugzeuge verschiedener Typen [16]. Die Abschätzung des Fahrzeugbedarfs für ein Flug- bzw. Bahn- oder Busnetz wird zum Teil noch mit einfachen Faustformeln gemacht [24, 31], z. B. als

$$\left\lceil \frac{\text{Linienlänge} \times \text{Taktzeit}}{\text{Durchschnittsgeschwindigkeit}} \right\rceil + \text{Reserve},$$

zum Teil werden hierbei die Instrumente der operativen Planung eingesetzt. In der Regel erfolgen die Entscheidungen durch Kompromisse auf der politischen Ebene, bei denen mathematische Überlegungen nicht immer eine wichtige Rolle spielen.

Linien-, Fahr- und Preisplanung. Bei der Linien- und Fahrplanung geht es darum, die Verkehrslinien festzulegen, mit Takten zu versehen und die konkreten Abfahrtszeiten zu bestimmen. Diese beiden Planungsschritte gehören eigentlich zusammen, werden derzeit aber nacheinander durchgeführt, weil die mathematischen Methoden zur gleichzeitigen Behandlung großer Praxisprobleme nicht vorhanden sind. Zur Lösung beider Einzelprobleme gibt es inzwischen Forschungscode, mit denen mittelgroße Nahverkehrsbetriebe behandelt werden können, siehe z. B. [7]. Etwas weiter fortgeschritten ist die periodische Taktfahrplanung, mit deren Hilfe z. B. der Fahrplan der U-Bahn in Berlin berechnet wurde [25]. Diese Ansätze gehen von einer fest vorgegebenen Nachfrage aus.

In der Preisplanung wird untersucht, wie sich die Nachfrage nach ÖV in Abhängigkeit von den geforderten Preisen verhält. Zu diesem Zweck müssen Modelle des Nachfrageverhaltens von Nutzern entwickelt werden, die eine Vorhersage der Nachfrage bei Preisvariationen erlauben. Sehr einfache Modelle der Nachfragemodellierung basieren auf Preiselastizitäten. Das Problem ist, dass diese Elastizitäten nur punktuell (für gewisse Orte zu gewissen Zeiten) empirisch geschätzt wurden und sich durchaus voneinander unterscheiden. Konstante Preiselastizitäten eignen sich nicht für die Vorhersage von Substitutionseffekten. Modernere, auf diskreten Entscheidungsmodellen basierende Ansätze sind inzwischen entwickelt worden und können in moderatem Umfang auch gelöst werden [29]. Sie sind bisher aber noch nicht in der Praxis angewendet worden. Eine Aus-

nahme bildet der Luftverkehr. Hier wird das bereits besprochene Revenue Management mit großem Erfolg eingesetzt.

Das langfristige Ziel muss sein, Linien-, Fahr- und Preisplanung integriert zu behandeln, da alle Komponenten des Systems stark voneinander abhängen. Da man die Einzelprobleme aber mathematisch nicht vollständig beherrscht, ist man von einer Systemoptimierung noch weit entfernt.

3.4 WETTBEWERBSBEDINGUNGEN

Die Angebotsplanung, wie im letzten Abschnitt diskutiert, beruht auf der gegenwärtigen Marktordnung. Überall in der Welt wird diskutiert, wie weit der Markt des öffentlichen Verkehrs dereguliert werden soll und auf welche Weise in diesem System Wettbewerb erzeugt werden kann. Die Deregulierung des Flugverkehrs hat zu einem enormen Boom geführt. Es kam zu einer starken Verbilligung der Flugpreise und hohen Steigerungsraten beim Fluggastvolumen. Ähnliche Effekte erhofft man sich im Bahn- und Busverkehr, jedoch ist nicht klar, was die richtigen Deregulierungsrezepte sind. Die bisherigen Erfahrungen in unterschiedlichen Teilen der Welt ergeben ein gemischtes Bild. Die Deregulierung der Eisenbahn in England wurde zunächst als großer Wurf gefeiert, führte danach zu einem Desaster, nach Anpassungen der Regelungen scheint sich nun ein Erfolg abzuzeichnen. Ein anderes Beispiel ist die Umgestaltung des Busverkehrs in Santiago de Chile, der im Frühjahr 2007 vollzogen wurde. Dies führte zu einem Verkehrskollaps, der beinahe den Sturz der Regierung nach sich gezogen hätte. Die mathematische Analyse von Regulierungsmaßnahmen kann ein Mittel sein, unerwünschte Auswirkungen von Systemwechseln zu erkennen und ihnen zu begegnen.

Ein Beispiel, wie dies in der Praxis aussehen könnte, gibt die Eisenbahn-Infrastruktur-Benutzungsverordnung (EIBV), die den wettbewerblichen Zugang zum bundesdeutschen Schienennetz regeln soll. Die EIBV stellt in § 9 Abs. 5 fest: „Bei der Entscheidung ... hat der Betreiber der Schienenwege die Entgelte für die streitigen Zugtrassen gegenüberzustellen und ... bei einem Konflikt zwischen mehr als zwei Zugtrassen den Zugtrassen den Vorrang einzuräumen, bei denen in der Summe das höchste Regelentgelt zu erzielen ist.“ Dies ist gesetzlich verordnete Optimierung. Sie verpflichtet den Netzbetreiber DB Netz AG, ein *Trassenallokationsproblem* zu lösen, und die Bundesnetzagentur als Aufsichtsbehörde des Schienenverkehrs darauf zu achten, dass eine Optimallösung bestimmt wird. Ohne mathematische Optimierung ist ein im Sinne dieser Vorschrift diskriminierungsfreier Zugang überhaupt nicht möglich. Dies wird von allen Beteiligten „übersehen“, in der Regel mit der Begründung, dass Verhandlungslösungen besser geeignet sind. Eine im Grundsatz ähnliche Problematik stellt sich bei der Slotvergabe im Flugverkehr und bei der Ausschreibung von Linienbündeln im Busverkehr. Diese Themen führen aus mathematischer Sicht auf Fragen der kombinatorischen Auktionen und der algorithmischen Spieltheorie, die in der ökonomischen Diskussion äußerst aktuell sind und für die mehrere Ökonomie-Nobelpreise, zuletzt 2007, vergeben wurden.

Es gibt inzwischen erfolgreiche Anwendungen dieser Methodik bei Ausschreibungen von Transportvolumina in der Logistik [23] und es sind Firmen entstanden, die entsprechende Auktionsplattformen vermarkten (combine.net). Im öffentlichen Verkehr ist die Methodik noch nicht eingesetzt worden.

4 STÄRKEN/SCHWÄCHEN-ANALYSE, HERAUSFORDERUNGEN

In Abschnitt 3 wurden die Planungsprobleme im ÖV aus betrieblicher Sicht dargestellt. Dieser Abschnitt bietet eine Gesamtschau aus mathematischer Perspektive, erläutert Stärken und Schwächen und beschreibt einige der bestehenden Herausforderungen.

4.1 RAHMENBEDINGUNGEN FÜR DEN EINSATZ VON MATHEMATIK

Zentrale Organisationsformen. Die öffentlichen Verkehrsträger Bus, Bahn und Flugzeug zeichnen sich durch relativ zentrale Organisationsformen aus. Bis auf den Bereich der Betriebssteuerung steht für alle Planungsprobleme ausreichend Vorlaufzeit zur Verfügung, so dass auch Alternativ- und Variantenrechnungen durchgeführt werden können. Damit sind grundsätzlich exzellente Voraussetzungen für die Verwendung mathematischer Planungsverfahren gegeben.

Datenverfügbarkeit und Informationstechnik. Um Optimierungsmethoden anwenden zu können, werden detaillierte und präzise Daten benötigt. Durch den Fortschritt in der Informationstechnik werden Daten in immer größerem Umfang und immer größerer Genauigkeit verfügbar. Datenfunk, RBL, GPS, elektronisches Ticket sind ergiebige Quellen über das Nutzerverhalten und den Zustand des Betriebes. Für eine vorausschauende Planung sind aber noch weitere Daten in erheblichem Umfang erforderlich. So benötigt man z. B. für die Umlaufplanung tausende von Streckenlängen für Alternativwege, die man potentiell benutzen könnte. Diese sog. Freiheitsgrade sind nicht automatisch bzw. in ausreichender Qualität in den Systemen vorhanden, und ihre Beschaffung ist häufig ein wesentlicher Aufwand bei der Einführung von Optimierungssystemen. In einigen Bereichen ist es auch grundsätzlich schwer, die notwendigen Daten zu erheben. Bei der Preis- und Angebotsplanung im Nahverkehr wäre es eigentlich notwendig, das Nutzerverhalten bei Veränderungen von Preis und Angebot zu prognostizieren. Hier arbeitet man in der Regel mit statistischen Nachfragematrizen. Die gelegentlich benutzten Preiselastizitäten beruhen auf wenig gesicherten Daten. In diesen Bereichen muss die Datenbasis erst noch geschaffen werden, um eine zielgenaue Angebotsplanung durchführen zu können.

Der öffentliche Verkehr und die Planungsmethoden profitieren in großem Maße von den atemberaubenden Entwicklungen in der Informationstechnik. Mobilfunk, Datenfunk, etc. haben den Einsatz von quantitativen, mathematischen Planungsmethoden überhaupt erst möglich gemacht.

Komplexität. In kleinen Verkehrsbetrieben und bei konstanten Rahmenbedingungen kann manuelle Planung durchaus gute Lösungen bieten. Aber auch dort wachsen die Anforderungen an Qualität und Kosteneffizienz. Bei großen Verkehrsbetrieben mit tausenden von Bussen, zehntausenden von Mitarbeitern und zigtausenden von Verbindungen ist Planung durch „Draufschauen“, durch Auswertung von Excel-Tabellen oder durch Argumentation anhand wichtiger Einzelaspekte weder Stand der Technik noch der Forschung. Gleiches gilt für die Planung kostenintensiver Infrastrukturmaßnahmen, die die Verkehrslogistik in ganzen Regionen verändern. Es ist immer wieder frappierend zu sehen, dass beim Bau von Flughäfen die Anbindung von U- und S-Bahn „vergessen“ wurde. Dies liegt daran, dass „globale Sichten“ fehlen und die Netzwerkwirkungen einzelner Entscheidungen nicht ausreichend verstanden werden. Infrastrukturpolitische Entscheidungen müssten eigentlich in umfassende Netzwerkmodelle eingebettet werden, die Verkehrsbedarfe einer Region über lange Frist berücksichtigen und dabei insbesondere auch Intermodalitäten in Betracht ziehen. Planungen dieser Art erfolgen heute in großem Maße durch Ingenieurbüros des Bauwesens, die die technischen Einzelaspekte zwar hervorragend verstehen, aber nicht über die zur Beherrschung der komplexen Zusammenhänge notwendigen mathematischen Methoden verfügen. Daneben wird Planung durch Gesetze, Betriebsvereinbarungen und andere Regelwerke, bei denen nicht an die Auswirkungen auf die Verkomplizierung der Planung bedacht wird, erschwert, siehe Abb. 8 für ein typisches Beispiel.

ÖV-Planungsprobleme sind hochkomplex. Mathematik kann Komplexität nicht zum Verschwinden bringen, bietet aber Methoden, sie zu „beherrschen“. In der Tat kann man sogar beweisen, dass die im ÖV auftretenden Planungsprobleme in einem mathematisch präzisierbaren Sinne schwierig (\mathcal{NP} -schwer [19]) sind. Auch wenn mathematische Methoden für derartige Probleme nicht notwendig Optimallösungen liefern können, so können doch Abweichungen von der Optimalität oder ähnliche Gütegarantien berechnet werden, etwas, das keine andere Vorgehensweise erbringt. In einfachen Worten: Ohne Mathematik „geht nichts“.

Erschreckend ist, dass diese Einsicht noch nicht von allen Verkehrsbetrieben realisiert wird. Vielfach dominiert selbst bei großen Betrieben immer noch manuelle Planung. Wirklich erstaunlich ist, dass sogar einige Beratungsbüros und Softwareanbieter in diesem Marktsegment argumentieren, dass Optimierung nur spezielle Aspekte, manuelle Planung jedoch „das Große und Ganze“ berücksichtigen könne – aus unserer Sicht eine vollkommen absurde Behauptung. In Wahrheit muss das Ziel darin bestehen, die Komplexität der ÖV-Systeme dadurch zu beherrschen, dass man das Know-How der Verkehrsplaner mit der mathematischen Methodik verknüpft. Nur so kann man den großen Herausforderungen bei der integrierten Planung und dem Entwurf gesamter Verkehrssysteme gerecht werden. Die Investitionen in mathematische Planungsverfahren sind im Vergleich zu den Kosten für Infrastruktur, Fahrzeuge und Personal marginal, die Effekte, die man damit erzielen kann, allerdings erheblich.

Abb. 8: Aus der 2. Durchführungsverordnung zur Betriebsordnung für Luftfahrtgerät (2. DVLuftBO)

§ 8 Flugdienstzeiten der Besatzungsmitglieder

(1) Die uneingeschränkte Flugdienstzeit jedes Besatzungsmitgliedes zwischen zwei Ruhezeiten beträgt 10 Stunden. Innerhalb 7 aufeinanderfolgender Tage ist eine viermalige Verlängerung der Flugdienstzeit nach Satz 1 bis 4 Stunden zulässig, wobei die Summe der Verlängerungen innerhalb jeweils 7 aufeinanderfolgender Tage 8 Stunden nicht überschreiten darf. Der Zeitraum von 7 aufeinanderfolgenden Tagen beginnt jeweils um 00.00 Uhr Mittlere Greenwich Zeit (MGZ) des ersten und endet um 24.00 Uhr MGZ des siebten Tages. Bei einem Luftfahrzeugführer, der während der Flugzeit nach Satz 1 ganz oder teilweise ohne Unterstützung durch ein weiteres Flugbesatzungsmitglied als Luftfahrzeugführer tätig wird, finden die Absätze 2 und 3 keine Anwendung.

(2) Bei Flugbesatzungsmitgliedern verringert sich die nach Absatz 1 höchstzulässige Zeitverlängerung von 4 Stunden

1. um 1 Stunde, wenn der Flugdienst mehr als 2, jedoch weniger als 4 Stunden,
2. um 2 Stunden, wenn der Flugdienst 4 oder mehr Stunden zwischen 01.00 Uhr und 07.00 Uhr Ortszeit des Startflugplatzes (Winterzeit)

ausgeübt wird.

(3) Eine nach Absatz 2 verringerte Zeitverlängerung ist

1. bei mehr als 3, jedoch weniger als 6 Landungen um eine weitere Stunde,
2. bei mehr als 5 Landungen um 2 weitere Stunden

zu kürzen.

(4) Bei einer Verstärkung der vorgeschriebenen Mindestflugbesatzung und bei Vorhandensein geeigneter Schlafgelegenheiten in einem von dem Führerraum und der Kabine abgetrennten Raum kann die Aufsichtsbehörde auf schriftlichen Antrag eine zweimalige Verlängerung der Flugdienstzeit nach Absatz 1, Satz 1 bis zu 8 Stunden innerhalb 7 aufeinanderfolgender Tage zulassen. Die mit der Führung und Bedienung des Luftfahrzeugs verbrachte Zeit jedes Flugbesatzungsmitglieds darf hierbei 12 Stunden nicht überschreiten.

Standardisierung. Viele Verkehrsbetriebe haben die Tendenz, sich als einzigartig zu sehen und zu glauben, dass die bei Ihnen auftretenden Probleme so speziell sind, dass sie einer Sonderbehandlung bedürfen. Daneben ist die Vielfalt der Wünsche von Planern, Managern und Betriebsräten an Softwaresysteme nahezu unbegrenzt. Eine derartige Atomisierung der Problemstellungen ist ein wesentliches Hindernis für die Weiterentwicklung der Planungsmethodik. Nur da, wo Planungsprobleme standardisiert werden und Daten in Standardformaten verfügbar sind, kann man beginnen, sich über einen längeren Zeitraum nachhaltig mit den zugrundeliegenden mathematischen Problemen zu beschäftigen, deren Struktur zu analysieren und Spezialverfahren zu entwickeln, so dass auch bei den hier vorliegenden Größenordnungen Hoffnung auf eine Lösung be-

steht. In Bereichen, in denen Standardisierung gelungen ist, haben sich im Laufe der Zeit leistungsfähige Methoden etabliert. Die Frage der „richtigen“ Definition der Fragestellungen ist teilweise kontrovers. Der Gewinn durch Standardisierung ist aber meistens so groß, dass es sich für Unternehmen auf lange Sicht eher lohnt, sich in ihren Prozessen anzupassen, als mit individuell zugeschnittenen Speziallösungen zu arbeiten, bei denen die Entwicklung nicht im gleichen Tempo voranschreitet.

Wie üblich ist die Standardisierung im Flugverkehr am weitesten fortgeschritten. Dort ist der Planungsprozess in allgemein akzeptierte Teilprobleme gegliedert, für die Datenschnittstellen und leistungsfähige Optimierungskomponenten zur Verfügung stehen. Im Busverkehr gibt es mittlerweile ähnliche Standards im Bereich der operativen Planung. In der Angebotsplanung entsteht teilweise Standardisierung durch Marktbeherrschung einzelner Softwareanbieter wie z. B. durch das Produkt VISUM der pvt AG. Diese Quasistandards bilden prinzipiell eine mögliche Basis für die Entwicklung von Optimierungsmethoden, müssten aber ausgebaut werden, da sie primär eher für Simulations- als für Optimierungszwecke konzipiert worden sind. Bei der Bahn hingegen verfolgt fast jedes Unternehmen einen individuellen Ansatz, so dass sich Ergebnisse nur sehr schwer von einer Bahn zur anderen übertragen lassen. Es gibt auch dort erste Standardisierungsbestrebungen, z. B. durch das *OpenTrack*-Projekt der ETH Zürich.

Regulierung und Deregulierung. Es ist allgemeiner Konsens, dass öffentlicher Verkehr zur Daseinsvorsorge des Staates gehört. Strittig ist jedoch, wie weit die Daseinsvorsorge reicht und wie das Verkehrsangebot organisiert wird, d. h. wie der Staat das angestrebte Wohlfahrtsmaximum für seine Bürger erreichen soll. In der Vergangenheit war man der Meinung, dass dem Bürger am besten dadurch gedient wird, dass staatlich betriebene, monopolistische Anbieter den ÖV durchführen. Diese Konzeption ist in Bedrängnis geraten. Unter anderem verfolgt die Europäische Union auch im ÖV eine eindeutige Deregulierungspolitik. Sie geht davon aus, dass ein Wohlfahrtsmaximum eher durch Marktmechanismen als durch Monopolstrukturen erreicht werden kann. Unterschiedliche Länder gehen dabei verschiedene Wege. Es ist vielfach unklar, welche Planungsfragen in Zukunft entstehen werden und welcher Marktteilnehmer aufgrund von Regulierungsmaßnahmen was planen soll.

Beim Busverkehr etwa werden folgende Szenarien diskutiert. Soll die Region das Verkehrsangebot im Detail festlegen und dann nur noch die Durchführung der Fahrten ausschreiben, sollen die Regionen Angebote von Nahverkehrsunternehmen prüfen und dem besten den Zuschlag geben, oder sollen die Nahverkehrsunternehmen im freien Wettbewerb als Mobilitätsanbieter tätig werden? Der Luftverkehr hat durch die von den USA initiierte Deregulierung einen enormen Aufschwung erlebt. Konkurrenz und dadurch notwendige Optimierungsmaßnahmen haben zu wesentlich höherer Effizienz geführt. „Sollen Rad und Schiene getrennt und was soll privat betrieben werden?“ sind kontrovers diskutierte Themen im Bahnbereich. Die Befürworter einer Bahnprivatisierung verweisen auf die Erfolge im Luftverkehr, die Gegner argumentieren, dass ein in-

tegriertes System als Gesamtheit besser geplant und gesteuert werden kann und damit kundenfreundlicher ist.

Bei monopolistischer Herangehensweise bietet sich die Chance auf eine integrierte Gesamtoptimierung, während bei Entstehen eines Marktes jeder Wettbewerber zur Optimierung seiner Aktivitäten gezwungen ist. Ob und wie reguliert und dereguliert wird, Mathematik ist in allen Fällen nützlich.

Regulierung und Deregulierung, Auktionen und Verteilungsmechanismen haben in der ökonomischen Theorie in den letzten Jahren große Aufmerksamkeit erfahren. Von einer mathematischen Theorie der Regulierung und Deregulierung kann man aber noch lange nicht sprechen. Bisher sind nur in sehr speziellen Einzelbereichen wie bei der Auktionierung von Transportleistungen, Schürfrechten oder Kommunikationsfrequenzen bereits quantitative Umsetzungen der theoretischen Überlegungen erfolgt – nicht immer zur Zufriedenheit aller Beteiligten.

4.2 MATHEMATISCHE MODELLE UND ALGORITHMEN

Die diskrete Optimierung ist zur Formulierung von sehr vielen der in der Verkehrsplanung auftretenden Planungsprobleme grundsätzlich ausgezeichnet geeignet, weil es sich vielfach um Ja/Nein-Entscheidungen handelt, die sich als 0/1-Optimierungsprobleme modellieren lassen. Die mathematischen Modellierer haben gelernt, auch komplexe Zusammenhänge in Form von linearen Gleichungen und Ungleichungen mit Ganzzahligkeitsbedingungen auszudrücken und so Modellierungswerkzeuge geschaffen, mit denen es im Prinzip möglich ist, Probleme des öffentlichen Verkehrs in größtmöglichem Detail mathematisch zu formulieren. Diese Vorgehensweise hat mehrere Vorteile. Verkehrsplanungsprobleme können so exakt, mit allen ihren Nebenbedingungen und Voraussetzungen erfasst und beschrieben werden. Diskrete Mathematik und Optimierung bieten sich als „Sprache der Verkehrsplanung“ an. Die Fragestellungen werden dadurch betriebs- und länderübergreifend kommunizier- und analysierbar. Mathematische Modelle helfen, die Problemkreise zu strukturieren, Spezialfälle und Verallgemeinerungen zu erkennen und zu benennen. Dabei sollte man nicht verschweigen, dass es für den Laien durchaus intransparent ist, welche der Modelle theoretisch oder praktisch einfach oder schwierig zu lösen sind. Selbst erfahrenen Optimierern ist oft nicht von vornherein klar, ob und in welchen Größenordnungen ein konkretes, schwieriges Problem gelöst werden kann.

Standardmodelle und -techniken. Im Laufe der Zeit haben sich eine Reihe von Standardmodellen der kombinatorischen Optimierung herauskristallisiert, die häufig im Verkehrsbereich Anwendung finden (Partitionierungs-, Pfadüberdeckungs- und Flussprobleme); hinzu kommen Basismethoden wie der Einsatz von Verfahren der Linearen und Ganzzahligen Optimierung. An diesen Modellen und Algorithmen ist in den letzten Jahrzehnten intensiv gearbeitet worden. Heute können Größenordnungen bearbeitet werden, von denen man vor wenigen Jahren noch nicht geträumt hätte. Eine Studie von Bob Bix-

by aus dem Jahr 2004 hat gezeigt, dass Lineare Programme, die 1988 zwei Monate Rechenzeit zu ihrer Lösung benötigten, 2004 in einer Sekunde gelöst werden konnten. Die Beschleunigung der Rechner trug hierzu einen Faktor von etwa 1.600, die Verbesserung der LP-Algorithmen einen Beschleunigungsfaktor von 3.300 bei. Das schlichte Aufschreiben von mathematischen Modellen und die Verwendung von Standardtechniken reicht aber nur bei kleinen Verkehrsbetrieben und bei speziellen Fragestellungen zur Lösung der anstehenden Aufgaben aus. Häufig führen geringfügig erscheinende, aus praktischen Gegebenheiten herrührende Modifikationen zu deutlichen Erhöhungen des Schwierigkeitsgrades, welchen nur durch gezielte Forschungsanstrengungen begegnet werden kann. Die Berücksichtigung von Zeitfenstern, Reihenfolgebedingungen, Weglängenbeschränkungen usw. sind typische Beispiele hierfür.

Daten für akademische Forschung. Anwendungsnahe algorithmische Forschung zu \mathcal{NP} -schweren Problemen ist nur erfolgversprechend, wenn praxisgerechtes Datenmaterial zur Verfügung steht, an dem die algorithmischen Alternativen experimentell ausgetestet werden können. Die Verfügbarkeit von Praxisdaten für akademische Forschung ist von größtem Wert. Man muss feststellen, dass die Situation in diesem Bereich sehr unbefriedigend ist. Selbst staatliche Verkehrsbetriebe, die mit zusätzlichen staatlichen Mitteln Daten erheben, stellen diese Daten (mit ganz wenigen Ausnahmen) nicht einmal vertraulich für akademische Forschung zur Verfügung. Diese Furcht vor der Bekanntgabe von Betriebsgeheimnissen behindert den Fortschritt bei der mathematischen Lösung von Planungsproblemen im öffentlichen Verkehr erheblich.

Theorie und Algorithmen. Die im Verkehrsbereich auftretenden Planungsprobleme führen in der Regel zu kombinatorischen Optimierungsproblemen mit einer extrem großen Anzahl von Variablen und/oder Nebenbedingungen, die von kommerziellen Standardcodes der linearen oder ganzzahligen Programmierung nicht bearbeitet werden können. In den letzten 10 Jahren wurden Techniken wie Lagrangean Pricing oder Separierungsalgorithmen entwickelt, die mit derartigen Problemgrößen umgehen können, indem sie nicht alle Details des Problems von Anfang an behandeln, sondern die jeweils erforderlichen Variablen und Nebenbedingungen dynamisch generieren. Mit diesen Techniken kann man klassische Probleme der Verkehrsplanung in ausreichender Qualität lösen.

Es schließen sich eine Reihe weiterführender Fragen an, an denen heute wissenschaftlich gearbeitet wird:

- *Koppelung von diskreten Modellen (Integrierte Planung).* Bisher ist es nur in Einzelfällen geringer Größe und hoher Struktur gelungen, die Koppelung von zwei diskreten Modellen (z. B. Fahrereinsatzplanung und Fahrzeugumlaufplanung) erfolgreich zu behandeln. Das Fernziel besteht darin, die künstliche Zerlegung des ÖV-Planungsprozesses in hierarchisch geordnete Einzelprobleme zu überwinden

und mit Hilfe eines Baukastens von Modellen und Methoden, die sich miteinander kombinieren lassen, eine Gesamtplanung in einem Zug vorzunehmen. Man versucht, diese Kombination mit Methoden der nichtlinearen und nichtdifferenzierbaren Optimierung (Subgradienten- und Bündelverfahren) zu erreichen. Hier besteht noch erheblicher Forschungs- und Entwicklungsbedarf.

- *Robuste und Stochastische Planung.* Ein Aspekt, der erst durch die Erfolge der Optimierung verstärkt ins Bewusstsein gerückt ist, ist, dass sehr ausgefuchste Pläne besonders anfällig gegenüber Störungen sind. Das übergeordnete Ziel muss deshalb sein, einen Plan zu berechnen, der nicht nur auf dem Papier gut aussieht, sondern auch in der tatsächlichen Umsetzung. Hier gilt es, Puffer auf optimale Weise so einzubauen, dass kleinere Störungen ohne große Umplanungen gut abgefangen werden können. Es gibt derzeit mehrere, zumindest theoretisch konkurrierende Ansätze zur Behandlung dieser Frage, insbesondere Methoden der robusten und stochastischen Optimierung scheinen hierzu geeignet zu sein. Praktische Erfahrungen liegen jedoch kaum vor.
- *Echtzeit-Planung.* Verkehrssysteme geraten gelegentlich aus dem Ruder, weil man nicht alle potentiellen Störungsmöglichkeiten in die Planung miteinbeziehen kann. Genau diese Störungen sind es, die den Fahrgast am meisten beeinträchtigen, insbesondere wenn er keine Informationen über den weiteren Verlauf seiner Reise erhält. Ein solches Szenario ist ein Fall für die sog. Echtzeit-Optimierung, die versucht, entsprechende Sachverhalte in den Griff zu bekommen. Man muss allerdings feststellen, dass die theoretischen Werkzeuge nicht sonderlich weit entwickelt sind und insbesondere kaum brauchbare Aussagen für das praktische Verhalten von Online-Algorithmen liefern. Ein wichtiges Anliegen besteht darin, für die Echtzeit-Optimierung von Verkehrssystemen theoretisch fundierte, praxisadäquate Methoden zu entwickeln.
- *Koppelung von diskreten, stochastischen und nichtlinearen Modellen (Infrastrukturplanung).* Ein Bereich, der aus unserer Sicht nahezu mathematikfrei ist, ist die Infrastrukturplanung von Verkehrssystemen. Wenn irgendwo noch traditionelle Planung dominiert, dann ist dies hier der Fall. Entscheidungsfindung ist politischer Natur, wobei einige technische Details, Erfahrungswissen und in guten Fällen der Einsatz von Simulationssoftware Berücksichtigung finden. Die Entwicklung mathematischer Methoden zur Infrastrukturplanung ist ohne Frage schwierig, weil sie die Kombination von Modellen aus unterschiedlichen Bereichen der Mathematik erfordert, insbesondere von solchen, die mit verschiedenen mathematischen Werkzeugen arbeiten. Nutzerverhalten kann nur prognostiziert und damit stochastisch modelliert werden, Größen wie Einnahmen = Preis \times Nachfrage führen zu Nichtlinearitäten, Angebote von Monatskarten zu Treppenfunktionen und damit zu Unstetigkeiten, die Betrachtung des Tradeoffs zwischen Qualität und Preis führt zu Überlegungen der multikriteriellen Optimierung, und diese Liste lässt sich noch beliebig fortsetzen.

Betrachtet man jedoch die langfristige Wirksamkeit der Entscheidungen und die Höhe der Investitionskosten, wäre es angebracht, ernsthafte Versuche einer modellgestützten Langfristplanung zu unternehmen. Die Forschung beschäftigt sich mit einigen hier auftretenden Basisproblemen wie nichtlinearen Netzwerkflussproblemen und ganzzahligen Programmen mit speziellen Nichtlinearitäten, der grundsätzlichen Beschleunigung von Ansätzen der stochastischen Optimierung z. B. durch Szenarioreduktion, der algorithmischen Spieltheorie usw. Einige erste Erfolge sind beim Einsatz von Discrete Choice Modellen bei der Preisplanung im ÖPNV oder beim Revenue Management im Flugverkehr auf Netzbasis zu verzeichnen.

Am Ende können Entscheidungen über Infrastrukturmaßnahmen natürlich nicht mit einem einzigen mathematischen Modell getroffen werden. Alternativrechnungen, komplexe Simulationen von Szenarien müssen gleichfalls als Entscheidungskriterien herangezogen werden.

4.3 TRANSFER UND AUSBILDUNG

Kommunikation und Ausbildung. Ein Hindernis bei der Erzielung von Synergien zwischen Praktikern, Algorithmenentwicklern und Theoretikern ist ein oft bestehender Mangel an Bereitschaft, aufeinander zuzugehen. Mathematiker wollen sich ungern in die Niederungen der Datenbeschaffung begeben, Praktikern erscheinen die Überlegungen der Mathematiker zu abgehoben. Das fängt sogar schon bei der Sprache an. Wenn ein Praktiker von Optimierung spricht, meint er häufig etwas anderes als ein Mathematiker. Dieses Kommunikationsproblem verstärkt sich durch Unkenntnis über die Beiträge, die die jeweils andere Gruppe leisten könnte. Im Rahmen dieses Beitrags sticht insbesondere ein Defizit in der Ausbildung der im Verkehrsbereich tätigen Ingenieure in Bezug auf moderne Methoden der Optimierung, Graphentheorie und diskreten Mathematik hervor. Jeder Ingenieur weiß, was eine Differentialgleichung ist, viele haben aber nicht gelernt, dass man Entscheidungsprobleme als ganzzahlige Programme formulieren und lösen kann, obwohl dies der moderne Zugang zu einer guten Verkehrsplanung ist. Dies ist (übrigens weltweit) ein klarer Mangel, der in der Hochschulausbildung behoben werden muss.

Firmenlandschaft. Im Luftverkehr entstanden bei Beginn der Mathematisierung der Planung in vielen Unternehmen Operations Research (oder ähnlich genannte) Abteilungen, die sich mit den in diesem Artikel genannten Planungsproblemen beschäftigten und proprietäre Lösungen entwickelten. Im Laufe der Jahre sind die meisten dieser Abteilungen ausgesourct worden. Die besten unter ihnen haben sich zu Softwareanbietern entwickelt, die standardisierte Lösungen anbieten, welche viele Luftverkehrsgesellschaften nutzen. Im Gegensatz dazu haben sich bei Bussen und Bahnen ähnliche OR-Abteilungen nur in geringem Maße entwickelt. Softwareentwicklung wird im Bus- und Bahnbereich

von Anbietern betrieben, die aus dem universitären Umfeld stammen. Alle diese Firmen sind klein im Vergleich zu ihren Kunden und haben das Problem, ihre Ideen und internen Standards gegen die vielen Sonderwünsche ihrer großen Kunden durchzusetzen. Diese Sonderwünsche führen durch Fragmentieren des Softwaremarktes zu hohen Kosten, weil überall kostenträchtige Spezialentwicklungen und entsprechende Wartungen erfolgen müssen. Kleine Anbieter haben es auch schwer, genügend Werbung zu machen, um ihre Innovationen im Markt erfolgreich zu platzieren.

4.4 FAZIT

Probleme der Infrastrukturplanung, der Personaleinsatzplanung etc. sind seit über 100 Jahren jeweils zeitaktuelle Fragestellungen in der Verkehrswissenschaft. Für ihre Lösung wurden die zur jeweiligen Zeit verfügbaren Werkzeuge der Technik und Mathematik eingesetzt. Im Nachruf auf den bedeutenden Verkehrswissenschaftler Rüger schrieb der Dekan der verkehrswissenschaftlichen Fakultät Dresden:

„Die Nutzung der wissenschaftlichen Erkenntnisse zur Verbesserung der Arbeit der Nahverkehrsbetriebe, ihre Umsetzung und Handhabung in der Planungs- und operativen Arbeit waren stets seine Maxime.“

Rüger [31] in der DDR und Lehner [24] in der Bundesrepublik haben in den siebziger Jahren richtungsweisende Werke zur Verkehrsplanung geschrieben, in denen im Prinzip ein Verkehrsbetrieb mathematisch modelliert wird. Damals mussten die Entscheidungsprobleme auf (aus heutiger Sicht) einfache, praxisadäquate Formeln reduziert werden, die auch heute noch vernünftige Handlungsanweisungen für die Praxis darstellen. Die Fortschritte der Mathematik ermöglichen nun aber völlig neue Herangehensweisen, bei denen Entscheidungen als Lösungen komplexer Netzwerkmodelle entstehen. Ziel muss es sein, diese Mathematik konsequent weiterzuentwickeln und damit im Sinne Rügers und Lehnners die Planungsabläufe kontinuierlich zu verbessern, so dass Ressourcen geschont, der Einsatz von Steuermitteln reduziert und die Kunden mit bedarfsgerechten und nutzerfreundlichen Verkehrsangeboten versorgt werden.

Wir sehen den Übergang von der formelgestützten Planung zu der hier skizzierten modellgestützten mathematischen Planung als einen ähnlichen Sprung wie von der Braunschen Röhre zum Transistor, vom Reißbrett zum CAD-System und sind davon überzeugt, dass sich Mathematik in immer stärkerem Maße als ein wesentlicher Produktionsfaktor im öffentlichen Verkehr etablieren wird.

Trotz der genannten Defizite in der Kommunikation zwischen den Disziplinen, in der Ausbildung und beim Transfer in die Industrie bietet Deutschland aus unserer Sicht die besten Voraussetzungen, dem Einsatz von mathematischen Planungsmethoden im öffentlichen Verkehr zum Durchbruch zu verhelfen. Die deutschen Verkehrssysteme gehören zu den besten der Welt, deutsche Softwarefirmen sind international führend, und

an deutschen Hochschulen beschäftigen sich in verschiedenen Fachbereichen Wissenschaftler auf höchstem internationalen Niveau mit der quantitativen Behandlung von Planungsproblemen im öffentlichen Verkehr. Im ÖV sind erhebliche Optimierungspotentiale vorhanden. Das Know-How, um diese zu heben, ist verfügbar. Wenn es gelingt, die unterschiedlichen Akteure zu einem koordinierten und zielgerichteten Zusammenspiel zu bringen, könnte Deutschland sich zu einem Schaufenster für effizienten Nahverkehr entwickeln.

5 VISIONEN UND HANDLUNGSEMPFEHLUNGEN

Unsere Vision ist, diskrete Mathematik und Optimierung als einen wesentlichen Produktionsfaktor in allen Bereichen der Verkehrsplanung zu etablieren. Die mathematischen Modelle und Methoden müssen so ausgebaut werden, dass sie die verschiedenen Planungsprozesse in der Praxis nutzerfreundlich und anwendungsadäquat unterstützen. Hier sind Fortschritte (in einigen Teilbereichen durchaus erhebliche) in der Mathematik erforderlich, aber auch die Schnittstellen zu den beteiligten Ingenieuren und Praktikern müssen durch gemeinsame Anstrengungen verbessert werden.

Wie wir im Vorhergehenden dargestellt haben, gibt es in fast allen Bereichen der ÖV-Planung Handlungsbedarf. Dieser erstreckt sich von mathematischer Modellierung, der dazu zu entwickelnden Theorie, über die Umsetzung dieser Theorie in Algorithmen bis hin zur Einführung von Optimierungssystemen in die Praxis. Dies sind Aufgaben für Jahrzehnte, bei denen nicht nur mathematische und innerbetriebliche Hindernisse überwunden, sondern auch politische Gegebenheiten berücksichtigt werden müssen.

Wir wollen hier zwei konkrete Maßnahmen vorschlagen, die relativ kurzfristig gestartet werden und vom BMBF/BMWi oder der DFG zusammen mit der Industrie finanziell auf den Weg gebracht werden können. Beide Maßnahmen zeichnen sich durch verkehrstypische Interdisziplinarität aus. Sie verknüpfen mathematische Forschung mit ingenieurwissenschaftlichem Know-How, betriebswirtschaftlichen Kenntnissen, Erfahrungen von Praktikern und beziehen die innovative Nutzung von IT-Systemen mit ein.

5.1 DISKRETE OPTIMALSTEUERUNG: UMLANUNG VON VERKEHRSSYSTEMEN BEI STÖRUNGEN IN ECHTZEIT

Störungen im Betriebsablauf von Verkehrssystemen sind praktisch unvermeidlich. Größere Störungen erfordern Umplanungsmaßnahmen, die unmittelbar nach dem Eintritt der Störungen eingeleitet werden müssen. Aus mathematischer Sicht ist dies ein Thema der Online- bzw. der Echtzeit-Optimierung. Um Online-Optimierung sinnvoll einsetzen zu können, müssen viele Voraussetzungen erfüllt sein. Alle Daten über den geplanten und den aktuellen Zustand des Verkehrssystems müssen vorhanden sein. Daneben benötigt man Prognosen über die möglichen Entwicklungen des Systems und vorhandene Alternativen. Das Ziel ist, das Verkehrssystem wieder in einen Zustand zu bringen, so dass

Plan und Stand weitgehend übereinstimmen und die „Schäden“ so gering wie möglich sind.

Online-Planung wird heute in den Leitzentralen der Verkehrsbetriebe durch erfahrene Disponenten vorgenommen, die durch Betriebsleitsysteme und deren Visualisierungsmöglichkeiten unterstützt werden. Die Disposition erfolgt aber weitgehend „per Hand“. In den meisten Fällen wird hier mit Hilfe von Entscheidungstabellen bzw. einem Regelwerk gearbeitet, das klare Handlungsanweisungen gibt, aber keine Optimierung vornimmt. Die Einbeziehung von Zielen und nicht nur die Abarbeitung von Regeln soll durch Echtzeit-Optimierung verfolgt werden. Ein Umstieg auf derartige mathematische Planung erfordert insbesondere eine Anpassung der IT-Infrastruktur und die Verknüpfung verschiedenster Datenbestände, ein nicht einfaches informationstechnisches Unterfangen.

Ein grundsätzliches Problem der Online-Optimierung ist die Aufzeichnung von Daten, mit deren Hilfe sich der Ablauf analysieren und rekonstruieren lässt. Idealerweise müssten solchen Daten nicht nur den Betriebsablauf aufzeichnen, sondern auch die Ursachen für Störungen und die Dispositionsmaßnahmen vermerken. Solche Daten sind in der Regel nicht vorhanden. Ein Forschungsprogramm zur Online-Optimierung sollte bereits auf dieser Ebene ansetzen und eine qualitativ hochwertige Datengrundlage schaffen.

Diese Daten müssen mit Simulationssystemen verknüpft werden können, mit denen sich exemplarisch Verifikationen einzelner Szenarien durchführen lassen. Solange es noch keine in der Praxis erprobte Online-Theorie gibt, ist der Einsatz von Simulationswerkzeugen zur Überprüfung von Lösungsansätzen erforderlich. Im Bahnbereich ist die Betriebssimulation eine langjährig geübte Praxis, die mit einer ganzen Reihe von Systemen, wie RailSys, OpenTrack, etc. realitätsnah durchgeführt werden. Im Bus- und Flugverkehr fehlen solche Simulationssysteme noch.

Die zu entwickelnden mathematischen Echtzeit-Algorithmen sollten die langjährigen Erfahrungen der Disponenten berücksichtigen und natürlich ihre heuristischen Regeln zur Systemumplanung einbeziehen. Zusammenarbeit ist hier erforderlich zwischen Mathematikern, Datenverarbeitungsfachleuten und erfahrenen Disponenten. Darüber hinaus sehen wir einen möglichen Fortschritt darin, dass mathematische Algorithmen den Zustand von Verkehrssystemen im Sinne eines Regelkreises überwachen und steuernd auch präventiv eingreifen, wenn Zustandsänderungen sichtbar oder prognostiziert werden, die den Betrieb beeinträchtigen können. Ideen dieser Art könnte man unter dem Begriff „Diskrete Optimalsteuerung“ fassen. Gemeint ist dabei der fortlaufende Einsatz von Optimierungsalgorithmen, um Verkehrssysteme im Plan zu halten. Solche Fragestellungen werden z. B. in der Mechanik und der Verfahrenstechnik betrachtet und mathematisch unter den Begriffen Kontrolltheorie und Optimalsteuerung behandelt. Sie werden typischerweise mit Methoden der Differentialgleichungstheorie bearbeitet. Das neue im vorliegenden Fall ist, dass auch Methoden der diskreten Optimierung maßgeblich zum Einsatz kommen.

Projekte dieser Art hätte man schon vor 10 Jahren starten können. Damals waren jedoch die IT-technischen Voraussetzungen noch nicht gegeben. Heute besteht eine Chance, informationstechnisches Know-How, mathematische Methoden und betriebliche Erfahrung so zusammenzubringen, dass sie für alle am Verkehrsgeschehen Beteiligten Nutzen bringen. Die Spannweite dieses Projektes kann von der Echtzeit-Optimierung von U-Bahnen, Straßenbahn- und S-Bahn-Systemen, über die Echtzeit-Optimierung von Nahverkehrsbussen bis hin zur Echtzeit-Optimierung des gesamten deutschen Bahnnetzes oder des europäischen Luftraums führen, wobei klar ist, dass letztere sicherlich noch Zukunftsmusik sind. Für den Kunden ist wichtig, dass Echtzeit-Optimierung mit Fahrgastinformationssystemen verbunden wird, so dass die Fahrgäste einen direkten Nutzen davon haben.

5.2 MODELLINTEGRATION: ANGEBOTSPANUNG IM BUS- UND BAHNBEREICH

Wir haben über verschiedene Aspekte der Modellintegration gesprochen und wollen hier vorschlagen, in dem speziellen Bereich der Angebotsplanung im Bus- und Bahnverkehr verschiedene mathematische Modelle so zu integrieren, dass diese algorithmisch lösbar sind und zu verkehrsplanerisch sinnvollen Aussagen führen.

Genau wie im vorhergehenden Fall der diskreten Optimalsteuerung liegen die eigentlich benötigten Daten nicht in hinreichender Exaktheit vor.

Das Besondere an der Angebotsplanung ist die enge Verzahnung verschiedener Aspekte. Wir haben bereits ausgeführt, dass die Prognose von Nachfrage ein stochastisches Problem ist, während der Entwurf von Netzen auf diskrete Optimierungsprobleme führt und die Festlegung von Tarifen ein Thema der nichtlinearen gemischt-ganzzahligen Optimierung ist. Die Integration von Modellen aus verschiedenen mathematischen Disziplinen ist deshalb eine notwendige Bedingung für die Entwicklung erfolgreicher Ansätze zur Angebotsplanung. Modellintegration ist in der Regel hoffnungslos schwierig, jedoch erscheinen uns in diesem konkreten Fall aufgrund der speziellen Strukturen erfolgreiche Lösungsansätze möglich zu sein.

Erste Schritte in diese Richtung wurde z. B. bei Straßenmautsystemen durch den Einsatz von sog. Bilevel-Optimierung erzielt. Im Bereich der Preisplanung gibt es Ansätze, nichtlineare Optimierungsmodelle mit stochastischen Nachfragemodellen (Discrete Choice Logit-Modelle) zu kombinieren.

Das Vorgehen würde sich an historisch gewachsenen Planungsstrukturen orientieren, aber über die Durchrechnung einiger weniger Szenarien hinaus können diese Modelle Rückkopplungseffekte berücksichtigen und Substitutionseffekte z. B. durch Betrachtung des Individualverkehrs mit einbeziehen.

6 LITERATUR

- [1] R. E. Anbil, B. Patty Gelman, and R. Tanga. Recent advances in crew-pairing optimization at american airlines. *Interfaces*, 21:62–74, 1991.
- [2] Cynthia Barnhart, Amy M. Cohn, Ellis L. Johnson, Diego Klabjan George L. Nemhauser, and Pamela H. Vance. Airline crew scheduling. In Randolph W. Hall, editor, *Handbook of Transportation Science*, pages 517–560. Kluwer, Boston, 1999.
- [3] Cynthia Barnhart, Ellis L. Johnson, George L. Nemhauser, Martin W. P. Savelsbergh, and Pamela H. Vance. Branch-and-price: Column generation for solving huge integer programs. *Operations Research*, 46(3):316–329, 3 1998.
- [4] Berliner Verkehrsbetriebe AöR. Geschäftsbericht 2006. <http://www.bvg.de/index.php/de/Common/Document/field/file/id/1409>, 2006.
- [5] Andreas Bley and Marcus Pattloch. Modellierung und Optimierung der X-Win Plattform. *DFN-Mitteilungen*, 67:4–7, 2005.
- [6] Ralf Borndörfer, Martin Grötschel, and Andreas Löbel. Duty scheduling in public transit. In Willi Jäger and Hans-Joachim Krebs, editors, *MATHEMATICS – Key Technology for the Future*, pages 653–674. Springer Verlag, Berlin, 2003.
- [7] Ralf Borndörfer, Martin Grötschel, and Marc E. Pfetsch. A column-generation approach to line planning in public transport. *Transportation Science*, 41(1):123–132, 2007.
- [8] Sven Brieger. Workflow oriented and integrated optimization. Vortrag auf der AGI-FORS Konferenz, Mai 2005.
- [9] Thomas M. Cook. Sabre soars. *ORMS Today*, 25(3):27–31, 6 1998.
- [10] J. R. Daduna and A. Wren, editors. *Computer-Aided Transit Scheduling*, Lecture Notes in Economics and Mathematical Systems. Springer Verlag, Berlin, 1988.
- [11] Joachim R. Daduna, Isabel Branco, and José M. Pinto Paixão, editors. *Proc. of the Sixth International Workshop on Computer-Aided Scheduling of Public Transport (CASPT), Lisbon, Portugal, 1993*, volume 430 of *Lecture Notes in Economics and Mathematical Systems*, Berlin, Heidelberg, 1995. Springer Verlag, Berlin.

- [12] Camil Demetrescu, Andrew Goldberg, and David Johnson, editors. *9th DIMACS Implementation Challenge – Shortest Path*, Piscataway, New Jersey, 2006. DIMACS.
- [13] Guy Desaulniers and Jaques Desrosiers Marius M. Solomon, editors. *Column Generation (Gerad 25th Anniversary)*. Springer Verlag, Berlin, 2005.
- [14] M. Desrochers and J.-M. Rousseau, editors. *Computer-Aided Transit Scheduling*, Lecture Notes in Economics and Mathematical Systems. Springer Verlag, Berlin, 1992.
- [15] Deutsche Bahn AG. Geschäftsbericht 2006. http://www.db.de/site/bahn/de/unternehmen/investor_relations/finanzberichte/geschaeftsbericht/geschaeftsbericht_2006.html, 2006.
- [16] Deutsche Lufthansa AG. Geschäftsbericht 2006. http://konzern.lufthansa.com/de/downloads/presse/downloads/publikationen/lh_gb_2006.pdf, 2006.
- [17] Edsger Wybe Dijkstra. A note on two problems in connection with graphs. *Numerische Mathematik*, 1:269–271, 1959.
- [18] Forschungsgesellschaft für Strassen- und Verkehrswesen. *Heureka '05: Optimierung in Verkehr und Transport*, Köln, 2005.
- [19] Michael R. Garey and David S. Johnson. *Computers and intractability*. Freeman, 1979.
- [20] Martin Grötschel, Sven O. Krumke, Jörg Rambau, and Luis M. Torres. Making the yellow angels fly: Online dispatching of service vehicles in real time. *SIAM News*, 35(4):10–11, 2002.
- [21] Karla L. Hoffman and Manfred Padberg. Solving airline crew scheduling problems by branch-and-cut. *Management Science*, 39(6):657–682, 1993.
- [22] Diego Klabjan, Ellis L. Johnson, George L. Nemhauser, Eric Gelman, and Srini Ramaswamy. Solving large airline crew scheduling problems: Random pairing generation and strong branching. *Computational Optimization and Applications*, 20(1):73–91, 2001.
- [23] John O. Ledyard, Mark Olson, David Porter, Joseph A. Swanson, and David P. Torma. The first use of a combined-value auction for transportation services. *Interfaces*, 32(5):4–12, 2002.

- [24] Friedrich Lehner. *Der maximale Wirkungsgrad des Personaleinsatzes*. Alba Verlag, Düsseldorf, 1978.
- [25] Christian Liebchen. *Periodic Timetable Optimization in Public Transport*. Dissertation, Technische Universität Berlin, 2006.
- [26] K. Littlewood. Forecasting and control of passenger bookings. In *AGIFORS Symposium Proceedings*, volume 12, Nathanya, Israel, 1972.
- [27] Andreas Löbel. *Optimal Vehicle Scheduling in Public Transit*. Shaker Verlag, Aachen, 1997. Dissertation, Technische Universität Berlin.
- [28] Martin Müller-Elscher. Die MICROBUS-Optimierungskomponenten im Überblick, Vortrag auf dem 1. PSYRUS-Anwenderforum, 30. 6. 2005, Fulda.
- [29] Marika Neumann. Mathematische Preisplanung im ÖPNV. Diplomarbeit, TU Berlin, 2005.
- [30] Partner für Berlin Gesellschaft für Hauptstadt-Marketing GmbH. Berlinbrief, 2003. ISSN 1611-3284.
- [31] Siegfried Rüger. *Transporttechnologie städtischer öffentlicher Personenverkehr*. Transpress, Berlin, 1986.
- [32] Barry C. Smith, John F. Leimkuhler, and Ross M. Darrow. Yield management at american airlines. *Interfaces*, 22(1):8-31, 1992.
- [33] Andreas Sturmowski. Notwendigkeit der effizienzsteigerung unter einsatz von it-tools. In *Proceedings der IVU-Konferenz IT im ÖPNV am 05.10.2007*. IVU Traffic Technologies AG, Berlin, 9 2007.
- [34] Kalyan T. Talluri and Garrett J. van Ryzin. *The Theory and Practice of Revenue Management*, volume 68 of *International Series in Operations Research and Management Science*. Kluwer, Boston, MA, 2005.
- [35] Stefan Voss and Joachim Daduna, editors. *Computer-Aided Scheduling of Public Transport*, Lecture Notes in Economics and Mathematical Systems. Springer Verlag, Berlin, 2001.
- [36] N. H. M. Wilson, editor. *Computer-Aided Transit Scheduling*, Lecture Notes in Economics and Mathematical Systems. Springer Verlag, Berlin, 1999.

- [37] Gang Yu, editor. *Operations Research in the Airline Industry*. Kluwer, Dordrecht, 1998.
- [38] Gang Yu, Michael Argüello, Gao Song, Sandra M. McCowan, and Anna White. A new era for crew recovery at continental airlines. *Interfaces*, 33(1):5-22, 2003.

> MIT MATHEMATIK ZU MEHR INTELLIGENZ IN DER LOGISTIK

ROLF H. MÖHRING UND MICHAEL SCHENK

1 EXECUTIVE SUMMARY

Die Lösung logistischer Probleme ist ein wichtiger Aspekt menschlichen Handelns seit Menschen gemeinsam zielgerichtet tätig wurden. Die Grundlagen dessen, was wir heute Logistik nennen, entstammen dem militärischen Bereich. So basierte z. B. das römische Imperium in starkem Maße auf militärisch-logistischen Glanzleistungen. Ob damals bereits mathematische Überlegungen eine Rolle spielten, wissen wir heute nicht. Jedoch versuchte z. B. Napoleon, der mit den bedeutendsten Mathematikern seiner Zeit befreundet war, den Transport seiner Truppen und die Verbreitung von Informationen zu optimieren und strategisch einzusetzen.^{1, 2}

Seit der Zeit Napoleons hat sich die Funktion der Logistik tiefgreifend verändert und in den letzten vierzig Jahren hat sie sich von einer stark auf physische Abläufe eines Unternehmens konzentrierten Funktion zu einem ganzheitlichen, prozess- und kundenorientierten Managementkonzept und Führungsinstrument entwickelt. Womit sich dann auch die Aufgaben der Mathematik in der Logistik stetig verändert haben. Eine Übersicht über die zeitliche Entwicklung der Logistik und den damit verbundenen Optimierungsbereichen ist in Abb. 1 aufgeführt.

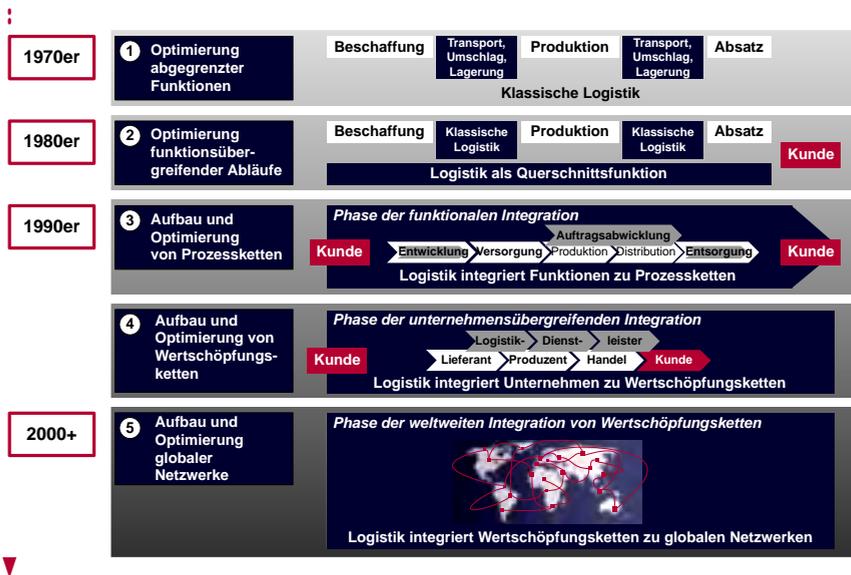
Die Logistik ist heute mit einem Brutto-Umsatzvolumen von über 166 Mrd. EUR und mehr als 2,6 Mio. Beschäftigten in Deutschland die drittgrößte Branche nach dem Handel und der Automobilindustrie ([6]). Das jährliche Wachstum beträgt dabei 3–10 %. Dieses Wachstum ist nicht nur quantitativ sondern auch qualitativ, da die Logistik vor immer neuen Herausforderungen steht, welche am Ende des Artikels noch genauer dargestellt werden.

Um verstehen zu können, welche Beiträge die Mathematik zur Logistik leistet, ist es nützlich, zwei der vielen Definitionen der Logistik, die durch die stetige Entwicklung entstanden sind, zu kennen. Zum einen ist es die Seven-Rights-Definition von Plowman,

¹ <http://www.geophys.tu-bs.de/geschichte/laplace.htm>, online: 19. 6. 2008

² <http://www.napoleon-online.de>, online: 19. 6. 2008

Abb. 1: Entwicklung der Logistik (Quelle: [2, S. 2])



welche besagt, dass Logistik die Verfügbarkeit des richtigen Gutes, in der richtigen Menge, im richtigen Zustand, am richtigen Ort, zur richtigen Zeit, für den richtigen Kunden, zu den richtigen Kosten sichert.³ Und zum anderen ist es eine in Anlehnung an die Definition von Baumgarten, dabei „umfasst die Logistik in einem Unternehmen die ganzheitliche Planung, Steuerung, Koordination, Durchführung und Kontrolle aller unternehmensinternen und unternehmensübergreifenden Güter- und Informationsflüsse. Die Logistik stellt für Gesamt- und Teilsysteme in Unternehmen prozess- und kundenorientierte Lösungen bereit.“⁴

Im folgenden Abschnitt werden exemplarisch einige Beispiele aufgeführt, bei denen der Einsatz von neuen mathematischen Methoden die Logistik verbessert und damit das jeweilige Unternehmen zum Erfolg geführt hat. Dabei wird zu Anfang ein Prozesskettenmodell eines Wertschöpfungsprozesses dargestellt, um die jeweiligen Beispiele und somit auch die Aufgaben der Mathematik besser einordnen zu können. In Abschnitt 3 wird dann detaillierter und ein wenig abstrakter auf den Status quo der Logistik und ihren

³ <http://www.fml.mw.tum.de/fml>, online: 6. 6. 2008

⁴ <http://www.bvl.de>, online: 6. 6. 2008.

Bezug zur Mathematik eingegangen. Am Ende werden noch die momentanen Stärken und Schwächen und einige Visionen und Herausforderungen genannt.

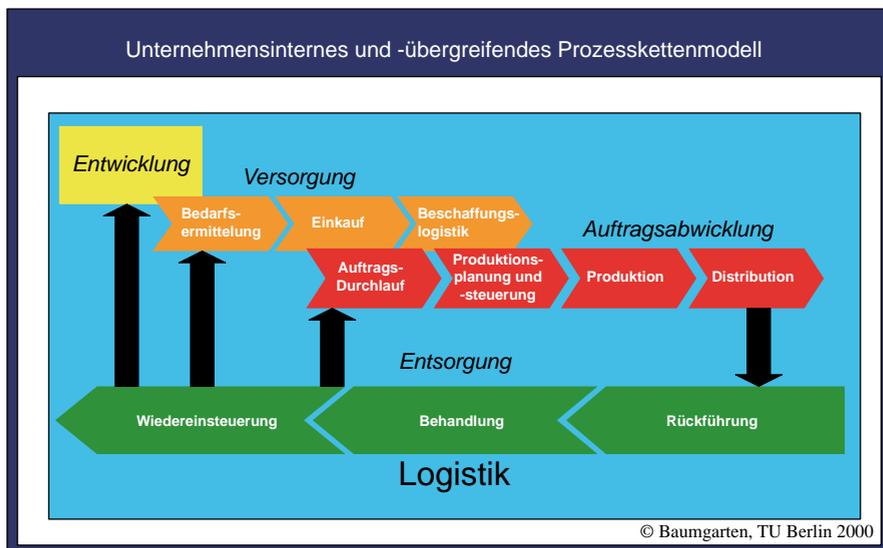
2 ERFOLGSBEISPIELE

Ein Wertschöpfungsprozess besteht in fast jedem Unternehmen aus den vier idealtypischen Prozessketten Entwicklung, Versorgung, Auftragsabwicklung und Entsorgung. Diese bilden zusammen einen Prozesskreislauf, bei dem sie sequenziell und parallel ineinander greifen. Der Aufbau des Prozesskettenmodells und die Unterteilung der Prozessketten sind in Abb. 2 dargestellt.

2.1 LOGISTIKKETTEN

Nicht nur die Versorgung eines Unternehmens besitzt ein Versorgungsnetz, sondern auch jede andere Prozesskette hat als Basis ein Versorgungsnetz, das aus Lieferstellen und Empfangsstellen besteht. Diese einzelnen Stellen sind jeweils durch Lieferketten miteinander verbunden, und werden je nach Prozesskette Versorgungs-, Beschaffungs-, Transport-, Fracht-, Beförderungs-, Entsorgungs- oder auch allgemein eine Logistikkette genannt ([11]). Um ein optimales Versorgungsnetz zu bekommen, müssen die Liefer-

Abb. 2: Unternehmensinternes und -übergreifendes Prozesskettenmodell (Quelle: [2, S. 7])



und Empfangsstellen bestmöglich durch die Lieferketten miteinander verbunden werden, was eine Optimierungsaufgabe für die Mathematik darstellt. Die folgenden Beispiele kommen aus verschiedenen Bereichen, die zu einem optimalen Versorgungsnetz beitragen.

Eine große Herausforderung entsteht bei der Planung regelmäßig wiederkehrender sehr großer Transportleistungen, wie sie bei den Verbunddienstleistern der Post, dem Gütertransport oder dem Personentransport im Bahn- und Flugverkehr anfallen. Die große Aufgabe dieser Dienstleister besteht darin, Entscheidungen zu Standorten und Flotten mit Entscheidungen zum Transport integriert zu optimieren (Routing und Scheduling).

Die Konzeption dafür geeigneter Verteilstrukturen und die effiziente Organisation des Ablaufs in diesen Ketten basieren auf der Kombination von neuen Techniken zur Datenerfassung und Kommunikation (GPS-, Barcodes-, RFID-, EDV-Vernetzung) und ausgefeilten mathematischen Optimierungsverfahren wie Netzwerkalgorithmen und (ganzzahlige) lineare Optimierung.

Die konsequente Ausnutzung der oben genannten Erkenntnisse ist ein bedeutender ökonomischer Faktor. Durch Einsatz solcher Methoden (und der Liberalisierung des Marktes) hat sich z. B. die Deutsche Post in weniger als 15 Jahren von einer Bundespost zum Weltkonzern gewandelt. Durch die Übernahme des britischen Logistikunternehmens Exel im Dezember 2005 ist die Deutsche Post AG Weltmarktführer in den Bereichen Luft-, See- (beide DHL Global Forwarding) und Kontraktlogistik (DHL Exel Supply Chain) mit einem Jahresumsatz von 55 Mrd. Euro geworden.⁵ Andere Erfolgsbeispiele sind die Optimierung der Transportleistungen im Luftverkehr⁶, in der Containerlogistik⁷ und im Bahnverkehr. So wurde der renommierte Edelman Award der INFORMS, der amerikanischen Gesellschaft für Operations Research, in diesem Jahr für die Optimierung des Fahrplans der Niederländischen Bahn (Nederlandse Spoorwegen) verliehen.⁸

Große Unternehmen haben die Bedeutung der Optimierung von Logistiknetzen erkannt und hierfür eigene Lehrstühle an Universitäten gestiftet, etwa den Deutsche Post Lehrstuhl an der RWTH Aachen oder das DB Logistics Lab an der Technischen Universität Berlin.

2.2 TRANSPORT

Im Transportbereich hat die Logistik ihren Ursprung und hier gab es bereits früh Anwendungen von mathematischen Methoden. Diese Methoden werden, da sehr praxisrelevant, oft auch dem Operations Research zugeordnet.

⁵ http://www.dpwn.de/dpwn?skin=hi&check=yes&lang=de_DE&xmlFile=2004705, online: 2. 7. 2008

⁶ <http://web.mit.edu/airlines/www/index2.htm>, online: 2. 7. 2008

⁷ <http://www.hhla.de/Altenwerder-CTA.64.0.html>, online: 2. 7. 2008

⁸ <http://www.informs.org/index.php?c=401&kat=Franz+Edelman+Award>, online: 02.07.2008

Ein klassisches Beispiel ist die *Tourenplanung*, deren Aufgabe es ist, für einen kleinen Zeitraum die verschiedenen Transportvorgänge mit dem Ziel der Fahrwegoptimierung aufeinander abzustimmen ([12]).

Um dieses Problem lösen zu können, führt man es auf ein mathematisches Modell zurück, das dann optimiert werden kann. Dabei werden zwei Aspekte betrachtet, zum einen die Clusterung, die beschreibt, welche Aufträge zu einer Tour zusammengefasst werden, und zum anderen das Routing, das die Reihenfolge der Punkte innerhalb einer Tour beschreibt. Zielsetzung einer Tourenplanung ist z. B. die Minimierung der Anzahl der eingesetzten Fahrzeuge, der zurückgelegten Strecke, der Einsatzzeit oder einer komplexeren Kostenfunktion. Beim Standardproblem der Tourenplanung liegen alle Start- oder Zielpunkte in einem Depot und es steht dort eine begrenzte oder unbegrenzte Zahl von identischen Fahrzeugen mit begrenzter Kapazität zur Verfügung. Andere Varianten betrachten mehrere Depots oder beliebige Start- und Zielpunkte (sog. Pickup-and-Delivery-Probleme).

Die mathematischen Techniken für die Routenplanung haben sich in den letzten Jahrzehnten stark weiterentwickelt, da sich durch die rasante Entwicklung der Informationstechnologie neue und schnellere Möglichkeiten zur Implementierung boten. Somit sind die mathematischen Modelle Basis vieler Softwarepakete, siehe <http://www.wior.uni-karlsruhe.de/bibliothek/Vehicle/com/> (18. 6. 2008). Die Verfügbarkeit digitaler Straßennetze und die Nutzung von GPS haben in Kombination mit mathematischen Verfahren den Erfolg solcher Systeme begründet.

2.3 PRODUKTIONSLOGISTIK

Die Produktionslogistik beschäftigt sich mit der Gesamtheit aller logistischen Tätigkeiten, Maßnahmen und Themenstellungen, welche sich aus der Waren- bzw. Leistungserbringung ergeben. Die Aufgaben der Produktionslogistik sind Planung, Steuerung, Transport und Lagerung von Rohmaterialien, Hilfs- und Betriebsstoffen, Kauf- und Ersatzteilen oder Halbfertig- und Fertigprodukten.⁹ Somit bildet sie den Kern eines jeden Industrieunternehmens. Im Vordergrund steht die Optimierung des gesamten Produktionssystems.

Daher werden wie in der Transportlogistik bereits seit langem mathematische Methoden des Operations Research eingesetzt. Beispiele sind Scheduling, Lagerhaltung, Lineare (ganzzahlige) Optimierung und Flussmodelle. Oft sind diese Modelle sehr speziell auf einen bestimmten Produktionstyp zugeschnitten, etwa bei der Steuerung von Werkzeugmaschinen, der Minimierung von Rüstzeiten oder der Steuerung von Hochregallagern. Mit zunehmender Verfügbarkeit von Daten, speziell über ERP-Systeme, geht der Trend zum Einsatz allgemeiner Techniken unter dem Stichwort APS-Systeme (Advan-

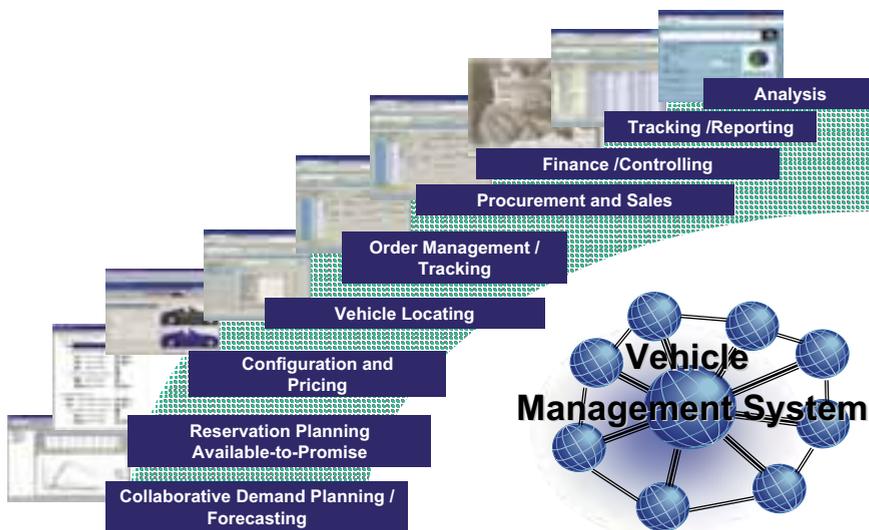
⁹ <http://www.bvl.de/taschenlexikon/>, online: 2. 7. 2008

ced Planning and Scheduling Systeme) entlang der gesamten Produktionskette. Abb. 3 zeigt diese Kette am Beispiel der Porsche AG.

Meist geht es dabei um die zeitliche Planung unter komplexen Nebenbedingungen und unsicheren Daten. APS-Systeme suchen Lösungen auf Basis von Plandaten oder geschätzten mittleren Daten und adaptieren diese Lösungen, sobald die realen Daten vorliegen. Die komplexen Nebenbedingungen verlangen nach einer sehr allgemeinen Lösungstechnik, hierfür werden in der Regel ganzzahlige lineare Optimierung, Constraint Programming und Heuristiken herangezogen. Die Unsicherheit in den Ausgangsdaten verlangt nach Vorhersagetechniken und einer Validierung der Lösungstechniken, die in der Regel durch Simulationsmodelle geschieht, mit der die Eignung der Optimierungsmodelle bzgl. zufälliger Variation der Daten getestet wird. Diese Kombination von Vorhersage, Optimierung und Simulation bringt oft signifikante Erfolge.

So berichtet die Firma ILOG über den Einsatz ihres Tools ILOG Plant PowerOps bei Danone in einer komplexen Produktionsumgebung. Ein typischer Danone-Standort produziert zehn Zwischenprodukte, aus denen dann 120 Fertigprodukte auf 15–25 Produktionslinien erzeugt werden. Die einzelnen Produktionsschritte benutzen mehr als 100 Tanks für die Vorbereitung, Pasteurisierung, Fermentierung und Lagerung. Alle Schritte in dieser Produktionskette sind aus hygienischen und Haltbarkeitsgründen sehr

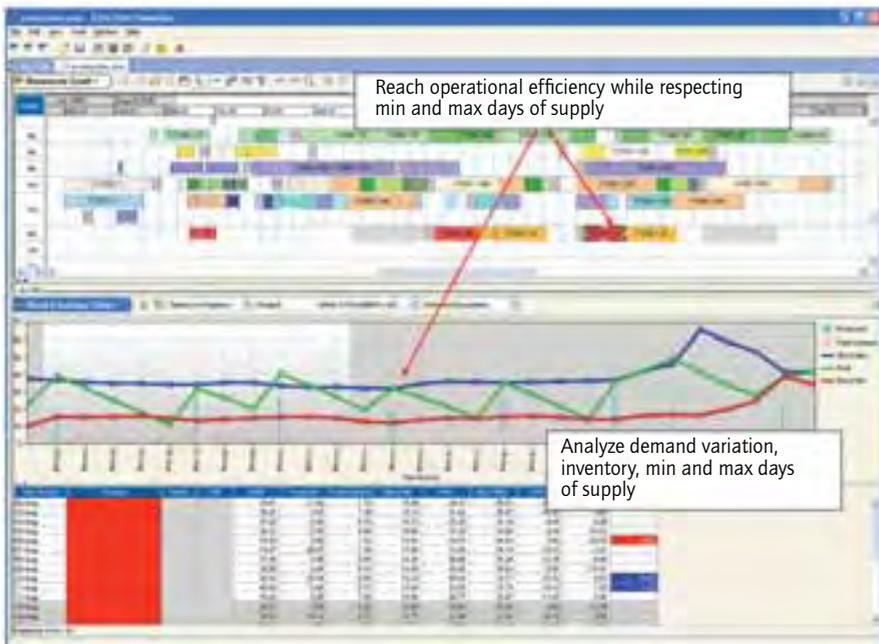
Abb. 3: SAP VMS – VehicleManagement System (Importeur) (Quelle: [15, S. 13])



zeitsensitiv und erfordern ein sorgfältiges Scheduling, bei dem Umrüstvorgänge (inklusive Reinigung), Kapazitäten der Tanks, Lieferzusagen und Änderungen der Nachfrage gleichzeitig zu beachten sind ([13]). Kriterien in solchen komplexen Situationen sind oft gegenläufig (z. B. Zeitersparnis beim Umrüsten versus Einhalten der Lieferfristen). Das Tool basiert auf ganzzahliger linearer Optimierung, Constraint Programming und Heuristiken und berechnet gute Produktionspläne in weniger als 15 min. Es wird unterstützt durch eine vollständige Anbindung an SAP, Möglichkeiten zur Sensitivitätsanalyse (what if analysis) und eine gute graphische Aufbereitung der Ergebnisse (siehe Abb. 4, [13, S. 16]).

Die Pharma-Firma Merck KGaA hat den Materialfluss für eine neu zu errichtende Fabrik zur Feststoffproduktion auf diese Weise geplant. Die Fabrik umfasst 10 000 m² Produktionsfläche auf sechs Ebenen und ist für maximal 1 000 t Tabletten pro Jahr ausgelegt. Mit Hilfe einer Simulationsstudie wurden verschiedene Modellvarianten bei unterschiedlichen Produktionsprogrammen untersucht. Auf dieser Basis wurde ein adäquates Produktionslogistikkonzept erstellt. Kern der Produktionsanlage ist die innerbetriebliche Logistik, die hoch automatisiert ist. Aufgrund dieses Logistikkonzeptes konnten

Abb. 4: Schedule in ILOG Plant PowerOps (Quelle: [13, S. 16])



über 80 % an Stellflächen eingespart werden und die manuellen Transporttätigkeiten um mehr als 70 % reduziert werden ([16]).

Die Beumer Maschinenfabrik GmbH und Co KG hat eine neue Leitstandssoftware für die Blech- und Sägebearbeitung implementiert, die eine optimierte Schnittstelle zwischen dem PPS-System, dem Lager und der Bearbeitungsmaschine geschaffen hat. Dadurch ergaben sich eine Verbesserung der Maschinenauslastung im Bereich Sägen um bis zu 30 % und im Bereich Lasern um bis zu 60 %, eine Reduzierung der Durchlaufzeit um 20 % und mehr als 2000 Stunden mannl. Betrieb der Laserschneidanlagen ([19, S. 52]).

2.4 STEUERUNG LOGISTISCHER NETZE

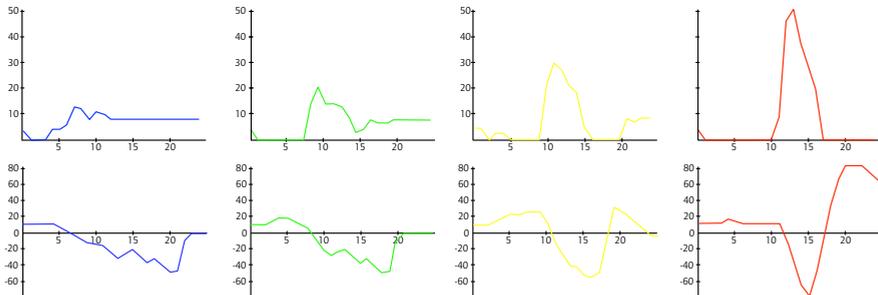
Die bisherigen Anwendungen waren weitgehend noch isoliert und global, also von einem einzigen Unternehmen oder Planer gesteuert. Die Berücksichtigung verschiedener, miteinander selbstständig agierender überbetrieblicher Teilnehmer schafft eine deutlich höhere Komplexität und führt zum Begriff der logistischen Netzwerke.¹⁰ Diese Netzwerke enthalten vielfältig verflochtene und übereinander geschichtete Ebenen mit vielen Wechselwirkungen, darunter Flüsse von Materialien und Produkten, Informationsflüsse von zugehörigen Auslöse-, Steuerungs-, Kontroll- und Dokumentationsinformationen und andere Aktivitätenketten, die allesamt die bereitgestellten Ressourcen beanspruchen.

Ein typisches Beispiel hierfür ist eine Lieferkette (Supply Chain) über mehrere Stufen mit autonomen Planern auf jeder Stufe. Dabei tritt ein Problem auf, dass von Procter & Gamble zuerst erkannt wurde. Der Großhändler bestellte bei Procter & Gamble sehr unterschiedlich große Mengen an Pampers-Windeln, obwohl die Nachfrage des Endkunden nahezu konstant war. Dadurch schwankte auch die Nachfrage von Procter & Gamble nach Vorprodukten sehr stark. Diese Schwankungen schaukelten sich immer weiter auf, weswegen Procter & Gamble begann, die Ursachen für diesen Effekt zu suchen. Dabei ergab sich, dass die Bestellungen der einzelnen Stufen keinen Zusammenhang mehr mit dem ursprünglichen Bedarf des Endkunden hatten, dessen Nachfrage ja konstant war. Durch den Vergleich der Nachfrage und Lagerbestände der einzelnen Stufen miteinander stellte sich heraus, dass die Schwankungen immer größer wurden, je weiter man sich in der Lieferkette vom Endkunden entfernte. Diesen Effekt in der Supply Chain nennt man Bullwhip Effekt. Die Gründe dafür liegen in den komplexen und dynamischen Abhängigkeiten der Supply Chain. Die „Akteure“ innerhalb der Kette treffen aufgrund falscher Wahrnehmungen und verzerrter Systeminformationen¹¹ lokal scheinbar rationale Bestell- und Produktionsentscheidungen, welche aber letztendlich nichts mit dem eigentlichen Bedarf des Endkunden zutun haben (siehe Abb. 5, [17]).

¹⁰ <http://www.logistik.wiso.uni-erlangen.de/german/profil/selbst/index.htm>, online: 2. 7. 2008.

¹¹ <http://beergame.uni-klu.ac.at/bullwhip.htm>, online: 2. 7. 2008

Abb. 5: Bullwhip Effekt. Oben: Nachfrage von links nach rechts. Unten: Lagerbestand von rechts nach links. (Quelle: [17])



Dieser Effekt ist mathematisch mit Mitteln der Wahrscheinlichkeitstheorie gut untersucht ([7]) und liegt darin begründet, dass die Teilnehmer entlang der Kette unterschiedliche zufällig fluktuierende Nachfrage- bzw. Bestellmengen beobachten (*Informationsasymmetrien*) und daher unverträgliche Entscheidungen treffen, die sich aufschaukeln können. Abhilfe schafft eine bessere Koordination wie sie in den Logistik-Konzepten JIT (Just in Time, Fertigung und Anlieferung von Teilen/Erzeugnissen erfolgen zum exakt geforderten Termin, ohne auf Lagerbestände zuzugreifen), ECR (Efficient Consumer Response, stellt dem Kunden Waren anhand seiner Informationen zur Verfügung) und APS (Advanced Planning and Scheduling, nutzt Optimierung und Vorhersage, berücksichtigt knappe Kapazitäten). Barcodes und RFID Tags spielen eine wichtige Rolle um zutreffende Vorhersagen zu machen und in Echtzeit korrigierend eingreifen zu können.

Das ITWM hat Methoden entwickelt, um in komplexen logistischen Netzen auch auf strategischer Ebene Entscheidungsunterstützung zu liefern. LibStrat-SCM wurde in der APO Software der SAP AG eingesetzt und bildet das Modul Supply Chain Design. Es nutzt effiziente Optimierungsalgorithmen und hat eine Schnittstelle zu CPLEX, dem Mixed Integer Programming Solver von ILOG.¹²

Die folgenden Beispiele zeigen eindrucksvoll, wie groß die Spareffekte durch bessere Steuerung von logistischen Netzen sind.

Für sein Konzept zur Marktversorgung im Bereich Pneumatik bekam die Festo AG & Co. KG 2003 den deutschen Logistikpreis. Festo liefert Komponenten und Systeme für Automatisierung mit Pneumatik und befindet sich auf Wachstumskurs mit zurzeit ca. 1,5 Mrd. EUR Umsatz und weltweit ca. 12 000 Mitarbeitern. Mit dem prämierten Logistikkonzept gelang es Festo, die Liefertreue für Europa, wo eine Lieferung inner-

¹² http://www.itwm.fhg.de/opt/Dokumente/LibStrat-SCM_de.pdf, online: 2. 7. 2008.

halb von 19 Stunden garantiert wird, von 86,3 % im Jahr 1996 auf 97,7 % im Jahr 2002 zu erhöhen. Weiterhin wurden bei einem Umsatzwachstum von immerhin 55 % die dezentralen Bestände zwischen 1996 und 2002 um 83 % und die Gesamtbestände um 19 % gesenkt. Teil des Konzeptes war neben einer Standardisierung von Produkten und Prozessen, einer Klassifizierung der Produkte nach Logistiktypen und einer lernenden Organisation auch die Einführung von hierarchisch strukturierten mathematischen Planungs- und Prognoseverfahren, die zu einer optimalen Auslastung führen ([8]).

Die Würth Logistics AG koordiniert einen Großteil der Beschaffungslogistik zwischen den Lieferanten und den Gesellschaften des weit verzweigten Mutterkonzerns Würth, die mit mehr als 51 000 Beschäftigten ein Sortiment von mehr als 100 000 Produkten an rund 2.800.000 Handwerker und Werkstätten in 81 Ländern liefert. Die Würth Logistics AG besitzt keine eigenen LKWs, keine Flugzeuge und keine Schiffe. Die Transportaufträge werden vom Kunden in der Regel online in das Softwaresystem von Würth Logistics eingegeben und dort automatisch unter Berücksichtigung aller Artikel-, Kunden-, Termin- und Preisdaten nach einer langen, komplexen Reihe von Optimierungsschritten, u. a. Routenoptimierung, berechnet und müssen von den Mitarbeitern nur noch per Knopfdruck freigegeben werden. Das Bestellsystem und das Transportsystem sind voll integriert. Das Gesamtsystem hat zu Einsparungen in Millionenhöhe geführt. Die Würth-Tochter Mepla-Alfit spricht von ca. 20 % Einsparungen bei den Logistikkosten ([10]).¹³

Die ersten SCM-Aktivitäten der Daimler Chrysler AG im Werk Sindelfingen wurden durch Versorgungsengpässe für Türinnenverkleidungen der C-Klasse im Zuge der Modellpflege 1997 gestartet. Aufgrund der unerwartet hohen Nachfrage konnten vom Systemlieferanten nicht genügend Türinnenverkleidungen bereitgestellt werden. Es musste eine zusätzliche Anlage installiert werden, deren Inbetriebnahme jedoch einige Wochen dauerte, so dass in den Werken Sindelfingen und Bremen ca. 2 000 Fahrzeuge der C-Klasse nicht produziert werden konnten und weitere 3 000 Fahrzeuge nachgerüstet werden mussten. Aufgrund dieser Erfahrung und der daraus gewonnenen Erkenntnis, dass reaktive Methoden in einstufigen Beschaffungsprozessketten nicht mehr ausreichend sind, wurde das SCM-Projekt, mit dem Ziel proaktive Logistikmethoden für die Produktion in Beschaffungsnetzwerken zu entwickeln, in die Wege geleitet. Als Pilotbeispiel diente die Türinnenverkleidung der E-Klasse. Die Datenaufnahme ergab, dass hinter dem Systemlieferanten noch ca. 100 weitere Unterlieferanten auf insgesamt sieben Ebenen standen. Das dynamische Verhalten des Netzwerkes für die Türinnenverkleidung der E-Klasse wurde mit einer Simulation dargestellt und analysiert. Die Auswirkungen von Programmveränderungen auf die einzelnen Stationen des Netzwerkes konnten untersucht werden und kritische Beschaffungsketten mit einer Sensitivitätsanalyse identifiziert werden. Mit Hilfe eines internetbasierten Softwaretools, dem IC-Tool (Information Control Tool), wird seit 1999 die gesamte Prozesskette vom Kunden bis zum 6th Tier Lieferanten abgebildet. Jeder Lieferant hat Zugriff auf Bruttobedarfe, Bestände und Kapazitäten in der

¹³ <http://www.wurth-logistics.com>, online: 18. 5. 2008.

gesamten Beschaffungskette. Zu erwartende Lieferschwierigkeiten werden online sofort angezeigt und der Systemlieferant hat somit frühzeitig die Möglichkeit zusammen mit den Unterlieferanten entsprechende Maßnahmen einzuleiten. Die Analyse des IC-Tools mündet in einer prozesskettenübergreifenden Kosten- und Nutzenanalyse und einem Beteiligungsmodell, so dass alle Lieferanten an den Einsparungen teilhaben. Es wurden Einsparungen von über 20 % der gesamten Logistikkosten entlang der Prozesskette erreicht. Es wird geschätzt, dass dieses SCM-Tool ca. 5–10 % der gesamten Beschaffungsprozessketten effizienter steuern kann ([9]).

3 LOGISTIK UND MATHEMATIK: STATUS QUO

3.1 ZUR ENTWICKLUNG DER LOGISTIK

Die Logistik hat sich seit den 70er Jahren zu einer wichtigen Kernkompetenz in Deutschland entwickelt. Durch die immer neu entstanden Aufgaben und Probleme musste sie sich ständig neu orientieren.

Wie in Abb. 1 zu sehen, beinhaltete die sogenannte klassische Logistik in den 70er Jahren überwiegend material- und warenflussbezogene Aufgaben. Schwerpunkte waren die Sicherstellung der Verfügbarkeit von Materialien und Waren innerhalb des Produktionsprozesses. Da die Logistik zu dieser Zeit in vorherrschende Unternehmensorganisationen eingebettet war, ergab sich eine Zersplitterung der material- und warenflussbezogenen Leistungsbereiche, was effizienzmindernde Arbeitsstrukturen und unwirtschaftliche Teilprozesse zur Folge hatte. In den 80er Jahren begann man den Schwerpunkt der Logistik auf die Schnittstellen zwischen den einzelnen Teilbereichen zu verlegen. Es wurden jetzt nicht nur, wie in den 70er Jahren, abgegrenzte Funktionen, sondern auch funktionsübergreifende Abläufe optimiert. In der dritten Entwicklungsstufe (90er Jahre) veränderte sich die Logistik von einer funktionsorientierten zu einer flussorientierten Betrachtungsweise. Sie entwickelte sich zu einer unternehmensweiten und -übergreifenden Koordinationsfunktion und konzentrierte sich hauptsächlich auf die Optimierung von Material- und Warenflüssen. Durch die neuen Entwicklungen in der Informationstechnologie war es nun möglich Informationsflüsse zu verbessern und Informationsdefizite abzubauen. Bei der Planung und Koordination von Waren-, Material- und Informationsströmen wurden nicht mehr nur die Bereiche Beschaffung, Produktion und Distribution betrachtet, sondern auch deren Entwicklung und Entsorgung. Seit dem Jahr 2000 beginnen die Unternehmen mit ihren Wertschöpfungspartnern zu kooperieren, um die gesamte Wertschöpfungskette vom Zulieferer bis zum Endkunden zu optimieren (Supply Chain Management). Dabei werden integrierte Wertschöpfungsketten zu globalen Netzwerken ausgebaut und optimiert ([2, S. 3, 4]).

Heute agieren im Bereich der Logistik-Dienstleistungen ca. 60 000 Unternehmen, darunter Weltkonzerne wie die Deutsche Post (15,6 Mrd. EUR Jahresumsatz) oder die Deutsche Bahn (5,5 Mrd. EUR), aber überwiegend mittelständische Unternehmen un-

ter einem Jahresumsatz von 10 Mio. EUR. Diese Unternehmen sind zu 55 % aus den Bereichen Industrie und Handel und zu 45 % Logistik-Dienstleister ([6, S. 22]).

Durch diese Entwicklung ist die Logistik von einem strategischen Instrument der Unternehmensführung zu einer Managementphilosophie und einem wichtigen Wettbewerbsfaktor geworden. Sie zeigt je nach Leistungsbereich ein überdurchschnittliches Wachstum von 3–10 % pro Jahr und investiert ca. 15 Mrd. EUR jährlich.¹³

Doch noch ist die Entwicklung der Logistik nicht zu Ende, und es werden in den nächsten Jahren viele neue Herausforderungen und Aufgaben auf die Logistik zu kommen. Dabei werden die Theorien der Mathematik weiterhin eine wichtige Funktion einnehmen. Der momentane Status Quo der Mathematik in der Logistik wird im Folgenden dargestellt.

3.2 MATHEMATIK IN DER LOGISTIK

Die Mathematik hat schon früh über das Operations Research Einzug in die Logistik gefunden. Abb. 6 stellt anhand der Bereiche und Tätigkeiten der Logistik die dafür relevanten Gebiete der Mathematik dar.

Abb. 6 zeigt, dass die Mathematik bereits in vielen Aufgabenstellungen der Logistik eingesetzt wird. Die Leistungsfähigkeit der heutigen Algorithmen – speziell in der ganzzahligen linearen Optimierung –, ihre Verfügbarkeit in kommerziellen Solvern wie CPLEX oder XPRESS und nicht zuletzt die enorme Verbesserung der Datenlage haben entscheidend dazu beigetragen.

Dennoch zeigen sich auch klare Grenzen, z. B. bei Steuerungsprozessen, Frühwarnsystemen oder bei der Reaktion auf Störungen (Disruption Management).

Eine Ursache hierfür liegt in der unsicheren oder unvollständigen Information zum Planungszeitpunkt. Für die Mathematik bedeutet dies den Übergang von deterministischen zu stochastischen Modellen, für die es zwar bereits Theorien und Methoden gibt, die aber einen wesentlich höheren Berechnungsaufwand erfordern. Darüber hinaus sind die Ergebnisse mit Wahrscheinlichkeit behaftet (wahrscheinliche Szenarien, Mittelwerte), wofür Entscheidungsträgern oft die richtige Vertrautheit fehlt.

Eine weitere Ursache liegt in der Notwendigkeit, bei vielen Prozessen Entscheidungen in Echtzeit – also in wenigen Sekunden – zu treffen. Dies bedeutet eine starke Einschränkung für Lösungsverfahren, da sie in so kurzer Zeit in der Regel keine optimalen Entscheidungen treffen können. Das ist jedoch kein Defizit der Mathematik; sie kann (in der Komplexitätstheorie) sogar nachweisen, dass ein Tradeoff zwischen Lösungsgüte und Rechenzeit besteht. Wenig Rechenzeit wirkt sich demnach nachteilig auf die Qualität der Ergebnisse auf.

Es zeichnen sich jedoch auch hier Fortschritte ab. Zum einen können in der Regel durch mathematische Analyse Qualitätsgarantien für bestimmte Algorithmen gegeben werden (beweisbar oder empirisch), zum anderen entwickeln sich neue Gebiete wie Approximations- und Online-Algorithmen sowie die Robuste Optimierung, die den

Abb. 6: Mathematik in logistischen Aufgabenstellungen

Bereiche und Tätigkeiten der Logistik → Logistische Aufgabenstellungen → Bereiche der Mathematik

Gestaltung	Planung	Betrieb		
A	B	C		
1. gesamtes Netzwerk			a. Sortimentsanalyse und Artikelklassifizierung - 1A,1B	• lineare (ganzzahlige) Optimierung - c, d, e, f, g, k, l, m, o, p
2. einzelner Standort			b. Bedarfsprognose - 9A, 9B	• heuristische Optimierungsmethoden - c, d, e, f, g, k, l, m, o, p
3. Transportsystem			c. Standortproblem - 1A, 1B	• Scheduling Algorithmen - l, p
4. Lieferkette			d. Auswahl von Transportwegen - 1A, 1B	• spezielle analytische Modelle - a, b, h, i, l, m
5. Lieferant			e. Gestaltung von Lieferketten - 1A, 1B	• Graphentheorie - c, d, e, f, k
6. Produktionsstätte			f. Tourenplanung - 3B, 3C, 4B, 4C	• Automatentheorie und Petri-Netze - o
7. OEM			g. Layoutplanung - 6A, 7A	• mathematische Statistik - a, b, i, k, n
8. Lager oder Umschlagpunkt			h. Lagerhaltungs- und Nachschubstrategien - 5C, 6C, 7C	• Warteschlangenmodelle - n
9. Kunde			i. Lieferstrategien - 5C, 6C, 7C	• kontinuierliche Simulationsmodelle - e, h, i
			j. Lagerorganisation und betriebsstrategien - 8B, 8C	• ereignisdiskrete Simulationsmodelle - g, h, i, k, l, n, o, p
			k. Gestaltung und Dimensionierung von Fahrzeugsystemen - 6A, 7A, 8A	
			l. Gestaltung von Kommissionierprozessen - 8A, 8B	
			m. Bildung von Logistikeinheiten - 8B, 8C	
			n. Grenzleistungen und Staufeffekte - 6A, 6B, 8A, 8B	
			o. Auftragsdisposition und Produktionsplanung - 5B, 5C, 6B, 6C, 7B, 7C	
			p. Reihenfolgeproblem - 6B, 6C, 7B, 7C	

besseren Umgang mit unvollständiger Information in Echtzeit erforschen bzw. kleine Störungen der Daten bereits in der Optimierung berücksichtigen und für Störungsfälle Recovery Methoden vorsehen. Solche Methoden halten bereits Eintritt in die Planung von Flug- und Bahnverkehr ([1, 14]).

4 ZUKÜNFTIGE HERAUSFORDERUNGEN

In vielen Bereichen der operativen Logistik wird der Zuwachs der so genannten Echtzeitreaktionsfähigkeit beobachtet. Diese Tendenz wird vor allem durch die schnelle Entwicklung von automatischen Identifikations- und Lokalisierungssystemen stimuliert. Ein System ist im logistischen Kontext echtzeitfähig, wenn es synchron zu der Zeit, die die Abläufe in der „realen Welt“ verbrauchen, Entscheidungen treffen kann. Im Idealfall werden beliebige Zustandsänderungen im System (Ereignisse) momentan erfasst, in Form von Protokollen gespeichert und mit Hilfe einer entsprechenden Software interpretiert und präsentiert. Die eigentliche Reaktionszeit hängt dann nur davon ab, ob der Mensch oder die Maschine für die Auswahl und Umsetzung von notwendigen Maßnahmen zuständig sind.

Die sogenannten autonomen Objekte der Logistik – Produkte, Verpackungen, Ladungsträger – erhalten eine elektronische Identität und werden auf Basis der RFID-Technologie drahtlos mit ihrer Umwelt vernetzt. Mit Hilfe von RFID-Sensoren und -Tags werden Daten während des gesamten Produktionszyklus – angefangen von der Herstellung über den Transport bis hin zu Auslieferung und Lagerung am Verkaufs- bzw. Verbrauchsort – dokumentiert und in regelmäßigen Abständen an weiterverarbeitende IT-Systeme zur Auswertung übermittelt.

Der Mangel an klaren und allgemein anerkannten theoretischen Konzepten und mathematischen Modellen zur Erfassung und Interpretation von Echtzeitdaten führt oft dazu, dass der Umfang und die Inhalte von erfassten Daten nicht in ausreichendem Maße auf die Ziele und Techniken der Prozessanalyse in logistischen Netzen und Materialflusssystemen orientiert sind. Nicht selten entsteht die Situation, bei der ein modernes (z. B. RFID-basiertes) Datenerfassungssystem sehr umfangreiche Datenmengen liefert, wobei sich die beiden betroffenen Seiten – Informatiker und Logistiker – nicht glücklich damit fühlen. Die Informatiker wissen nicht, wohin mit den Daten, und die Logistiker fragen sich „Was bringen uns, den Disponenten und Managern, diese Daten?“. Es fehlt noch immer eine mathematische Modellbasis, die einen erfolgreichen Dialog zwischen Logistikern und Informatikern bei der Konzipierung und Umsetzung von Echtzeitsystemen ermöglicht. Diese Basis kann z. B. geschaffen werden, wenn die zeitlichen Aspekte von zu beobachtenden und zu analysierenden Prozessen aus einer ereignisorientierten Sicht und die räumlichen Aspekte aus einer objektorientierten Sicht von beiden Seiten betrachtet werden ([18, S. 222–226]). Bei der Entwicklung von Grundkonzepten zu logistischen Echtzeitsystemen – Monitoring, Event Management und Frühwarnung – wird eine Mitarbeit durch Mathematiker dringend erforderlich.

Eine weitere Herausforderung in den nächsten Jahren wird das Supply Chain Event Management sein. Das SCEM beschäftigt sich nicht nur mit der Effizienz und Effektivität einer Lieferkette, sondern hauptsächlich mit der Optimierung und Steuerung, um eine nachhaltig stabile Lieferkette zu erhalten. Dabei ist ein wichtiges Element das ereignisorientierte Messen, was durch neue DV-Systeme ermöglicht wird. Hierbei wird nicht mehr in unregelmäßigen Abständen die Leistung überprüft, sondern das System benachrichtigt den Prozessverantwortlichen über Ausnahmestände und Handlungsnotwendigkeiten ([3]). Jedoch fehlen hier noch geeignete Methoden und Modelle um die relevanten Daten von den irrelevanten zu unterscheiden, zu erkennen und auszuwerten. Um neue Modelle zur Entwicklung von solchen Datenverarbeitungsprogrammen entwickeln zu können, werden Theorien aus der Mathematik benötigt, wie beispielsweise die der Mustererkennung oder des Data Mining.

5 VISIONEN UND HANDLUNGSEMPFEHLUNGEN

Die Auswirkungen der modernen RFID-Technologie auf logistische Prozesse in der näheren Zukunft beschreibt ausführlich der Reader „Internet der Dinge“ ([5]). Autonome Objekte, agentenbasierte Steuerung, selbstständig gesteuerte Systeme, intelligente Ladungsträger werden die Logistik in den nächsten Jahren bestimmen. Diese zukunfts-trächtigen Technologien sorgen dafür, dass Waren und Behältnisse immer eigenständiger und „intelligenter“ werden. Sie wissen, wohin sie müssen und steuern die Systeme, in denen sie sich bewegen.

Ein Beispiel für einen intelligenten Ladungsträger ist die im Fraunhofer Institut für Fabrikbetrieb und -automatisierung IFF Magdeburg im RFID- und Telematiklabor *Log-MotionLab* entwickelte Smart Box. Diese ist eine wiederverwendbare Kiste, die RFID-Antennenstrukturen und eine autarke Stromversorgung besitzt. Sie erkennt das Hinein- und Herausnehmen von Objekten, im Sinne einer ständigen Inventur. Auch kann diese über Zugangssysteme den Zugriff auf die Waren kontrollieren. Die Position des intelligenten Ladungsträgers sowie alle Zugriffsoperationen werden über GSM-Module an die Zentrale gesendet. Jedoch ist hier noch weiterer Forschungsbedarf vorhanden, um diese Technik auch auf andere Behälter, wie zum Beispiel der Europalette zu übertragen ([18, S. 77 ff.]).

Weitere Ideen und Beispiele zum Einsatz von RFID in verschiedenen Branchen sind zum einen die Verbesserung der Fälschungssicherheit von Medikamenten durch den Einsatz von RFID-Labeln auf den Verpackungen. Zum anderen die Nutzung von RFID in der Luftfahrt, beispielsweise zur Ortung von verloren gegangenen Gepäck oder aber den verkürzten Check-in Zeiten, indem man Flugtickets mit RFID bestückt ([4, S. 157 ff.]).

Mit der stets besser werdenden Datenlage, der besseren Leistung der Rechner und der Nutzung neuer mathematischer Einsichten und Verfahren ist das Potenzial für weitere Erfolge in der Logistik gegeben. Allerdings bedarf dies einer gemeinsamen Anstrengung von Logistikern, Mathematikern und Informatikern, denn nur durch Interdisziplinä-

rität können die komplexen Anforderungen logistischer Systeme bewältigt werden. Einen geeigneten Rahmen hierfür bieten Projekte des BMBF/BMWi oder der DFG zusammen mit der Industrie.

6 LITERATUR

- [1] Algorithms for Robust and online Railway optimization: Improving the Validity and reliability of Large scale systems, EU Projekt ARRIVAL, <http://arrival.cti.gr/>, online: 2. 7. 2008.
- [2] Baumgarten, H.; Walter, S. (2000): Trends und Strategien in der Logistik 2000+.
- [3] Becker, T. (2007): Ereignisorientiertes Prozessleistungsmanagement; In: 9/17/08 7:31:34 PM Ijioui, R., Emmerich, H. und Ceyy, M.: Supply Chain Event Management; Physica-Verlag Berlin Heidelberg 2007.
- [4] von Bögel, G.: Technologische Trends bei RFID-Systemen für den Einsatz im Internet der Dinge. In: Bullinger, H.-J. and ten Hompel, M. (2007): Internet der Dinge, Springer, Berlin 2007.
- [5] Bullinger, H.-J. and ten Hompel, M. (2007): Internet der Dinge, Springer, Berlin 2007.
- [6] Bundesvereinigung Logistik e. V. (Hg.) (2005). Wachstum schaffen – Zukunft gestalten, Logistik als Motor für Wachstum und Innovationen in Deutschland; Thesen und Handlungsempfehlungen.
- [7] Chen, F., Drezner, Z., Ryan, J. K. and Simchi-Levi, D.: Quantifying the bullwhip effect in a simple supply chain: The impact of forecasting, lead times, and information. *Management Science*, 46(3) pp. 436–443, 2000.
- [8] Gericke, E. (2004): Klassenbester in der Marktversorgung im Bereich Pneumatik. In: Schenk, M. (Hrsg.): Tagungsband Gastvortragsreihe Logistik 2004 – Logistik als Arbeitsfeld der Zukunft – Potenziale, Umsetzungsstrategien und Visionen. Fraunhofer IFF, 2004.
- [9] Graf, H. u. Putzlocher, S. (2004): DaimlerChrysler: Integrierte Beschaffungsnetzwerke. In: Corsten, D. u. Gabriel, C.: Supply Chain Management erfolgreich umsetzen – Grundlagen, Realisierung und Fallstudien. Springer. 2004.

- [10] Großer, T. (2006): Die Strippenzieher. In: Antrecht, R. (Hrsg.): McK Wissen 16 – Logistik. McKinsey.
- [11] Gudehus, T. (2007): Logistik 2. Netzwerke, Systeme, Lieferketten; Springer-Verlag Berlin Heidelberg.
- [12] Günther, H.-O. und Tempelmeier, H. (2005): Produktion und Logistik; 6. Auflage; Springer-Verlag Berlin Heidelberg.
- [13] ILOG (2007): Improving Plant Performance and Flexibility in Process Manufacturing With an Example from the Food and Beverage Industry, White Paper.
- [14] Kohl, N., Larsen, A., Larsen, J., Ross, A., and Tiourine, S. (2004): Airline Disruption Management – Perspectives, Experiences and Outlook, Technical Report, Carmen Systems, 2004. http://www.carmen.se/research_development/articles/crtr0407.pdf, online: 2. 7. 2007.
- [15] Kühlwein, R.: Einsatz von SAP-Lösungen entlang der gesamten Automobil-Prozesskette, Präsentation Branchenforum Automobillogistik, Stuttgart 2006.
- [16] Münch, G. und Seltmann, B. (2001): Produktionslogistik als Maßanzug für die Pharmaproduktion. In: LOGISTIK für Unternehmen 6/2001.
- [17] Nienhaus, J.: What is the Bullwhip Effect caused by? Supply Chain World Europe, Amsterdam 2002. http://www.beergame.lim.ethz.ch/Bullwhip_Effect.pdf.
- [18] Schenk, M., Tolujew, J.; Barfus, K., Reggelin, T. (2007): Grundkonzepte zu logistischen Echtzeitsystemen: Monitoring, Event Management und Frühwarnung. In: Jahrbuch Logistik 2007.
- [19] Ältere Lager- und Fertigungsanlagen mit Software zur Produktionslogistik aufrüsten. In: LOGISTIK für Unternehmen 10/2003.

> OPTIMIERUNG VON KOMMUNIKATIONSNETZEN

JÖRG EBERSPÄCHER, MORITZ KIESE UND ROLAND WESSÄLY

1 EXECUTIVE SUMMARY

Die weltweiten Telekommunikationsnetze (Telefon-, Mobilfunknetze, Internet) bilden heute das Nervensystem von Wirtschaft und Gesellschaft. Durch den anhaltenden technologischen Fortschritt in der Elektronik, der Computer- und Kommunikationstechnik steigt nicht nur die Zahl der Teilnehmer in den Netzen ständig, sondern es entstehen auch laufend neue Anwendungen, wie Radio und Fernsehen im Internet, elektronischer Handel und interaktive Spiele im Internet. Auf den Kupfer- und Glasfaserkabeln sowie den Funkstrecken werden gigantische Informationsmengen transportiert. Das Verkehrswachstum ist nach wie vor exponentiell.

Aufgebaut und betrieben werden diese Netze heute weltweit fast durchweg von kommerziellen privaten, miteinander konkurrierenden Netzbetreibern und Diensteanbietern. Sie müssen wegen des scharfen Wettbewerbs und der steigenden Kundenanforderungen ihre Netze permanent auf den neuesten Stand bringen und so effizient wie möglich betreiben. Eine systematische Planung ist hierbei unverzichtbar, insbesondere weil Prognosen über Verkehrswachstum, künftige Anwendungen und vor allem die Technologieentwicklung mit großen Unsicherheiten verbunden sind. Standorte für Netzknoten, Kabeltrassen und Antennen müssen festgelegt, Übertragungsstrecken dimensioniert und alles möglichst gut gegen Ausfälle abgesichert werden. Verkehrsströme unterschiedlichster Art (Sprache, Daten, Multimedia) sind günstig zu lenken (Routing), und im Fehlerfall sollten die Kunden möglichst keine Beeinträchtigung ihres Dienstes erleiden (Quality-of-Service, Ausfallsicherheit). Dominiert werden diese Anforderungen durch das Gebot der Kosteneffizienz. Entwurfs- und Planungsverfahren zur Unterstützung dieser Prozesse müssen sich also vor allem durch Robustheit, Flexibilität und Skalierbarkeit auszeichnen und möglichst kostenoptimale Lösungen produzieren. Ein Problemfeld, wie geschaffen für den Einsatz anspruchsvoller mathematischer Methoden! Es zeigt sich, dass hier an vielen Stellen noch erheblicher Handlungsbedarf für Forscher, Entwickler und Anwender besteht.

Dieser Beitrag gibt einen Überblick über den Stand des Einsatzes mathematischer Optimierung in der Planung von Telekommunikationsnetzen und identifiziert eine Reihe zukünftiger Herausforderungen. Ausgehend von der Technologie und Architektur der

Netze erläutert der Beitrag die Teilaufgaben der strategischen, der taktischen und der operativen Planung, die sich vor allem im Zeithorizont unterscheiden, über den sie sich erstrecken. Methodisch im Vordergrund stehen diskrete Optimierungsverfahren, wobei eine wesentliche Herausforderung die Größe der Probleme (Zahl der Netzknoten, Verbindungswege etc.) darstellt. Die Komplexität wird dadurch weiter gesteigert, dass Nebenbedingungen für Ausfallsicherheit und Dienstqualität mit berücksichtigt werden müssen.

Zukünftig ist zu erwarten, dass vor allem das Anwachsen der Netzgröße, die Berücksichtigung neuartiger Netztechnologien und das „Leben mit der Unschärfe der Eingangsparameter“ im Vordergrund des Interesses stehen werden. Zur Lösung dieser Problemstellungen sollten Mathematiker und Netzingenieure (Technologen) enger als bisher zusammenarbeiten. Hier ist auch die Industrie (Betreiber, Geräte-, Toolhersteller) aufgerufen, sich in größerem Umfang moderner Methoden zu bedienen. Als Vehikel zur Beschleunigung der Fortschritte auf dem Gebiet der Planung von Telekommunikationsnetzen wird in diesem Artikel neben der stärkeren Betonung der einschlägigen theoretischen Methoden in der Ingenieurausbildung ein konkretes wissenschaftlich-technisches Verbundforschungsvorhaben „Modellierung, Planung und Bewertung von Kommunikationsnetzen“ empfohlen, bei dem Mathematiker, Kommunikationsingenieure, Ökonomen und industrielle Anwender gemeinsam die „Grand Challenges“ der Netzplanung angehen. Querbezüge zu anderen Planungsaufgaben in der Logistik oder im Verkehr (Bahn, Straße, Flug) sind dabei vorhanden.

2 OB TELEFONNETZ ODER INTERNET: MATHEMATIK IST IMMER DABEI!

Der Einsatz von mathematischen Methoden in der Telekommunikation hat eine lange Historie. So geht die Planung von Telefonnetzen zurück auf die bahnbrechenden Arbeiten des dänischen Mathematikers A. K. Erlang (1878–1929) zur wahrscheinlichkeitstheoretischen Analyse von Leitungsbündeln. Seine Ergebnisse werden bis heute für die Planung von Telefon-, Daten- und Mobilfunknetzen verwendet.

In ähnlicher Weise profitiert das Internet von Methoden der angewandten Graphentheorie und Stochastik. Mathematische Optimierungsmethoden haben als zentrales Werkzeug Eingang in die Netzplanung gefunden.

2.1 ROUTING IM INTERNET

Das weltweite Internet ist ein riesiges Netz, welches aus mehr als 27 000 einzelnen Netzen, den so genannten *Autonomen Systemen*, zusammengesetzt ist. Wie finden die unzählbar vielen täglich gesendeten Datenpakete ihren Weg? Wie wird der Weg bestimmt und von wem? In aller Regel wird ein kürzester Weg von der Quelle zum Ziel gewählt – was nicht immer gleichbedeutend mit einem schnellsten oder günstigsten Weg ist. Festgelegt wird der Weg nicht durch eine zentrale Instanz, sondern ergibt

sich durch Kombination der Informationen der einzelnen Vermittlungsknoten entlang des sich ergebenden Weges (Pfades). Abb. 1 illustriert das weltumspannende Internet.

Für jede Art von Datenaustausch, wie E-Mail, Videokonferenz, Musikdownload, existieren unzählige Wegemöglichkeiten zwischen Quelle und Ziel. Trotzdem müssen die Netzknoten in Millisekunden entscheiden, ob und wie sie ein Datenpaket weitersenden. Bei diesem Vorgang der *Verkehrslenkung* (meist als *Routing* bezeichnet) wird im einfachsten Fall ein Weg ausgewählt, dessen „virtuelle Kosten“ im mathematischen Sin-

Abb. 1: Routen im Internet: Die Farbe zeigt die Top-Level-Domain (z. B. rosa für .de, .uk, .it, .pl und .fr), die Länge der Kanten repräsentiert die Paketlaufzeit [12]



ne minimal sind. Dabei gehen Parameter wie die Paketlaufzeit, die Zahl durchlaufener Netzknoten oder auch reale monetäre Kosten in die virtuellen Kosten ein. Die Router berechnen – jeder für sich – zu allen im Subnetz erreichbaren Zielen einen kürzesten Pfad. Dies geschieht im Internet in jeder Sekunde millionenfach unter Verwendung von Algorithmen aus der Graphentheorie. Zumeist kommt dabei eine Variante des in den fünfziger Jahren entwickelten Algorithmus von Dijkstra [8] zum Einsatz, einer der in praktischen Anwendungen am häufigsten verwendeten Algorithmen überhaupt.

2.2 QUALITÄTSKONTROLLE IM TELEFONNETZ UND IM INTERNET

Telefonieren heißt: Verbindungen schalten zwischen beliebigen Teilnehmern in der ganzen Welt, zu jeder Zeit. Für die Netzbetreiber ist das insofern eine große Herausforderung, als die Verbindungswünsche der Teilnehmer mehr oder weniger zufällig sind, die Kommunikationsaktivität über den Tag bzw. die Woche hinweg stark schwankt und extreme Lastspitzen auftreten. Man denke nur an die Silvesternacht oder an besondere Ereignisse wie Katastrophen oder Fernsehsendungen mit interaktiven Zuschauerabstimmungen. Die Verbindungen benötigen auf den Übertragungstrecken im Netz und in den Netzknoten Ressourcen, z. B. Leitungskapazitäten, die vorab geplant werden müssen. Diese *Netzdimensionierung* ist seit den Anfängen der Telefonie eine der Schlüsselaufgaben der Netzbetreiber. Werden die Kapazitäten zu gering ausgelegt, gibt es zu viele „Belegfälle“, in der Telekommunikation auch *Blockierungen* genannt. Wird überdimensioniert, bezahlt der Betreiber zu viel für sein Leitungsnetz, was beim heutigen wirtschaftlichen Druck auf die Betreiber besonders unerwünscht ist. Schon in den Frühzeiten der Telefonie entwickelten Mathematiker, allen voran der oben genannte A. K. Erlang, eine mathematische Theorie des Fernsprechverkehrs, mit der auf Basis gewisser Annahmen über den statistischen Charakter der Verkehrsanforderungen die zu erwartende Blockierungswahrscheinlichkeit im Netz bzw. auf Leitungen abgeschätzt werden kann. Die Betreiber können damit ihre Leitungen verkehrsgerecht dimensionieren. Dazu müssen nur die mittlere Anrufrate der Nutzer und die mittlere Rufdauer gegeben sein – zwei Werte, die sich hinreichend gut aus Messungen bestimmen lassen.

In Datennetzen gewinnt eine andere Größe an Bedeutung: die Verzögerungszeit eines Datenpakets bis zum Erreichen seines Ziels. Entlang des Pfades entstehen Verzögerungen, vor allem, weil in den Routern Zwischenspeicher vorhanden sind, um kurzfristige Überlasten abzpuffern. In diesen Puffern entstehen Wartezeiten, ähnlich wie im Straßenverkehr bei Staus, und wegen der endlichen Größe dieser Pufferspeicher, je nach Verkehrslast auch Paketverluste. Die mathematische Theorie der *Warteschlangensysteme* (*Queueing Theory*) ist eine wesentliche Grundlage für quantitative Aussagen über die Qualität in modernen Kommunikationsnetzen. Der Pionier dieser Theorie war der US-Wissenschaftler Leonard Kleinrock [13], einer der Väter des Internets. Die Warteschlangentheorie stellt heute ein ganzes Gebäude dar, das nicht mehr nur in Telekommunikationsnetzen angewendet wird. Wegen des rasanten Fortschritts der technischen

Infrastrukturen und neuen Anwendungen wird sie laufend weiterentwickelt und bildet eine wichtige Voraussetzung für alle modernen Planungsverfahren in Netzen.

2.3 KOSTENOPTIMIERTE KOMMUNIKATIONSNETZE

Zur Planung der Infrastruktur eines Kommunikationsnetzes sind unzählige, sich gegenseitig beeinflussende Entscheidungen zu treffen. So müssen in der strategischen Planung die Standorte des Netzes festgelegt werden, die Netzstruktur bestehend aus Zugangs-, Aggregations- und Kernnetz geplant werden, es müssen Netzelemente für viele Technologien an ausgewählten Standorten platziert werden, es müssen Verbindungen dimensioniert werden und Dienste über diese Netze so geroutet werden, dass auch bei Ausfall einzelner Netzkomponenten eine hohe Verfügbarkeit zugesichert werden kann. Der Aufbau und Betrieb einer nationalen Infrastruktur kostet dabei schnell viele Milliarden Euro. Mit den Methoden der ganzzahligen, linearen Optimierung hat die Mathematik bei der Infrastrukturplanung viele Erfolge erzielt. Bei der X-WiN Planung des Deutschen Forschungsnetzes (siehe auch Abb. 4 im Abschnitt 3.2) konnten im Jahre 2005 Aufgaben mit Trillionen an Variablen beweisbar kostenoptimal gelöst werden. Basierend auf den so berechneten Netzen sind heute nahezu alle deutschen Forscher an ein außerordentlich kostengünstiges Netz angebunden, welches ausfallsicher ist und sehr hohe Anbindungskapazitäten bereitstellt. Detailliertere Aussagen zu Kosteneinsparungen sind von den meisten Betreibern kaum zu erhalten. Daher ist hier noch die Planung von Ausfallsicherheit für das Netz des amerikanischen Betreibers AT&T erwähnenswert. Nach eigener Aussage hat AT&T im Jahr 2000 mit derartigen Verfahren „hundert Millionen Dollar“ eingespart [11].

3 NETZE, PLANUNG UND METHODEN

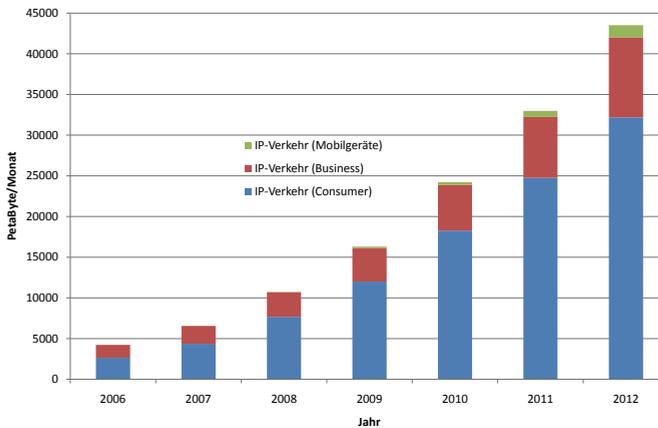
Die private und geschäftliche Nutzung der weltweiten Kommunikationsnetze nimmt stetig zu; das Verkehrsaufkommen im Internet verdoppelt sich etwa alle 18 Monate (siehe dazu auch Abb. 2). Die damit verbundene gesamtwirtschaftliche Bedeutung der Kommunikationstechnologie zeigt sich auch am Umsatz von 60 Milliarden Euro, den allein die Deutsche Telekom AG im Geschäftsjahr 2007 erwirtschaftet hat.

Aufgrund der hohen Investitionskosten und des intensiven weltweiten Wettbewerbs, ist es für alle Netzbetreiber von vitalem Interesse, Dienste möglichst kostengünstig zu realisieren. Eine sorgfältige Planung mit fundierten Methoden kann dazu einen wichtigen Beitrag liefern.

3.1 NETZARCHITEKTUR UND PLANUNGSASPEKTE

Heutige Kommunikationsnetze sind geprägt durch den Einsatz vieler verschiedener Netztechnologien, die ein Netz sowohl horizontal als auch vertikal gliedern.

Abb. 2: Verkehrswachstum im Internet (Quelle: Cisco [10])

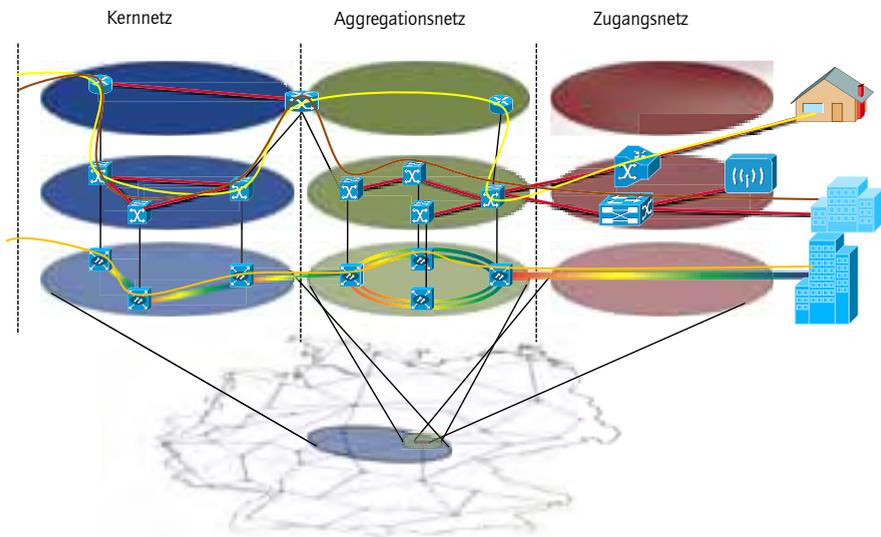


Horizontal wird das Gesamtnetz ausgehend von seiner geographischen Ausbreitung und seiner Funktionalität in *Zugangs-*, *Aggregations-* und in *Kernetze* untergliedert. In Aggregationsnetzen wird Verkehr von Privat- oder Geschäftskunden aus dem Zugangsnetz gesammelt und in das Kernnetz weitergeleitet. Das Aggregationsnetz dient als Zubringer zu den großen nationalen wie internationalen „Datenautobahnen“ des Kernnetzes, die die weltweite Konnektivität herstellen.

Vertikal ist ein Kommunikationsnetz in Schichten organisiert, die mit unterschiedlichen Technologien und Protokollen realisiert werden. Eine Analogie dazu aus dem täglichen Leben ist der Transport von Gütern. Das zu versendende Gut kommt in einen Karton, der Karton in einen Container, viele Container in einen LKW und der LKW transportiert das Gut zwischen den großen Sortierstandorten über die Infrastruktur bestehend aus Straßen und Autobahnen. In einem modernen Kommunikationsnetz werden Dienste wie Sprache, Daten und Video über verschiedene Technologien transportiert. Dazu gehören IP, Ethernet, ATM, SDH und WDM, deren technische Unterschiede hier nicht näher erläutert werden können. Diese Technologien nutzen Glasfasern, Kupferleitungen und Richtfunk als physikalische Infrastruktur.

Bei der Planung von Kommunikationsnetzen geht es um Aufgaben wie Architektur-, Standort- und Technologieauswahl, Technikplatzierung, Topologieplanung, Dimensionierung von Netzelementen und Verbindungen, und nicht zuletzt um Verkehrsprognose und Verkehrslenkung. Ziel ist es, ein qualitativ hochwertiges, robustes, skalierbares und kostengünstiges Netz aufzubauen und zu betreiben. Welche Themen und Ziele dabei im

Abb. 3: Horizontaler und vertikaler Netzaufbau



Vordergrund stehen, hängt davon ab, was im Detail mit welchem Zeithorizont geplant wird. Dabei kann man drei wichtige Bereiche zu unterscheiden:

- Strategische Planung (Infrastruktur und Netzarchitektur)
- Taktische Planung (Netzentwicklung)
- Operative Planung (Netzkonfiguration und Netzbetrieb)

In den folgenden Abschnitten werden diese Bereiche näher erläutert und der Stand der mathematischen Forschung bzw. der Grad der Anwendung in der Praxis dargestellt.

3.2 STRATEGISCHE PLANUNG

Die Gesamtnetzarchitektur ergibt sich aus dem ausgewählten Technologiemix zum horizontalen und vertikalen Aufbau des Netzes. In der strategischen Planung sind zur Vorbereitung der Architekturentscheidung verschiedene Alternativen zu bewerten und unter Berücksichtigung des Wechselspiels von technischen und ökonomischen Nebenbedingungen zu vergleichen. Dabei müssen sowohl Planungen für die Netzschichten als auch für die einzelnen Netze durchgeführt werden.

Die Bestimmung der Kernnetzstandorte mit der zugehörigen Aggregationsstruktur ist eine der wichtigsten, langfristig wirkenden Entscheidungen eines Netzbetreibers. Ausgehend von Verkehrsprognosen, den technologischen Möglichkeiten der potentiellen Netzhardware und zugehörigen Kostenstrukturen sind viele eng miteinander verknüpfte Fragen zu beantworten:

- Welche Kernnetzstandorte sind auszuwählen?
- Welche Kernnetzstandorte sollen Übergabepunkte zu anderen Netzbetreibern und für den internationalen Verkehr sein?
- Welche Aggregationsstandorte sind auszuwählen?
- Wie werden die Zugangs- und Aggregationsbereiche zugeschnitten?
- Welche Hardware soll in welcher Ausbaustufe zum Einsatz kommen?
- Welche Leitungskapazitäten sollen eingesetzt werden?
- Wie soll der Ausfallschutz realisiert werden?

Für solche Fragestellungen wurde eine Reihe von mathematischen Modellen entwickelt. Im Kern stecken oft *Facility Location Probleme*, die als ganzzahlige lineare Optimierungsprobleme (*Integer Linear Programming* – ILP) formuliert sind. Siehe dazu auch in Abb. 4 eine Modellvariante, die in ähnlicher Weise bei den Strukturplanungen des DFN-Vereins für die Deutschen Forschungsnetze G-WiN [6] und X-WiN [7] eingesetzt wurde. Zur Lösung solcher Aufgabenstellungen bieten sich sowohl Lagrange- als auch Branch-and-Cut-Verfahren an. Netze mit bis zu 1000 Standorten können mit solchen Methoden heute optimiert werden. Grundlegend dafür sind die Fortschritte bei der kompakten Beschreibung des Lösungsraums durch Polyeder.

Für die kommenden Jahre wird erwartet, dass die Netzbetreiber strategisch in den Ausbau von glasfaserbasierten Anschlussnetzen investieren. Für eine deutschlandweite Planung bedeutet dies eine Netzgröße von bis zu 40 Millionen Verkehrsquellen – was in etwa der Anzahl der Haushalte in Deutschland entspricht. Derartige Problemgrößen sind heute noch nicht lösbar – selbst für die trickreichsten mathematischen Verfahren und bei Einsatz von extrem leistungsfähigen Großrechnern. Hier steht die mathematische Forschung vor gewaltigen Herausforderungen.

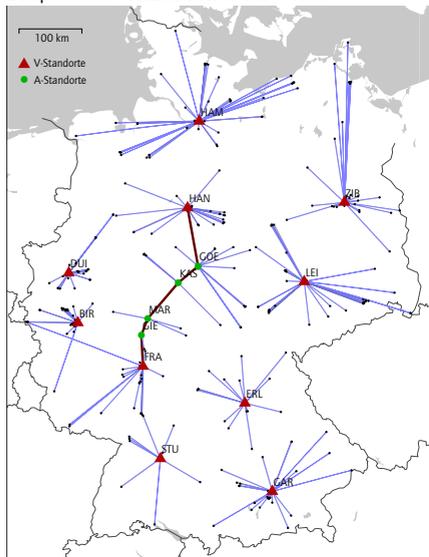
3.3 TAKTISCHE PLANUNG

In Abgrenzung zur strategischen Planung werden hier weniger weitreichende Entscheidungen getroffen. Basierend auf einer festen Netzarchitektur- und -hierarchie sowie einer zumindest teilweise existierenden Infrastruktur wird in der taktischen Planung die mittelfristige (ca. ein Budgetjahr) Netzentwicklung organisiert. Klassische Aufgaben der taktischen Planung sind:

Abb. 4: G-WiN Optimierungsmodell

Das Deutsche Forschungsnetz verbindet deutschlandweit die mehr als 700 wissenschaftlichen Institute und Universitäten. Zur mathematischen Abbildung der strategischen Planung von Aggregations- und Kernnetz für die Ausbaustufen G-WiN (1998) [6] und X-WiN (2004) [7] des Deutsche Forschungsnetzes wurden Varianten des *Facility Location Problems* verwendet. Die folgende entspricht in etwa der Formulierung für das X-WiN.

Beispiel einer Netzhierarchie des X-WiN:



Aufgabe: Aus der Menge aller Standorte N mit fester Teilmenge V von Kernnetzstandorten war eine Teilmenge der potentiellen Aggregationsstandorte A auszuwählen. Jeder Standort war mit ausreichender Bandbreite über eine Verbindung aus der Menge L an einen Aggregationsstandort anzubinden und dieser selbst war in einer ausfallsicheren Kette p von A

Standorten mit ausreichend Bandbreite an die Kernnetzstandorte anzubinden.

Variablen: $y_a \in \{0, 1\}$, $y_a = 1$ genau dann, wenn a zum A-Standort wird; $x_{ij} \in \{0, 1\}$, $x_{ij} = 1$ genau dann, wenn i ein Anwenderstandort ist oder wird und i an den A- oder V-Standort j angebunden wird; $z_p \in \{0, 1\}$, $z_p = 1$ genau dann, wenn a_1, \dots, a_m zu A-Standorten werden und diese über die Kette $p = (v_1, a_1, \dots, a_m, v_2)$ an die V-Standorte v_1, v_2 angebunden werden.

In einem Netz mit mehr als 700 Standorten, 100 potentiellen Aggregations- und Kernnetzstandorten ergeben sich über eine Trillion Variablen!

Mathematisches Modell: Die zugehörige Formulierung als ganzzahliges lineares Programm ist:

$$\begin{aligned} \min & \\ & \sum_{a \in A} k_a^A y_a + \sum_{p \in P} k_p^P z_p + \sum_{ij \in L} k_{ij}^L x_{ij} \\ & \sum_{ij \in L} x_{ij} = 1 \quad \forall i \in N \\ & \sum_{aj \in L} x_{aj} = 1 - y_a \quad \forall a \in A \\ & x_{ia} \leq y_a \quad \forall a \in A, ia \in L \\ & \sum_{p \in P, a \in p} z_p = y_a \quad \forall a \in A \\ & + \text{ausreichende Bandbreiten} \\ & + \text{Ausfallsicherheit} \end{aligned}$$

Nach Anwendung von komplexen Reduktions- und Aggregationstechniken verbleiben gemischt-ganzzahlige Programme mit *nur noch* 50 000 bis 100 000 Variablen und Ungleichungen. Diese wurden mit Schnittebenenverfahren und einem Branch-and-Cut Algorithmus optimal gelöst.

Ergebnis: Laut DFN-Verein, konnten „nur durch den Einsatz dieser mathematisch fundierten Lösungsansätze und weitgehend automatisierter Verfahren eine hohe Planungsqualität und -sicherheit gewährleistet werden und die vorhandenen Optimierungspotentiale voll ausgeschöpft werden“.

- Erweiterung der physikalischen Netztopologie
- Einführung neuer Netzelemente
- Ausbauplanung für Leitungskapazitäten

In dieser Planungsphase wird angestrebt, mit geringen Investitionen die mittelfristig prognostizierte Verkehrssteigerung aufzufangen. Die Prognosen sind für etablierte Dienste einigermaßen verlässlich. Gleiches gilt für die Kosten von Ausbaualternativen, wobei sich diese deutlich in den Kosten unterscheiden. So sind in urbanen Gebieten neue Trassen für den Faserausbau deutlich kostenintensiver als z. B. die einfache Erhöhung von Übertragungskapazitäten (bei vorhandenen Fasern) durch Aufrüsten der Netzknoten.

Das existierende Netz ist so auszubauen, dass die zukünftigen Verkehrsanforderungen mit den vorhanden Ressourcen geroutet werden können. Dabei sind Wege zum Routen des Verkehrs zu wählen, die den Qualitätsanforderungen der Netzbetreiber genügen: die Länge von Wegen in optischen Netzen darf die physikalischen Grenzen der Übertragung von Licht nicht überschreiten, die Wege in IP-Netzen dürfen zur Einhaltung von Paketlaufzeiten nicht zu viele Netzelemente passieren, etc. Dabei ist darauf zu achten, dass Ausfälle von Netzkomponenten nicht zu nennenswertem Verlust von Verkehr und damit unter Umständen zu Konventionalstrafen führen! Gerade in modernen Weitverkehrsnetzen ist dieses Thema aufgrund der enormen Leitungskapazitäten von besonderer Brisanz [9].

Aus mathematischer Sicht lassen sich diese Aufgaben der taktischen Planung mit Varianten von Mehrgüterflussproblemen mit diskreten Leitungs- und Knotenkapazitäten modellieren [14]. Durch *Spaltengenerierung* sind moderne Lösungsverfahren in der Lage, mit vielen Millionen von Wegealternativen umzugehen. Auch Bedingungen an die Ausfallsicherheit lassen sich gut abbilden. Gleiches gilt für Bedingungen, die sich aus den technologischen Randbedingungen der zu verwendenden Netzelemente ergeben.

Entsprechend der vertikalen Strukturierung sind die Netze über mehrere Schichten hinweg zu planen. Dabei werden Verbindungen der höheren Netzschichten über (ausfallsichere) Wege der darunter liegenden Netzschicht realisiert. Die integrierte Berücksichtigung von mehreren Schichten führt zu sogenannten *Multi-Layer-Problemen*. Aus mathematischer Sicht sind diese Aufgaben extrem komplex. Während die Planung von einzelnen Schichten für praktisch relevante Netzgrößen von bis zu 50 Standorten gut gelöst werden kann, sind Multi-Layer-Probleme heute selbst bei Vernachlässigung von Ausfallsicherheitsanforderungen nur für sehr kleine Netze von etwa zehn Standorten mit akzeptablen Qualitätsgarantien lösbar. Multi-Layer-Planung und andere Themen der Integration mehrerer Planungsaspekte gehören zu den zukünftigen Herausforderungen auf die in Abschnitt 4.6 näher eingegangen wird.

3.4 OPERATIVE PLANUNG

Die kurzfristigen Aspekte der Netzkonfiguration und des Netzausbaus sind Teil der operativen Planung. Neben der Konfiguration von Hardware ist dabei die Planung und Optimierung der Verkehrslenkung in allen Netzschichten von zentraler Bedeutung. Dazu existieren in heutigen Netzen zwei Ansätze: *verbindungsorientiertes* und *verbindungsloses* Routing.

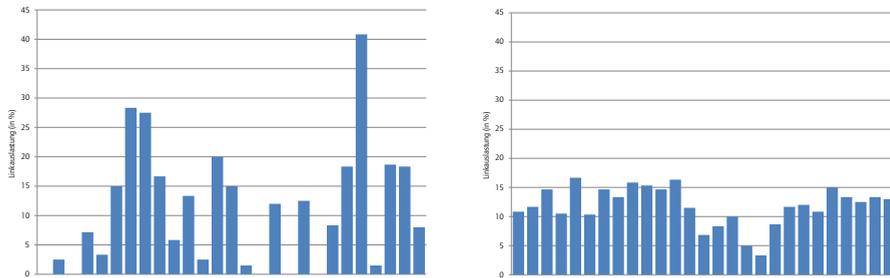
Bei verbindungsorientiertem Routing wird von den Netzelementen keine eigene „Intelligenz“ zur Wegefindung bereitgestellt, d. h. der Weg zu einem Ziel muss explizit geplant und konfiguriert werden. Diesen Ansatz findet man vorrangig in den unteren Netzschichten, d. h. in SDH-Netzen sowie in WDM-Netzen, in denen eine Wegefindung technisch einfach (noch) nicht möglich ist. Um einen so genannten *Lichtpfad*, d. h. einen Lichtstrahl, der unverändert durch mehrere Knoten geleitet wird, aufzubauen, müssen noch zahlreiche Einstellungen auf sehr physiknahen Schichten vorgenommen werden. Erst in der neuesten Gerätegeneration ist es in *Optical Cross Connects* oder *Reconfigurable Add Drop Multiplexern* gelungen, diese Art von Konfiguration „fernsteuerbar“ zu machen – von einer automatischen Wegesuche ganz zu schweigen.

Beim verbindungslosen Routing stellen Netzelemente ein autonomes Verfahren zur Wegefindung bereit. Auf der Basis von „virtuellen Distanzen“ werden beispielsweise Pakete durch die IP-Netze geroutet. Bei Technologien, die auf den Einsatz in Kernnetzen zugeschnitten sind, existieren aber auch Mischformen. So wird etwa bei MPLS im Normalfall entlang eines kürzesten Pfades geroutet, der Netzbetreiber kann aber auch explizit Routen vorgeben, die davon abweichen.

Mit den Methoden der ganzzahligen Optimierung ist die moderne Mathematik heutzutage für beide Routingansätze in der Lage, realistische Problemstellungen zu lösen. Beim verbindungsorientierten Routing haben sich Varianten von Mehrgüterflussproblemen etabliert. Mit ganzzahligen Programmen und Spaltengenerierungsverfahren lassen sich diese Aufgaben auch bei komplexen Nebenbedingungen zur Ausfallsicherheit und zu operativen Anforderungen gut lösen. Dies gilt für Netze mit mehreren hundert Netzelementen.

Auch im Falle der verbindungslosen Kommunikation können heutige Verfahren erhebliche Verbesserungen gegenüber vielfach verwendeten Standardkonfigurationen, die z. B. beim IP-Routing üblich sind, erzielt werden. Reichen die Werkseinstellungen vielleicht noch aus, solange alle Netzelemente wie vorgesehen funktionieren, so kann spätestens bei der Betrachtung von Ausfällen eine vorherige Optimierung der Verbindungsauslastung wertvolle Ressourcen sparen. Damit können Hardware-Upgrades vermieden werden – und das ohne die Nutzung spezieller Protokolle, sondern nur durch Verändern der Kantengewichte [7]. Wie Abb. 5 zu entnehmen ist, kann durch eine solche Optimierung eine deutlich bessere Lastverteilung erreicht werden. Das linke Bild zeigt die Lastverteilung auf den Verbindungen unter Verwendung von Standardgewichten und das rechte Bild die optimierte Variante. Auch ohne die Details zu kennen, lässt sich der

Abb. 5: Lastverteilung mit Standardgewichten (links) und optimierten Gewichten (rechts)



Vorteil der optimierten Lösung erahnen, da die maximale Last auf einer Verbindung – welche einer der entscheidenden Qualitätsparameter in IP/OSPF-Netzen ist – von 40,8 auf 17,0 Prozent gesenkt werden konnte. Da derartige Konfigurationen ohne Veränderungen an der Hardware auskommen, sind sie für zukünftigen Geschäftsmodelle, in denen Kunden hohe Bandbreiten für vergleichsweise kurze Zeiträume sehr schnell zur Verfügung gestellt werden sollen, von besonderem Interesse.

Für die zukünftigen Netzstrukturen, in denen alle Dienste über ein „All-IP-Netz“ transportiert werden, wird es bei der operativen Planung immer stärker darauf ankommen, die Stochastik der Dienste besser zu verstehen und in derartige Planungen – die heute auf statischen Verkehrsmessungen basieren – mit einzubeziehen.

3.5 STAND DER METHODEN: FAZIT

Die mathematischen Methoden zur Lösung von Planungs- und Optimierungsaufgaben der Telekommunikation haben sich in den vergangenen zwei Jahrzehnten gut entwickelt. Dies gilt insbesondere für die Methoden der ganzzahligen linearen Optimierung. Für viele Varianten von Topologie-, Routing und Dimensionierungsaufgaben auch unter Berücksichtigung von Ausfallsicherheitsanforderungen wurden leistungsfähige Modelle und Spezialverfahren entwickelt. Heute können Aufgaben mit vielen Millionen Routen bzw. Variablen durch Kombination von dynamischer Variablengenerierung, polyedrischen Methoden, Separierungsalgorithmen und Lagrangeverfahren gelöst werden. Trotz dieser offensichtlichen Fortschritte gibt es noch eine Fülle von Herausforderungen, auf die im nächsten Abschnitt näher eingegangen wird.

4 EINFLUSSFAKTOREN UND HERAUSFORDERUNGEN

4.1 MARKTEINFLÜSSE, WETTBEWERBSBEDINGUNGEN, REGULIERUNG

Die Telekommunikationsmärkte wurden in den 90er Jahren nahezu weltweit entmonopolisiert. So wurde das Monopol der Deutschen Telekom für Telefonie 1998 aufgehoben. Seitdem hat sich – nicht nur in Deutschland – ein dynamischer Telekommunikationsmarkt entwickelt. Der Preis- und Margendruck der Netzbetreiber nahm enorm zu und wird sich in den kommenden Jahren noch weiter verstärken. Heute existiert zunehmend ein Marktumfeld, in dem Betreiber von Mobilfunk-, Kabel-, klassischen Telefon- und Datennetzen sowie Diensteanbieter und Content-Provider mit einem breiten Produktspektrum (Telefonie, Datendienste, Fernsehen) um die Gunst der Kunden werben.

Es ist daher zu erwarten, dass das Interesse an kostengünstigen Lösungen für derartig integrierte *Multi-Service-Netze* weiter wächst und sich damit die Rahmenbedingungen für den erfolgreichen Einsatz von mathematischen Optimierungsmethoden weiter verbessern. Auch im Vorfeld von Regulierungsentscheidungen könnten technologische Analysen zu einer präziseren Markteinschätzung und Technologiebewertung führen.

4.2 DATENVERFÜGBARKEIT

Eine aussagekräftige Planung braucht eine verlässliche Datenbasis. Netzinformationen, Daten über die Hardware für die Netzknoten- und Verbindungsstrecken (Links), Kosten und Verkehrsprognosen sind die wesentlichen Eingangsdaten für die Netzplanung. Verfügbarkeit und Detaillierungsgrad dieser Informationen schwanken allerdings sehr stark. Am besten ist die Datenlage bei der Knoten- und Linkhardware. Die Systemhersteller stellen dazu wichtige Informationen, wie zum Beispiel die Bitrate einer Verbindung oder die Schaltkapazität eines Netzknotens zur Verfügung. Die Netzbetreiber pflegen zusätzlich ausführliche Netzinformationen in ihren Datenbanken: Welche Standorte hat das Netz und welche Hardware steht an welchem Standort? Mit welchen Medien (Kupfer- und Glasfaserkabel, Richtfunk) sind die Netzelemente miteinander verbunden und wie werden die Dienste über das Netz geroutet?

Ungünstiger stellt sich die Situation bei den Kosten und den Verkehrsdaten dar. Preislisten der Systemhersteller sind wegen unterschiedlicher oder nicht transparenter Rabattpolitik wenig vergleichbar. Außerdem spielen neben den Investitionskosten auch die Betriebskosten eine entscheidende Rolle zur Bewertung der *Total Cost of Ownership* (TCO). Mittlerweile existieren zwar erste Modellierungsansätze für Betriebskosten, jedoch sind die Zusammenhänge nicht ausreichend verstanden, um adäquat in die Planung einbezogen zu werden. Verkehrsprognosen sind per Definition ungenau. Für die operative und damit kurzfristige Planung können die Vorteile einer standardisierten Technologie mit integrierten Mess- und Protokollsystemen ausgespielt werden. Aus Messwerten lassen sich einigermäßen plausibel die Verkehrsbeziehungen inklusive Bandbrei-

tenanforderung ableiten. Die Langfristprognosen bleiben jedoch ein Problem. Hier muss man typische Alternativszenarien untersuchen. Aus Sicht der Betreiber lässt sich zusammenfassend eine gute Datengrundlage attestieren, die geeignet ist, mit Hilfe von modernen Optimierungsverfahren zusätzliche Potenziale zu heben.

Für den wissenschaftlichen Fortschritt bei Netzplanung und -optimierung ist es wichtig, dass nicht nur wenige privilegierte Forschergruppen den Zugang zu realistischen Referenznetzdaten haben. Erste Versuche, solche Daten bereitzustellen, wurden mit der SNDlib [1], sowie in den BMBF-Projekten EIBONE [2] und 100GET [3] oder auch im EU-Projekt NOBEL [4] gemacht. Weitere Anstrengungen sind hier notwendig, um für die Vielzahl der Planungsherausforderungen eine adäquate Datenbasis herzustellen. Dies kann sehr lohnenswert sein, wie man an einem Beispiel aus dem Bereich der ganzzahligen linearen Programmierung sehen kann: Seit die Referenzdaten der MIPLIB [15] publik gemacht wurden, hat sich die Leistungsfähigkeit der wichtigsten Algorithmen so sehr gesteigert, dass typische Probleme heute bis zu einer Million (!) mal schneller gelöst werden [16]. Die verbesserte Rechnerhardware leistet dabei mit einem Faktor von 1000 ihren Beitrag. Ein weiterer Faktor 1000 kann den heute deutlich leistungsfähigeren mathematischen Verfahren zugerechnet werden.

4.3 KONVERGENZ UND KOMPLEXITÄT DER SYSTEME

In der Vergangenheit wurde für nahezu jeden Dienst eine eigene Infrastruktur aufgebaut – zumindest oberhalb der physikalischen Übertragungsschicht. Dies wird sich mit dem Aufbau der so genannten *Next Generation Networks* (NGN) grundlegend ändern. Die Netze der Zukunft wachsen auf Basis der Internet- und Ethernet-Technologien zusammen und setzen dabei auf einer gemeinsamen optischen WDM-Infrastruktur auf. Diese *Konvergenz der Netze* führt dazu, dass große nationale und transnationale Multi-Service-Netze entstehen: ein Netz für alle Dienste. Ob diese technologische Vereinfachung die Planungsaufgaben erleichtern wird, ist angesichts der Vielfalt der gleichzeitig zu transportierenden Dienste und deren Anforderungen zweifelhaft. Waren die getrennten Netze der Vergangenheit auch separat planbar, so geht es jetzt darum, eine einzige, extrem komplexe Aufgabe in geeigneter Weise zu dekomponieren. Dieser Wandel wird ohne massiven Einsatz neuartiger mathematischer Methoden nicht zu bewältigen sein.

4.4 PLANUNGSPROZESSE UND PLANUNGSWERKZEUGE

Die heute eingesetzten Planungswerkzeuge sind mächtige Softwaretools, mit denen große Teile der Netzplanung unterstützt werden. Aufsetzend auf den Datenbanken zur Netzdokumentation erlauben diese Tools eine computergestützte, teilweise automatisierte, aber letztendlich im Wesentlichen manuelle Planung. Diese wird gestützt durch eine gute visuelle Darstellung der technologischen Schichten des Netzes. Aus Sicht der heute schon verfügbaren Mathematik zur Planung und Optimierung eines Netzes sind

die verwendeten Algorithmen und Methoden eher als hausbacken zu bezeichnen. Momentan ist aber die Integration von neuen Planungsmethoden auf Grund proprietärer oder nicht existenter Schnittstellen sehr aufwändig. Die Toolhersteller haben wenig eigenes Interesse innovative Methoden zu integrieren und für die Netzbetreiber ist der hohe initiale Aufwand hinreichend abschreckend, um die Wirksamkeit neuer Verfahren in realen Planungsumgebungen erst gar nicht zu testen. Eine mittelfristige Perspektive ergibt sich aus dem sich abzeichnenden Wandel der Software-Infrastruktur der Firmen hin zur so genannten *Service-Oriented Architecture* (SOA). Dies könnte zu quasi-standardisierten Software-Plattformen führen, mit deren Hilfe man leichter als heute komplexe Planungstools aus Bausteinen zusammensetzen kann. Offenere Schnittstellen würden es Ingenieuren und Mathematikern erlauben, neuartige innovative Methoden zu etablieren, mit denen Betreiber auf dem Weg zu kosteneffizienteren Netzen mit besserer Ressourcennutzung unterstützt werden.

4.5 ZUSAMMENARBEIT VON INDUSTRIE UND WISSENSCHAFT, INGENIEUREN UND MATHEMATIKERN

Die Telekommunikationsindustrie in Deutschland arbeitet traditionell eng mit den Universitäten zusammen. Aus Sicht der Mathematik erstreckt sich dies bedauerlicherweise vorrangig auf Ingenieursdisziplinen. Die großen Systemhersteller wie Alcatel-Lucent, Ericsson und Nokia Siemens Networks sowie Netzbetreiber wie die Deutsche Telekom haben aber auch eigene Forschungsabteilungen. Im Rahmen nationaler und internationaler vorwettbewerblicher Projekte kooperieren diese Abteilungen in erheblichem Umfang, so z. B. in den nationalen BMBF-Programmen EIBONE [2], 100GET [3] sowie in den Projekten NOBEL [4] und Euro NGI [5] auf europäischer Ebene. Daneben gibt es zahlreiche bilaterale Projekte der Industrie mit ingenieurwissenschaftlichen Universitätsinstituten. Der Transfer von Know-How und von Mitarbeitern funktioniert hier sehr gut. Beteiligung von Seiten der Mathematik gab es bisher aber kaum. Dies ist nicht den Ingenieuren anzukreiden; die Mathematik muss sich hier deutlich stärker öffnen. Die Einzelprojekte im DFG-Forschungszentrum MATHEON, erste Beteiligungen des Zuse-Instituts-Berlin an oben genannten BMBF-Programmen und vereinzelte mathematische Spin-Offs sind zu wenig, denn Ingenieure sind für mathematische Methoden durchaus begeisterungsfähig. Mehr und mehr verwenden sie bekannte Optimierungsverfahren zur Lösung komplexer Planungsaufgaben, wobei dann etablierte und teilweise in Programmbibliotheken verfügbare mathematische Lösungsverfahren zum Einsatz kommen – eine ideale Situation! Die Mathematiker arbeiten an den theoretischen und praktischen Aspekten von Lösungsverfahren, was den Ingenieuren wiederum erlaubt, Lösungen mit neuen Qualitäten zu berechnen.

4.6 AUSGEWÄHLTE HERAUSFORDERUNGEN

Will die Mathematik stärker als bisher eigene Beiträge zur Weiterentwicklung der Telekommunikationsnetze liefern, muss sie sich den Kernfragestellungen stellen, die sich bei der Migration hin zu den Netzen der Zukunft ergeben.

HERAUSFORDERUNGEN FÜR NETZBETREIBER

Die zentralen Themen für die Zukunft sind

- die Bewältigung des kontinuierlichen Wandels der Netze,
- die Berücksichtigung von techno-ökonomischen Zusammenhängen
- und die effizientere Nutzung der vorhandenen Netzressourcen.

Bewältigung des Wandels: Die Multi-Service-Netze der kommenden Generationen werden enorme Datenmengen transportieren. Schon heute existieren Anschlussbitraten von bis zu 50-100 Mbit/s. In weniger als zehn Jahren werden diese auf 1 Gbit/s und im Geschäftskundenbereich häufig bis auf 10 Gbit/s steigen. Zur Bewältigung dieser Übertragungskapazitäten müssen massive Investitionen in neue Technologien getätigt werden. Gleichzeitig existiert eine riesige installierte Basis. Die heutigen Netze bestehen aus vielen Hunderttausend Netzknoten, die über unterschiedliche Technologien komplex miteinander verbunden sind. Diese Netze funktionieren heute recht gut! Die Migration zu neuen Netzarchitekturen und Diensten ergibt neue komplexe Fragestellungen: Wann sollen welche Dienste über die neuen Netze realisiert werden? Wie sehen die optimalen Strukturen dieser Netze aus? Welche Dynamik wird der Verkehr haben? Wie lassen sich die Kosten der Migration minimieren?

Techno-ökonomische Bewertung. Das Zusammenwachsen der Netze, auch im Hinblick auf die Konvergenz von Fest- und Mobilfunknetzen, macht es für Betreiber immer wichtiger, die komplexen Zusammenhänge zu verstehen, die die *Total Cost of Ownership* (TCO) bestimmen und damit maßgeblichen Einfluss auf die Umsatz- und Gewinnerwartungen haben. Einzubeziehen sind dabei diverse Technologie- und Hardwareoptionen, Verkehrsprognosen, Erwartungen an die Preisentwicklung für alle Technologien, Entwicklungen an den Finanzmärkten, Umsatzerwartungen für einzelne Dienste und die Kosten für den Betrieb des Netzes. Die Netzmodelle müssen sich über alle Netzschichten erstrecken. Die Zugangsnetze mit alternativen technologischen Konzepten wie *Fiber-to-the-Home* (FttH) oder *Hybrid Wireless Optical Broadband Access Networks* (WOBAN) sind dabei genauso abzubilden, wie auch Metro- und Kernnetze zur Anbindung an die nationalen und internationalen Netze. Aus Sicht der Betreiber sollten die Modelle alle wichtigen technischen Aspekte (Beschreibung von Netzknoten für IP, Ethernet, WDM und von physikalischen Verbindungen) und alle wichtigen betriebswirtschaftlichen Aspekte (Investitionskosten

für Hardware, Glasfasern, Tiefbauarbeiten, betriebliche Kosten wie Mietleitungen, Energiekosten und Personal, sowie Finanzierungs- und Abschreibungsalternativen) adäquat abbilden. Werden die Handlungsoptionen mit solchen Modellen unter Einbeziehung von Erwartungen an die Entwicklung und den Umsatz einzelner Dienste techno-ökonomisch analysiert und bewertet, so können Managemententscheidungen auf einer transparenten, solide quantifizierten Grundlage getroffen werden.

Effiziente Nutzung von Netzressourcen. Der heutige Betrieb von Telekommunikationsnetzen ist aufwändig, unflexibel und wenig ressourcenschonend. Mit hohem Personalaufwand werden auf allen Transportnetzebenen Verbindungen geplant und weitgehend manuell konfiguriert, was logischerweise die Dauer zum Einrichten einer Verbindung in die Länge zieht. Eine drastische Verkürzung dieses Zeitraumes würde ein völlig neues Geschäftsfeld eröffnen: Bei „Bandwidth on Demand“, könnten sich Kunden bei Bedarf Netzkapazität für einen kurzen Zeitraum zumieten. In letzter Konsequenz könnte dann das Hinzufügen von zusätzlicher Bandbreite für einen Kunden so einfach und schnell gehen wie eine Telefonverbindung herzustellen. Weiterhin will man aus Kostengründen die Übertragungsmedien immer besser ausnutzen, seien es die Kupferkabel der Teilnehmeranschlussleitung, die Mobilfunkkanäle oder die Glasfasern der Weitverkehrsnetze. Zur Bereitstellung der künftig benötigten Terabit-pro-Sekunde-Kapazitäten will man auf den optischen Übertragungsstrecken die Zahl der genutzten Wellenlängen pro Faser immer weiter erhöhen; man erreicht bereits heute einige Hundert Wellenlängen pro Faser. Dabei treten allerdings zunehmend nichtlineare Störeffekte auf, die wiederum bei der Modellierung von Übertragungsstrecken beachtet werden müssen.

MATHEMATISCHE HERAUSFORDERUNGEN

Welche Herausforderungen für die Mathematik ergeben sich aufgrund der oben skizzierten technischen Entwicklungslinien?

Schnelle, skalierbare Optimierungsverfahren. Die Konvergenz der Netze hin zu einer einheitlichen Multi-Service-Plattform hat einen für die Planung entscheidenden Nachteil. Die Grenzen zwischen Zugangs-, Aggregations- und Kernnetz verschwimmen zunehmend und die natürliche Dekomposition in Fragestellungen bzgl. einzelner Netzbereiche entfällt. Die in der Planung zu analysierenden Netze werden entsprechend um ein Vielfaches größer. Die heute entwickelten Methoden sind für größere Netze aber nicht ausgelegt. Das Ziel muss sein, die theoretischen Grundlagen der Verfahren zu verbessern und diese in geeigneter Weise mit heuristischen Ansätzen zu koppeln, so dass zukünftig auch große Netze schnell und zuverlässig optimierbar werden. Hier sind neuartige, rechenfreundliche Modelle ebenso gefragt wie Optimierungsverfahren, die wegen der auftretenden physikalischen Effekte auch die Behandlung nichtlinearer Problemklassen erlauben.

Integration von Planungsaufgaben. Bisher können mit den oben skizzierten Verfahren Aufgaben der Netzplanung unterstützt werden, die sich auf einen Zeitpunkt, eine technologische Schicht, eine Verkehrsmatrix oder auch eine feste Netzhierarchie beziehen. Diese Einschränkung ist in der mangelnden Leistungsfähigkeit der mathematischen Methoden begründet. Gerade im Hinblick auf die zukünftigen *Next Generation Networks* sind die so genannten Multi-X Aufgaben zu lösen. X steht dabei unter anderem für „layer“, „period“, „hour“, „level“, d. h. für die Planung über mehrere technologische Schichten, über eine zeitliche Entwicklung des Netzes, über unterschiedliche Verkehrsprognosen bzw. Verkehrsverteilungen und Netzhierarchieebenen. Diese integrierten Planungsaufgaben werden heute schon vereinzelt modelliert und für kleine Netze gibt es teilweise auch erfolgsversprechende Ansätze. Das Ziel muss hier ein Baukasten von Modellen und Methoden sein, mit dem flexibel, schnell und skalierbar auf die Planungsfragen der Telekommunikation geantwortet werden kann.

Robuste und stochastische Optimierung und Simulation. Die Schätzungen zukünftiger Verkehrsmengen und Kosten werden immer unsicher sein. Eine ausgefeilte Netzplanung, die auf eine möglichst effiziente Nutzung der Ressourcen abzielt muss daher gegen Schwankungen dieser Eingangsgrößen abgesichert werden. Dies könnte mit Methoden der robusten und stochastischen Optimierung geschehen. Allerdings sind erhebliche Forschungsanstrengungen notwendig, um die Leistungsfähigkeit dieser Methoden so zu steigern, dass die zukünftigen Netze damit planbar werden. Die etablierten Methoden der Simulation und der Szenarienanalyse können dabei sicherlich nicht ersetzt werden. Vielmehr sollten letztere Methoden weiter erforscht werden. Je besser man beispielsweise die Stochastik von Datenverkehr analytisch beschreiben kann, umso besser lassen sich darauf robuste Planungsverfahren aufsetzen. In Zukunft wird ferner die so genannte *Selbstorganisation von Netzen* zunehmende Bedeutung haben. Für die Planung ergeben sich in diesem Zusammenhang interessante neue Fragestellungen: wie kann man bei der Infrastrukturplanung die Ressourcen gezielt und kosteneffektiv so bereitstellen, dass der sich selbstorganisierende Betrieb möglichst reibungslos verläuft?

EMPFEHLUNG

Der Einsatzbereich für innovative, mathematische Methoden ist groß. Die mathematische Forschung sollte besonderen Fokus auf die Behandlung strategischer Investitionsentscheidungen legen. Die erzielbare Hebelwirkung ist hier am größten, da die Telekommunikationsnetze sich in den kommenden zehn Jahren grundlegend wandeln werden. Sowohl der Ausbau der glasfaserbasierten Zugangsnetze mit Endnutzerbandbreiten von bis zu 10 Gbit/s als auch der Umbau der Produktionsplattformen hin zu Multi-Service-Netzen wird mit enormen Investitionen verbunden sein. Aufgaben der operativen und taktischen Netzplanung sind langfristig für die Mathematik auch interessant. Um hier jedoch einen nennenswerten Einfluss ausüben zu können, müssen sich Tools

und Prozesse so verändern, dass innovative Methoden leichter integriert werden können.

5 WAS ZU TUN IST – VORSCHLÄGE UND INITIATIVEN

Der effiziente Aufbau und Betrieb von Telekommunikationsinfrastrukturen ist von enormer Bedeutung. Fehler beim Entwurf und bei der Auslegung der Netze ziehen erhebliche Folgekosten nach sich oder schwächen ganze Regionen und Wirtschaftsbereiche. Eine sorgfältige, methodische Planung ist grundlegend für den erfolgreichen Einsatz der vielfältigen, laufend weiter entwickelten Netztechnologien. Die dabei auftretenden Aufgaben könnten durch engere Zusammenarbeit von anwendungsorientierten Mathematikern und methodisch ausgebildeten Ingenieuren deutlich besser gelöst werden.

Die Interdisziplinarität sollte daher in der Ausbildung, im Rahmen von Verbundforschungsvorhaben und in der Zusammenarbeit von Wissenschaft und Wirtschaft gefördert werden.

5.1 AUSBILDUNG

Dem Mathematikstudium fehlt es teilweise an Praxisnähe. Neben der zwingend notwendigen theoretischen Ausbildung, sollte es im Hauptstudium verstärkt Angebote geben, die sich mit der Modellierung, Simulation und Optimierung realer Prozesse beschäftigen.

Das universitäre Ingenieursstudium sieht heute schon eine sehr gute mathematische Ausbildung vor. Die Schwerpunkte liegen dabei auf analytischen und numerischen Methoden. Mit Blick auf die Herausforderungen bei der Planung, Konfiguration und Optimierung von Kommunikationsnetzen gibt es zusätzlichen Ausbildungsbedarf in der Graphentheorie, Kombinatorik, der linearen und ganzzahligen Programmierung sowie den Methoden des Operations Research.

5.2 WIRTSCHAFT

Schwierig zu realisieren ist die Forderung, dass auch in der Wirtschaft mehr methodisches Know-How aus dem mathematischen Umfeld genutzt wird. Das liegt zum einen an der unterschiedlichen Herangehensweise bei der Problemlösung und der fehlenden gemeinsamen Sprachbasis. Es gibt aber auch praktische Gründe. So ist z. B. die heute etablierte IT-Infrastruktur der Firmen, die für die Planungswerkzeuge genutzt wird, ein großes Hindernis bei der Integration innovativer Methoden in das IT-Umfeld von Netzbetreibern. Das im folgenden Abschnitt vorgeschlagene Verbundforschungsvorhaben soll dazu dienen, einen Schritt in die Richtung offenerer Prozesse für die Planung und den Betrieb von Kommunikationsnetzen zu gehen. Abhängig vom Grad und der Art der Industriebeteiligung könnte dieses Vorhaben daher auch dazu dienen, ein Referenzsystem für eine innovative Planungsinfrastruktur aufzubauen – zweifellos eine langfristige, aber

lohnende Vision. Damit verbunden könnte auch eine Standardisierung von Planungsprozessen oder -zielen (analog zu ISO 9001) den Service Providern helfen, entsprechend sorgfältig geplante und „mathematisch abgesicherte“ Netze zu vermarkten.

5.3 FORSCHUNGSSCHWERPUNKT

Für die Kommunikationstechnik halten wir ein Verbundforschungsvorhaben zwischen Instituten der Mathematik und des Ingenieurwesens für einen erfolgversprechenden Weg. Ein solches Vorhaben – je nach Ausprägung bei der DFG oder beim BMBF angesiedelt – könnte den Titel *Modellierung, Planung und Bewertung von Kommunikationsnetzen* tragen. Die zentrale Aufgabe wäre es, Ingenieure und Mathematiker gemeinsam eine umfassende Modellierung des Netzes sowie der Planungs- und Optimierungsaufgaben entwickeln zu lassen. Zur Bewertung sollte dies durch adäquate Simulations- und Analysemöglichkeiten ergänzt werden. Ziel wären u. a. Methoden für die in den Abschnitten 3 und 4 ausgeführten Herausforderungen vor allem im Bereich der großen Netze (Skalierung), der Dienstqualitätsplanung und der mehrschichtigen Netze. Auf Basis der Modelle sollten auch Prozesse entwickelt werden, mit denen flexibel Antworten auf die Fragen der strategischen, taktischen und operativen Planung gegeben werden können. Ohne wirklich eine Service-Oriented-Architecture-Infrastruktur zu schaffen, sollten die Modelle lose gekoppelt sein und die Daten sollten flexibel über offene Schnittstellen zwischen den Modellen ausgetauscht werden können.

Zur Sicherstellung der praktischen Relevanz sollte ein solches Vorhaben durch die strategischen Abteilungen von Betreibern oder Herstellern beraten werden. Mit deren Unterstützung könnten auch realistische Referenznetzdaten generiert werden.

6 LITERATUR

- [1] <http://sndlib.zib.de>.
- [2] <http://www.bmbf.de/de/6103.php>.
- [3] <http://www.celtic-initiative.org/Projects/100GET/>.
- [4] <http://www.ist-nobel.org/Nobel/servlet/Nobel.Main>.
- [5] http://eurongi.enst.fr/en_accueil.html.
- [6] Andreas Bley und Thorsten Koch. Optimierung in der Planung und beim Aufbau des G-WiN. *DFN-Mitteilungen*, 54:13–15, 2000.

- [7] Andreas Bley und Markus Pattloch. Modellierung und Optimierung der X-WiN Plattform. *DFN-Mitteilungen*, 67:4-7, 2005.
- [8] Edsger W. Dijkstra. A Note on Two Problems in Connexion with Graphs. *Numerische Mathematik*, 1:269-271, 1959.
- [9] Wayne D. Grover. *Mesh-based Survivable Networks*. Prentice-Hall, 2004.
- [10] Cisco Systems Inc. Cisco Visual Networking Index – Forecast and Methodology, 2007-2012, 2008.
- [11] Ken Ambts, Sebastian Cwilich, Mei Deing, David J. Houck, David F. Lynch, und Dicky Yan. Optimizing Restoration Capacity in the AT&T Network. *INTERFACES*, 30(1):26-44, January-February 2000.
- [12] Barrett Lyon et al., January 2005. <http://www.opte.org/maps>.
- [13] Leonard Kleinrock. *Queueing Systems*. John Wiley & Sons, 1975.
- [14] Michal Pioro und Deepankar Mehdi. *Routing, Flow, and Capacity Design in Communication and Computer Networks*. Elsevier, 2004.
- [15] Robert E. Bixby, E. Andrew Boyd, und Ronni R. Indovina. MIPLIB: A test set of mixed integer programming problems. *SIAM News*, 25(20), March 1992.
- [16] Robert E. Bixby. Solving Real-World Linear Programs: A Decade and More of Progress. *Operations Research*, 50(1):3-15, January-February 2002.

> MATHEMATIK IN DER DRAHTLOSEN KOMMUNIKATION

HOLGER BOCHE UND ANDREAS EISENBLÄTTER

1 EXECUTIVE SUMMARY

Die Mobilkommunikation besitzt heute große wirtschaftliche Bedeutung. Sie hat seit der Einführung des derzeit noch dominierenden Mobilfunkstandards GSM Anfang der neunziger Jahre einen enormen Einfluss auf das gesellschaftliche Leben genommen. Mit neuen Anwendungsfeldern, z. B. im Maschinen- und Anlagenbau, in der Automobilindustrie und im Wohnbereich, wird die Bedeutung der Mobilkommunikation weiter steigen. Um der Vielfalt der Anwendungen gerecht zu werden und die Wünsche nach neuen Diensten erfüllen zu können, werden fortwährend neue Mobilfunksysteme eingeführt.

Ohne die Digitalisierung des Mobilfunks wäre diese Entwicklung unmöglich gewesen. Der Mathematik kommt dabei eine tragende Rolle zu. Sie ist die Grundlage für die Informations- und Kommunikationstheorie, sie ist Wegbereiter für die Entwicklung neuer Übertragungsverfahren und wesentliches Instrument für die Planung und Optimierung von Mobilfunknetzen. Dabei werden z. B. die Wahrscheinlichkeitstheorie und die diskrete Mathematik in der Informationstheorie angewandt; die lineare Algebra, die konvexe Optimierung und die Spieltheorie bei der Entwicklung neuer Übertragungsverfahren; sowie die lineare, kombinatorische und stochastische Optimierung bei der Planung von Funknetzen.

Trotz der gewaltigen Fortschritte sind wesentliche Grundlagenprobleme der drahtlosen Kommunikation immer noch ungelöst. Einige Beispiele hierfür sind eine Netzwerkinformationstheorie für komplexere Mehrteilnehmersysteme (etwa Ad-hoc- und Sensornetze), eine Quanteninformationstheorie für miniaturisierte Systeme, eine Theorie für selbstorganisierende Netze und die Entwicklung neuer Paradigmen zur Nutzung des knappen Frequenzspektrums. In allen genannten Fällen muss das Zusammenspiel von Nachrichtentechnik und Mathematik die wesentlichen Beiträge zur Lösung liefern.

2 ERFOLGSGESCHICHTEN

Die ökonomische und gesellschaftliche Bedeutung der Mobilkommunikation und damit des Mobilfunks ist herausragend. Die Mobilfunknetze sind heute ein wichtiger Teil der Telekommunikationsinfrastruktur. In vielen Teilen der Erde bietet Mobilfunk überhaupt

die erste ökonomisch vertretbare Basis für eine großflächige Kommunikationsinfrastruktur.

Das Telefonieren per Mobiltelefon sowie die Kommunikation mittels kurzer Textnachrichten verbreitet sich seit Anfang der neunziger Jahre mit einer kaum vorstellbaren Dynamik. Die Grundlage hierfür war die Einführung der zweiten Mobilfunkgeneration (2G), die eine digitale Datenübertragung nutzt und damit im Vergleich zur zuvor üblichen Analogtechnik einen erheblichen Leistungssprung ermöglicht. Das *Global System for Mobile Communication*, kurz GSM, war eines der ersten 2G-Systeme und ist bis heute das erfolgreichste. Wurden in Deutschland im Jahr 1992 nur knapp eine Million Mobilfunkkunden (erste GSM-Nutzer, klassische Autotelefone, usw.) gezählt, waren es Ende des Jahres 2007 bereits mehr als 97 Millionen [29]. Im Durchschnitt verfügt also jeder Einwohner über mehr als einen Mobilfunkanschluss. Weltweit gibt es mehr als 3 Milliarden GSM-Kunden und täglich kommen über eine Million hinzu [6].

In vielen Industrieländern ist mit UMTS ein System der nächsten, dritten Generation (3G) eingeführt. Dieses soll langfristig GSM ablösen und wird bereits von mehr als 200 Millionen Nutzer in 73 Ländern genutzt [16]. UMTS steht für *Universal Mobile Telecommunications System* und ist mit einer maximalen Datenrate von 384 kbit/s insbesondere für mobile Datendienste ausgelegt. Im Vergleich zu GSM steigt die maximale Datenrate pro Verbindung etwa um den Faktor 30. In Europa sind zudem bereits Erweiterungen von UMTS installiert, die nochmals eine Steigerung der Datenrate sowie eine Verkürzung der Reaktionszeit ermöglichen. Die Reaktionszeit ist grob gesprochen, die Zeit die vergeht, bis auf eine Datenanfrage (z. B. dem Klicken mit der Maus auf einer Internetseite) die Daten geliefert werden (im Beispiel also die nächste Internetseite). Mit den UMTS-Erweiterungen *High-Speed Downlink Packet Access* (HSDPA) sind Datenraten bis 7,2 Mbit/s von der Basisstation zum Nutzer (Downlink) und mit *High-Speed Uplink Packet Access* (HSUPA) Raten von 1,4 Mbit/s in der umgekehrten Richtung (Uplink) möglich. Alternativ oder in Ergänzung zu UMTS wird in einigen Länder WiMAX (*Worldwide Interoperability for Microwave Access*) eingeführt, das für den Mobilfunk nochmals bessere Leistungseigenschaften als UMTS mit seinen Erweiterungen besitzt.

Mit UMTS *Long Term Evolution*, kurz LTE, wird zur Zeit ein weiterer internationaler Standard für ein Mobilfunksystem verabschiedet. LTE wird mindestens so leistungsfähig wie WiMAX sein, soll sich nahtlos mit GSM und UMTS kombinieren lassen und im Gegensatz zu den bestehenden Systemen wesentlich einfacher zu betreiben sein.

2.1 MOBILKOMMUNIKATION BRAUCHT MATHEMATIK

Der intensive Einsatz von Mathematik ist einer der Grundpfeiler dieser Entwicklung. Bedeutende Beiträge der Mathematik werden für drei Bereiche dargestellt:

- Grundlagen der Mobilkommunikation
- Entwicklung von Mobilfunksystemen
- Planung und Optimierung von Mobilfunknetzen

Die Entwicklung neuer Paradigmen zur fairen Nutzung des knappen Funkfrequenzspektrums lässt sich schlecht einem der Bereiche zuordnen. Hierauf wird in Abschnitt 5.4 eingegangen.

Die drei nächsten Abschnitte zeigen für jeden der Bereiche jeweils exemplarisch einen Beitrag auf, den Mathematik für die Entwicklung und den Betrieb moderner Mobilfunksysteme leistet. Damit wird auch der Anteil der Mathematik an effizienter mobiler Telekommunikation insgesamt sichtbar. Stellvertretend für die Ebene der Grundlagenforschung wird in Abschnitt 2.2 dargestellt, wie Mehrantennensysteme die Übertragungskapazität vervielfachen können. In Abschnitt 2.3 werden Mehrnutzerempfänger als Beispiel für den Beitrag bei der Entwicklung von Mobilfunksystemen ausgeführt. Auf der Ebene der Planung und Optimierung von Funknetzen werden in Abschnitt 2.4 die mathematischen Beiträge zur Frequenzplanung für GSM-Netze vorgestellt.

2.2 DAS FUNDAMENT DER KOMMUNIKATIONSTHEORIE

Die Shannonsche Informationstheorie bildet die mathematische Grundlage für die moderne Kommunikationstheorie. Sie befasst sich damit, wie eine Kommunikation zwischen Teilnehmern über einen verrauschten Kanal ermöglicht werden kann, wie sich Teilnehmer die Kommunikationsressourcen, beispielsweise Frequenz, Zeit, Raum, Leistung, teilen sollen.

Im Wesentlichen versucht die Informationstheorie, Antworten auf zwei fundamentale Fragen zu geben. Dies ist erstens die Frage, wieviel Information in einem gegebenen Kommunikationsszenario maximal übertragen werden kann und zweitens, welche praktischen Kommunikationstechniken angewendet werden müssen, um die maximalen Übertragungsraten tatsächlich zu erzielen. Ein zentrales Resultat der Informationstheorie ist das Shannonsche Kanalkodierungstheorem [31].

Das Kanalkodierungstheorem besagt, dass für jede Übertragungsrate R , die kleiner als die Kanalkapazität C ist, die Wahrscheinlichkeit fehlerhafter Dekodierung am Empfänger beliebig klein gemacht werden kann. Dagegen ist für Raten $R > C$ eine fehlerfreie Übertragung nicht möglich. Die Kanalkapazität gibt also eine obere Schranke für die maximal mögliche fehlerfreie Übertragungsrate an.

Abb. 1 erläutert diese Kapazitätsgrenze und zeigt auf, mit welchen methodischen Ansätzen diese Grenze klassischer Übertragungssysteme überwunden werden kann. Im Mittelpunkt steht die gleichzeitige, störungsarme Übertragung zwischen zwei Teilnehmern, einem Sender und einem Empfänger, über jeweils mehrere Antennen. Die dargestellte Theorie des Kanalkodierungstheorems für die Kommunikation zwischen zwei Teilnehmern wurde in großen Teilen bereits in Shannons Artikel von 1948 entwickelt.

Als sehr viel schwieriger hat es sich erwiesen, eine Informationstheorie für mehrere Teilnehmer zu entwickeln. Zwei wichtige Mehrteilnehmerkommunikationssysteme sind der Vielfachzugriffs- und der Verteilkanal. Beim Vielfachzugriffskanal wollen mehrere Teilnehmer gleichzeitig Nachrichten zu einem Teilnehmer übertragen. Dieses Kommuni-

Abb. 1: Kanalkapazität von Ein- und Mehrantennensystemen

Der Begriff der Kanalkapazität geht auf den Mathematiker C. E. Shannon zurück, der 1948 mit seiner wegweisenden Arbeit *A Mathematical Theory of Communication* [31] die Grundlagen der heutigen Informationstheorie schuf.

Die Kanalkapazität C ist definiert als das Supremum aller fehlerfrei erreichbaren Raten. Für einen diskreten, gedächtnislosen Kanal zeigte Shannon, dass die Kapazität durch das Maximum der Transinformation über alle möglichen Eingangswahrscheinlichkeitsverteilungen $p(x)$ gegeben ist:

$$C = \max_{p(x)} I(X; Y), \quad (1)$$

Die Transinformation $I(X; Y)$ misst die Stärke der statistischen Abhängigkeit zwischen dem Eingang X und dem Ausgang Y . Je größer die Transinformation ist, desto besser kann also vom empfangenen Signal auf das gesendete Signal geschlossen werden.

Für den AWGN-Kanal ist das empfangene Signal y durch

$$y = hx + n$$

gegeben. Dabei bezeichnet x das zu übertragene Signal, h den Kanaleinfluss und n das additive weiße Gaußsche Rauschen (additive white Gaussian noise, AWGN) am Empfänger. Ist die Leistung des Senders durch P beschränkt, so gibt

$$\begin{aligned} C &= \max_{p(x): \mathbb{E}\{|X|^2\} \leq P} I(X; Y) \\ &= \log \left(1 + \frac{|h|^2}{\sigma_n^2} P \right) \end{aligned} \quad (2)$$

die Kanalkapazität an, wobei $\frac{|h|^2}{\sigma_n^2} P$ das Signal-zu-Rauschleistungsverhältnis (SNR) beschreibt, siehe z. B. [11].

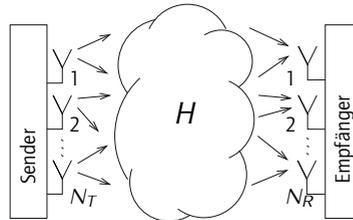
Gleichung (2) gilt für den Fall, dass Sender und Empfänger jeweils mit einer Antenne ausgestattet sind. Diese Konfiguration wird als SISO-Kanal (*Single Input Single Output*) bezeichnet. Betrachtet man den allgemeineren Fall mit N_T Sende- und N_R Empfangsantennen, spricht man von einem MIMO-Kanal (*Multiple Input Multiple Output*). Hier ist der Empfangsvektor Y durch

$$Y = HX + N$$

gegeben. Der Sendevektor X enthält die Sendesignale für die entsprechenden Sendeantennen. Die Kanalmatrix H beschreibt den Kanaleinfluss. Der (k, j) -te

Eintrag der Matrix gibt den Einfluss des Kanals auf ein Signal zwischen der j -ten Sende- und der k -ten Empfangsantenne wieder.

MIMO-Kanal:



Die Kapazität (1) dieses Kanals mit Sendeleistungsbeschränkung P ist

$$C = \max_{\text{tr}(Q) \leq P} \log \det \left(I + \frac{1}{\sigma_n^2} H Q H^H \right), \quad (3)$$

wobei $Q = X X^H$ die Sendekovarianzmatrix beschreibt [33, 34].

Mittels der Singulärwertzerlegung lässt sich der MIMO-Kanal als parallele SISO-Kanäle darstellen. Ist diese Zerlegung bekannt, so liefert die so genannte *waterfilling solution* die für (3) optimale Kovarianzmatrix

$$Q^* = V \Lambda_Q V^H,$$

wobei die Eigenvektoren von Q^* den rechtsseitigen Singulärvektoren des Kanals entsprechen. Das resultierende System kann durch diese Wahl von Q^* als $N = \min\{N_T, N_R\}$ parallele SISO-Kanäle interpretiert werden, die über ihre Sendeleistungen gekoppelt sind. Die vorhandene Gesamtsendeleistung wird auf die einzelnen Kanäle aufgeteilt. Dadurch ergibt sich schließlich die Kapazität des MIMO-Kanals

$$C = \sum_{i=1}^N \log \left(1 + \frac{\lambda_i}{\sigma_n^2} P_i \right),$$

wobei λ_i den Einfluss des i -ten SISO-Kanals und P_i die entsprechende zugeteilte Sendeleistung beschreibt. Die Kapazität des MIMO-Kanals entspricht der Summe der einzelnen SISO-Kanalkapazitäten. Durch mehrere Sende- und Empfangsantennen kann die Kapazität des Systems also deutlich erhöht werden.

kationsszenario modelliert informationstheoretisch den Uplink in einem zellularen Mobilfunksystem, d. h. mehrere Teilnehmer senden gleichzeitig ihre Signale zur Basisstation. Beim Verteilkanal sendet ein Teilnehmer zu mehreren Teilnehmern jeweils unterschiedliche Nachrichten. Der Verteilkanal ist damit das informationstheoretische Modell für den Downlink eines zellularen Mobilfunksystems, d. h. die Basisstation sendet unterschiedliche Informationen an mehrere Teilnehmer. Die Charakterisierungen der optimalen Übertragungsstrategien für den Vielfachzugriffs- und in wichtigen Spezialfällen für den Verteilkanal sind in den letzten 30 Jahren erzielt worden. Das gewonnene Verständnis über die Kapazitätsregionen dieser Mehrteilnehmersysteme hat großen Einfluss auf die Entwicklung von UMTS, HSDPA, HSUPA und LTE genommen.

Neue mathematische Techniken müssen entwickelt werden, um viele Fragestellungen der Informationstheorie in der Praxis zu lösen. Als Beispiel hierfür sei die von Lovász in [23] eingeführte Theta-Funktion genannt, siehe auch die sehr interessante Diskussion in [3]. Mit ihr gelang es Lovász, die Kapazität für eine Kommunikation ohne Fehler für bestimmte Klassen von Kommunikationsgraphen zu berechnen. Die Lovász-Theta-Funktion wurde von Grötschel, Lovász, Schrijver in [20] verallgemeinert. In dieser Verallgemeinerung wird sie heute in vielen Bereichen der Mathematik genutzt.

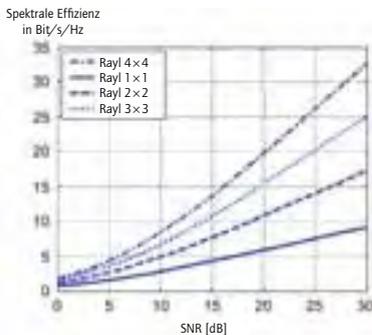
2.3 SCHLÜSSELTECHNIKEN

Damit ein neues Kommunikationssystem die vorgegebenen Durchsatzziele erreichen kann, muss es eine bestimmte spektrale Effizienz erzielen. Die spektrale Effizienz gibt an, wieviel Bit pro Sekunde pro Hertz übertragen werden können. Wie effizient das System werden kann, hängt auf der einen Seite von der physikalischen Signalausbreitung ab, andererseits kann dies aber auch der Systemdesigner beeinflussen, z. B. durch die Anzahl der verwendeten Antennen.

In Abb. 2 ist dieser Fall für ein Mehrantennensystem dargestellt. Der Systemdesigner kann z. B. analysieren, wieviele Sende- und Empfangsantennen die einzelnen Teilnehmer nutzen sollten. In Abhängigkeit von der Anzahl der verwendeten Antennen ist die erzielbare spektrale Effizienz durch Beziehung (3) in Abb. 1 gegeben. Als Kanalmodell wurde das Rayleigh-Fading-Modell benutzt [34]. Dieses Modell ist mit Hilfe von Grenzwertsätzen der Wahrscheinlichkeitstheorie aus den physikalischen Wellenausbreitungsbedingungen abgeleitet. Die Einträge der Kanalmatrix sind unabhängig und identisch nach einer Gauß-Verteilung mit dem Mittelwert Null verteilt. Das Modell beschreibt sehr gut, wie sich ein schmalbandiges Signal ausbreitet, wenn zwischen Sender und Empfänger keine Sichtverbindung besteht. Im Allgemeinen ist eine analytische Auswertung der spektralen Effizienz ein schwieriges Problem, weil hierfür zunächst ein geeignetes mathematisches Modell des Kanals entwickelt werden muss.

Die bisher in diesem Abschnitt diskutierten Resultate befassen sich im Wesentlichen mit der Optimierung der Kommunikation zwischen zwei Teilnehmern. Sendesignale anderer Teilnehmer können diese Kommunikation allerdings stören. Ein wichtiger An-

Abb. 2: Erzielbare Effizienz eines Mehrantennensystems in Abhängigkeit von dem Signal-zu-Rauschleistungs-verhältnis am Empfänger



satz der modernen Kommunikationstheorie besteht darin, die Störungen nicht als Rauschen zu interpretieren, sondern ihre spezielle Struktur auszunutzen, um den effektiven Einfluss der Störungen zu minimieren. Mehrnutzerempfänger lösen diese Aufgabe. Je nach Komplexität der Implementierung kann ein Mehrnutzerempfänger Störungen unterschiedlich gut unterdrücken. Der informationstheoretisch optimale Mehrnutzerempfänger für den Vielfachzugriffskanal (siehe Abschnitt 2.2) erlaubt sogar die vollständige Unterdrückung der durch ausgewählte Teilnehmer erzeugten Störungen.

Die neuen Konzepte für Mehrantennensysteme und Mehrnutzerempfänger haben die Entwicklung von UMTS, HSDPA, HSUPA, WiMAX und LTE wesentlich beeinflusst. Ohne diese Techniken wären die Steigerungen der maximalen Übertragungsraten nicht möglich gewesen. Für die Analyse und den Entwurf der entsprechenden Kommunikationssysteme waren unterschiedlichste mathematische Disziplinen bedeutsam. Beispiele hierfür sind die Kombinatorik, die diskrete Mathematik, die mathematische Optimierung, die Funktionalanalysis und die Theorie der zufälligen Matrizen.

2.4 OPTIMIERUNG VON GSM-NETZEN

Die Planung und Optimierung von GSM-Mobilfunknetzen ist äußerst anspruchsvoll. Der Netzaufbau ist kein kurzfristiger Prozess von zwei oder drei Jahren, sondern es findet ein steter Aus- und Umbau der Netze statt. In Deutschland werden noch im Jahr 2008 – und damit 16 Jahre nach dem Start – mehrere hundert Millionen Euro in den Ausbau der Infrastruktur investiert [19, 29]. Zusätzlich zu den einmaligen Kosten beim Aufbau eines Netzes fallen regelmäßig Betriebskosten an. Vor diesem Hintergrund wollen Netzbetreiber eine effiziente und kostengünstige Infrastruktur betreiben.

Methoden der kombinatorischen Optimierung, einem Teilgebiet der Mathematik, lassen sich auf allen Planungsebenen einsetzen: der langfristigen Planung der Netzarchitektur mit einem Horizont von mehreren Jahren, der mittelfristigen Planung der Netzstruktur und der kurzfristigen, operativen Planung der Konfiguration einzelner Netzelemente mit einem Horizont von wenigen Monaten.

Die mittelfristige Planung der Netzabdeckungsfläche und die kurzfristige Anpassung der Netzkapazität können für GSM entkoppelt betrachtet werden. Die Abdeckung wird maßgeblich durch die Auswahl der Standorte und der Ausrichtung der üblicherweise drei Sektorantennen bestimmt, die jeweils eine Zelle des Netzes definieren. Die Kapazität wird pro Zelle durch die Anzahl der vorgehaltenen Frequenzkanäle bestimmt.

Die Aufgabe der Frequenzplanung ist es, auszuwählen, welche Frequenzkanäle eine Zelle betreiben soll. So steht beispielsweise eine kleine Anzahl von etwa einhundert verschiedenen Kanälen der Nachfrage von mehreren zehntausend Zellen nach einem oder mehreren Kanälen gegenüber. Die Kanäle müssen wiederverwendet werden. Hierbei sind viele Bedingungen an eine zulässige Kanalzuteilung zu beachten. Deshalb werden

Abb. 3: Dargestellt ist ein Ausschnitt eines GSM-Netzes mit 19 Standorten. Die kleinen Pfeile zeigen die Position und Ausrichtung der Sektorantennen. Im graphentheoretischen Modell (a) werden nur Störungen ab einem einheitlichen Schwellwert berücksichtigt. Eine differenzierte Betrachtung der Störungsbeziehungen zwischen Zellen ist mit der Interferenzmatrix in (b) möglich. Hierbei stehen rote Linien für stärkere Störungsbeziehungen als gelbe. In (c) geben die Linien für eine konkrete Zuweisung an, welche der Störungsbeziehungen nicht durch unterschiedliche Frequenzen aufgelöst wurden. Die Farbwahl des Hintergrundes stellt die Stärke der Störungen ortsbezogen dar. Rot weist auf schlechte Qualität hin.

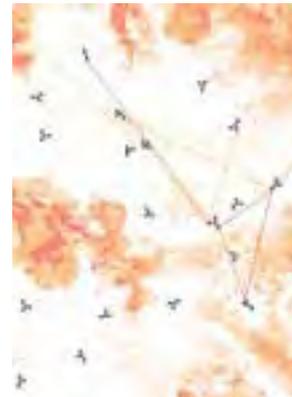
(a) uniformer Schwellwert



(b) Interferenzmatrix



(c) Qualität einer Zuweisung



bereits seit Anfang der neunziger Jahre Verfahren zur automatischen Frequenzplanung entwickelt.

Die ersten Modellierungen waren rein graphentheoretisch und unterscheiden Störungen nicht nach ihrer Stärke [17, 21], siehe Abb. 3a. Mit wachsendem Kapazitätsbedarf erwies sich die anfängliche Modellierung als zu einfach. Ab Mitte der neunziger Jahre wurden in enger Zusammenarbeit von Nachrichtentechnikern und Mathematikern detailliertere Modelle eingeführt, die Interferenz berücksichtigen [13, 28], siehe Abb. 3b. Auf der Basis dieser Modelle wurden neue Zuweisungsverfahren entwickelt [1, 10].

Seit Ende der neunziger Jahre verfügen die meisten Netzbetreiber über entsprechend leistungsfähige Frequenzplanungsverfahren, entweder in der Form von Eigenentwicklungen oder als kommerziell erworbene Produkte. Damit erreichen die Netzbetreiber die intensive Nutzung der GSM-Frequenzen, die heute notwendig ist.

3 WISSEN SCHAFFT FORTSCHRITT

In Abschnitt 2 wurde anhand von drei Beispielen die große Wirkung aufgezeigt, die Mathematik in der Nachrichtentechnik entfaltet. In diesem Abschnitt wird für konkrete Anwendungen gezeigt, wie diese Wirkung erzielt wird. Der Abschnitt endet mit einer Diskussion über den Zusammenhang zwischen erzielten technologischen Fortschritten in der Halbleiterindustrie und der Einsetzbarkeit komplexer Übertragungstechniken. Auch hier spielt die Mathematik eine wichtige Rolle. In den Beispielen zeigt sich die Bedeutung der Mathematik auf unterschiedlichen Ebenen. Sie spielt einmal die zentrale Rolle in der exakten Formulierung der nachrichtentechnischen Problemstellung; sie stellt dann die Techniken zur Verfügung, um die Struktur der Probleme zu analysieren; sie bietet Methoden, die entsprechende Lösungen effizient berechnen können, und bezüglich des axiomatischen Ansatzes für interferenzbeschränkte Mobilfunksysteme ist sie die Basis für eine universelle Theorie.

3.1 OPTIMIERTE RESSOURCENVERGABE IN ZELLULAREN FUNKNETZEN

Bei der Weiterentwicklung von UMTS zu HSDPA (High Speed Downlink Packet Access) und HSUPA (High Speed Uplink Packet Access) konnte durch die Technik der kanalangepassten Übertragung die spektrale Effizienz wesentlich erhöhen. Den Ausgangspunkt hierfür bildet die Beobachtung, dass die Qualität einer Funkverbindung wegen der physikalischen Natur von Funkwellen sehr stark davon abhängt, wo sich der Teilnehmer aufhält. Wenn sich ein Teilnehmer bewegt, verändert sich die Qualität des Kanals.

In einer Funkzelle mit vielen Teilnehmern ist die Wahrscheinlichkeit groß, einen Teilnehmer mit guter Kanalqualität zu finden. Die Idee der kanalangepassten Übertragung bei HSDPA ist, Nachrichten bevorzugt dann zum Teilnehmer zu senden, wenn der Kanalzustand gut ist. Auf diese Weise kann die Gesamtübertragungsrate erhöht werden. Für den Nachrichtentechniker tritt nun ein neues Problem auf, das Problem der fairen Res-

sourcenallokation. Wenn die Ressourcen so verteilt werden, dass alle Nutzer denselben Datendurchsatz erhalten, führt das in der Regel zu einem schlechten Gesamtdurchsatz für das Netz.

Ein neuer Ansatz zur Ressourcenallokation hat die Maximierung des Gesamtdurchsatzes zum Ziel. Eine Utility-Funktion bewertet aus Sicht des Netzes den aktuell möglichen Datendurchsatz zu jedem Nutzer. Die Ressourcen werden so aufgeteilt, dass die Utility-Funktion maximiert wird. Dieser Netzbetreiber-zentrierte Ansatz garantiert im Allgemeinen jedoch keine Fairness für die Teilnehmer.

Eine Herausforderung bei der Entwicklung von optimalen Schemen für HSDPA und HSUPA besteht darin, die Disziplinen der Informations- und Warteschlangentheorie miteinander zu verbinden. Abb. 4 stellt die kanalangepasste Übertragung für den Downlink dar und hebt die Bedeutung der Datenpufferung für dieses Konzept hervor. Auf Basis der bisher erzielten theoretischen Resultate (siehe Abb. 5) wurden Heuristiken für praktisch einsetzbare Scheduler entwickelt. Im Allgemeinen ist es aber ein offenes Problem, Informations- und Warteschlangentheorie mathematisch zufriedenstellend zu verbinden.

Die kanalangepasste Übertragung ist ein wichtiges Beispiel für die Optimierung von Mobilfunksystemen durch Nutzung von Crosslayer-Information. Dieser Ansatz ist in Abb. 6 beschrieben.

Abb. 4: Mehrnutzer-Downlink-System und die Pufferstruktur. H bezeichnet den Zustand des zeitabhängigen Verteilkanals. Die Vektoren (a_1, \dots, a_M) , (r_1, \dots, r_M) und (q_1, \dots, q_M) sind die momentanen Ankunftsdatenraten, Datenübertragungsraten und die Pufferzustände.

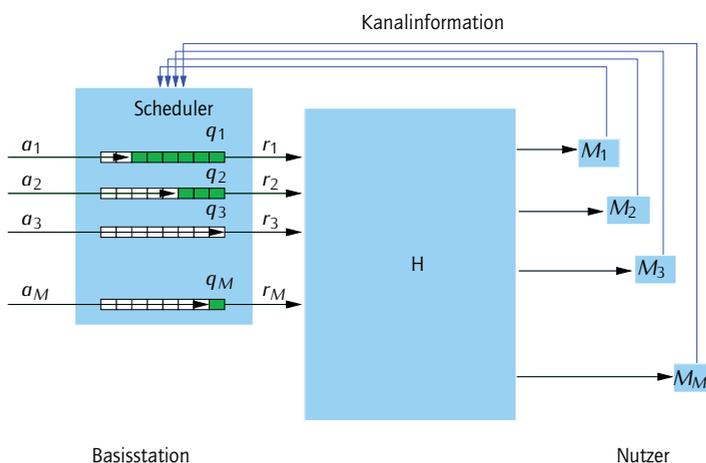


Abb. 5: Scheduling nach maximalem Nutzen

Das Scheduling-Problem kann wie folgt formuliert werden: Sei $t \in \mathbb{N}$ ein diskreter Zeitparameter des betrachteten, zufälligen Prozesses für M Nutzer. Um ganz allgemein unterschiedliche physikalische Gegebenheiten (Mehrantennensysteme, Kodierung Sw) abbilden zu können, wird die physikalische Schicht durch einen Ratenvektor $r_t = (r_t^{(1)}, \dots, r_t^{(M)}) \in \mathbb{R}_+^M$ dargestellt. Dieser Vektor muss in jedem Zeitintervall innerhalb einer Kapazitätsregion $C_t(p) \subset \mathbb{R}_+^M$ mit allgemeinen Leistungsbudgets $p = (p^{(1)}, \dots, p^{(M)}) \in \mathcal{P}_t$, wobei \mathcal{P}_t eine zulässige Leistungsregion ist, liegen. Es ist hervorzuheben, dass die Region $C_t(p)$ nicht notwendigerweise eine spezielle Struktur aufweisen muss (z. B. Konvexität). Um Stabilität des Systems zu gewährleisten, wird der Pufferzustand $q_t = (q_t^{(1)}, \dots, q_t^{(M)}) \in \mathbb{R}_+^M$ der Nutzer eingeführt und beim Scheduling berücksichtigt. Der Ratenvektor r_t ergibt sich dann als Lösung des folgenden Optimierungsproblems:

$$\begin{aligned} \max \quad & \mu(q_t)^T r_t \\ \text{s.t.} \quad & r_t \in C_t(p), \\ & p \in \mathcal{P}_t \end{aligned} \quad (4)$$

Hier ist $\mu = (\mu^{(1)}, \dots, \mu^{(M)}) : \mathbb{R}_+^M \rightarrow \mathbb{R}_+^M$ eine vektorwertige Abbildung, welche das jeweilige Ziel-

kriterium beinhaltet (z. B. minimale Verzögerung). Es kann gezeigt werden, dass die Beschreibung (4) in gewissem Sinne erschöpfend ist. Ein Beispiel für ein derartiges Problem ist das Maximum-Weight-Matching-Scheduling mit $\mu_t(q_t) = q_t$. Allerdings ist im Allgemeinen nicht klar, welche Abbildung $\mu_t(q_t)$ zur Stabilität des Gesamtsystems führt. Ein interessantes Teilproblem ergibt sich im informationstheoretischen Kontext, wobei $C_t(p)$ ein so genanntes Polymatroid ist. Bereits in diesem Fall ist die Lösung von (4) ein nicht triviales, kombinatorisches Problem.

In praktischen Systemen kann die Kapazitätsregion $C_t(p)$ dargestellt werden als die Vereinigung der realisierbaren Raten $r_{t,i}$, $i = 1, 2, \dots$. Damit ergibt sich ein diskretes Optimierungsproblem:

$$\begin{aligned} \max \quad & \mu(q_t)^T r_t \\ \text{s.t.} \quad & r_t \in \bigcup_i \{r_{t,i}\}, \\ & r_t^{(m)} \geq r_{min}^{(m)}, \quad \forall m \end{aligned} \quad (5)$$

Die geforderten minimalen Raten $r_{min} \in \mathbb{R}_+^M$ gewährleisten Mindestgüten für die Dienste. Auch dieses Problem ist für viele praktische Anwendungsfälle im Wesentlichen ungelöst.

Abb. 6: Optimierung von Mobilfunksystemen durch Nutzung von Crosslayer-Information

Um Breitbanddienste mit hohen Datenraten und individuellen Quality-of-Service- (QoS-) Anforderungen zu unterstützen, müssen die Funkressourcen in drahtlosen Kommunikationssystemen dynamisch zugewiesen werden. Geschieht diese Vergabe anhand der Systeminformationen von mehreren funktional getrennten Schichten [7], spricht man auch von einem Crosslayer-Problem.

Crosslayer-Probleme sind vielfältiger Bestandteil der aktuellen Mobilfunkforschung; insbesondere in Verbindung mit Orthogonal-Frequency-Division-Multiplexing- (OFDM-) Systemen bei denen eine besonders „feine“ Auflösung der Ressourcen Zeit, Frequenz, Raum möglich ist. Von besonderer Bedeutung sind speziell Ressourcenallokationsverfahren in der Media-Access-Control- (MAC-) Schicht, auch Scheduling-Verfahren genannt. Hierbei werden Leis-

tungsbudgets, Kanalzustand, Dienstanforderungen und minimale Datenraten in einem gemeinsamen Problemkontext betrachtet. Zu beachten ist, dass Datenpakete der jeweiligen Nutzer nicht kontinuierlich, sondern zufällig über die Zeit verteilt eintreffen und in Puffern zwischengespeichert werden. Dies kann die Stabilität des Gesamtsystems beeinträchtigen. Verschiedene Kriterien können zur Definition der Stabilität herangezogen werden (z. B. Rekurrenz beim Markovschen Modell), und beeinflussen das Systemverhalten entsprechend [36].

Im Allgemeinen kann das Scheduling-Problem als ein Utility-Optimierungsproblem dargestellt werden. Häufig verwendete Utility-Begriffe sind z. B. Systemdurchsatz und Reaktionszeit [37, 39]. Die Systemstabilität muss dann implizit berücksichtigt werden.

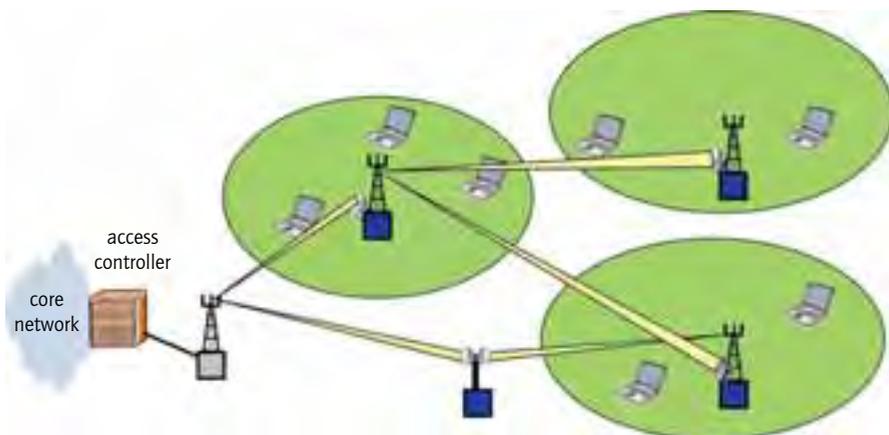
Im nächsten Abschnitt wird gezeigt, dass der Ansatz der Optimierung einer Utility-Funktion ebenfalls erfolgreich bei der Optimierung von Meshed Networks und Ad-hoc-Systemen eingesetzt werden kann.

3.2 STEUERUNG VON SENDELEISTUNGEN

Die Steuerung von Sendeleistungen ist ein Mechanismus zur Ressourcenallokation und zum Interferenzmanagement in Mobilfunknetzen. Utility-basierte Strategien zur Leistungskontrolle zielen darauf ab, Sendeleistungen von Basisstationen oder mobilen Kommunikationsgeräten so zu koordinieren, dass eine Netz-Nutzenfunktion ihr Maximum über einer gegebenen Menge zulässiger Sendeleistungen erreicht [8, 32, 38].

Abb. 7 zeigt ein vermaschtes Kommunikationsnetz. Das Netz besteht aus mehreren Basisstationen. Einige Basisstationen sind nicht direkt an die Festnetzinfrastruktur angeschlossen und bedienen sich anderer Basisstationen als Relaisstationen. Jede Basisstation verfügt über ein Mehrantennensystem, dessen Richtcharakteristik über den sogenannten Sende- bzw. Empfangsbeamformer gesteuert wird (elektronisches Steuern der Antenne in eine oder mehrere Vorzugsrichtungen). Die Sendeleistungen der Basisstationen sollen unter Verwendung eines verteilten Algorithmus ermittelt werden. Die optimale Leistungsallokation hängt von den Kanalzuständen ab. Ändern sich die Kanäle, so müssen die Sendeleistungen angepasst werden. Das gilt auch für Sende- und Empfangsbeamformer, die jedoch zusätzlich durch die Sendeleistungen beeinflusst werden können. Auf welche Weise die Leistungskontrolle mittels Utility-Optimierung erfolgen kann ist in Abb. 8 erläutert.

Abb. 7: Drahtlose Kommunikation mit Relaisstationen



Die Durchsatzmaximierung stellt hiervon einen wichtigen Spezialfall dar, bei dem die Netz-Nutzenfunktion eine gewichtete Summe der Datenraten der einzelnen Nutzer ist. Durchsatzoptimale Strategien können allerdings sehr unfair sein, da die Möglichkeit besteht, dass sie, in Abhängigkeit von Kanalzuständen, manchen Nutzern den Zugang zum Mobilfunkkanal über eine relativ lange Zeit verweigern. Weil das wiederum große Reaktionszeiten für bestimmte Ende-zu-Ende-Verbindungen nach sich ziehen kann, wer-

Abb. 8: Die Utility-basierte Leistungskontrolle für Mobilfunknetze

Angenommen, es gibt $K > 1$ Nutzer und $\mathbf{p} = (p_1, \dots, p_K) \in P \subset \mathbb{R}_+^K$ ist der Leistungsvektor, dessen Koordinaten die Sendeleistungen der einzelnen Nutzer beschreibt. Die Menge P enthält alle zulässigen Leistungsvektoren. Die Utility-basierte Leistungskontrolle läßt sich als Minimierungsproblem formulieren [8, 32]:

$$\mathbf{p}^* = \arg \min_{\mathbf{p} \in P} F(\mathbf{p}), \quad (6)$$

wobei $F : \mathbb{R}_+^K \rightarrow \mathbb{R}$ eine (negative) Netz-Nutzenfunktion bezeichnet und durch

$$F(\mathbf{p}) := \sum_{1 \leq k \leq K} w_k \psi(\text{SNR}_k(\mathbf{p})) \quad (7)$$

gegeben ist. Hier ist $\mathbf{w} = (w_1, \dots, w_K)$ ein gegebener positiver Gewichtsvektor und $\psi(x), x \geq 0$, eine (negative) Nutzenfunktion. Die Funktion SNR_k bezeichnet das Signal-zu-Interferenz-Verhältnis des k -ten Nutzers, so dass $\text{SNR}_k(\mathbf{p}) := p_k / I_k(\mathbf{p})$. Dabei ist $I_k : \mathbb{R}_+^K \rightarrow \mathbb{R}_+$ eine Interferenzfunktion und der spezielle Wert $I_k(\mathbf{p})$ somit die effektive Interferenzleistung (inklusive der Rauschleistung) am Ausgang des k -ten Empfängers, wenn die Sendeleistungen durch \mathbf{p} festgelegt sind. Es ist hier zu beachten, dass im Allgemeinen jedes SNR von allen Sendeleistungen beeinflusst wird. In Abb. 9 findet der Leser axiomatische Definitionen A1-3 allgemeiner Interferenzfunktionen. Für allgemeine Interferenzfunktionen, die lediglich A1-3 erfüllen, ist wenig über die Lösung \mathbf{p}^* des Problems (6) bekannt. Da eine geschlossene Lösung nur in wenigen trivialen Fällen möglich ist, kommen vor allem rekursive Algorithmen in Frage. Wegen der Ressourcenknappheit in drahtlosen Netzen, sollen derartige Algorithmen globale Konvergenz aufweisen und eine effiziente verteilte Implementierung ermöglichen. Die erste Bedingung ist erfüllt, wenn jedes lokale Minimum des Problems (6) zugleich auch globa-

les Minimum ist. Die wichtige Frage, nämlich für welche Funktionen ψ in (7) diese Forderung erfüllt ist, konnte bisher nur für eine Klasse von Interferenzfunktionen zufriedenstellend beantwortet werden. Alle lokalen Minima des Problems (6) sind zugleich global, wenn $\psi(e^x)$ für reelle x konkav und $I_k(e^{\mathbf{s}})$ mit $\mathbf{s} = \log(\mathbf{p}), \mathbf{p} > \mathbf{0}$, eine log-konvexe Funktion des logarithmischen Leistungsvektors $\mathbf{s} \in \mathbb{R}^K$ ist [32]. In dem vermaschten Netz mit Basisstationen ist die Log-Konvexitätsforderung erfüllt, wenn die Empfangsbeamformer für gegebene Kanalrealisierungen beliebig aber fest sind. Im Falle des optimalen Empfangsbeamformer ist die Interferenzfunktion nicht log-konvex, sondern konkav.

In [32] werden rekursive Algorithmen zur Leistungskontrolle präsentiert und analysiert. Für die Anwendungen in Mobilfunknetzen kommen sowohl einfache Gradienten-Algorithmen als auch Primal-Dual-Algorithmen zur Suche nach stationären Punkten einer zugehörigen Lagrange-Funktion in Frage. Im zweiten Fall werden neben den klassischen (linearen) Lagrange-Funktionen auch modifizierte nicht-lineare Lagrange-Funktionen betrachtet, um die Konvergenz zu lokalen oder globalen Minima zu beschleunigen. Ein wichtige Forderung hier ist, dass sich Algorithmen effizient in verteilten Netzen implementieren lassen. Im Anwendungsbeispiel in der Abb. 7 bedeutet diese Forderung, dass jede Basisstation ihre Iterationsschritte autonom anhand lokaler Messungen und Variablen berechnen kann. Das in [32] präsentierte und auf dem sogenannten adjungierten Netz basierende Verfahren erlaubt jedem Nutzer eine sehr effiziente Schätzung der Suchrichtung unter Verwendung von lokalen Messgrößen wie zum Beispiel dem SNR. Dieses Verfahren kann in Verbindung mit Gradienten- und Primal-Dual-Algorithmen verwendet werden.

den andere Netz-Nutzenfunktionen berücksichtigt, die einen besseren Kompromiss zwischen Fairness und Durchsatzperformanz gewährleisten [32].

3.3 AXIOMATIK FÜR INTERFERENZBESCHRÄNKTE FUNKSYSTEME

Bei vielen Ansätzen der Optimierung von drahtlosen Kommunikationssystemen geht es darum, die durch die Ausbreitung der Funkwellen verursachte Interferenz zwischen den Teilnehmern zu kontrollieren. Die Art der Interferenz, die ein einzelner Teilnehmer erfährt, hängt sowohl von dem von ihm genutzten Empfänger als auch von den Sendestrategien anderer Teilnehmer ab. Aufgrund dieser Vielfalt ergibt sich eine auf den ersten Blick nicht zu beherrschende Anzahl an möglichen Interferenzbedingungen.

Die axiomatische Theorie erlaubt eine elegante Lösung dieses Problems und hilft, die wesentlichen Eigenschaften der Interferenz zu verstehen (siehe hierzu Abb. 9). Die Theorie der Interferenzfunktionen erweist sich auch als leistungsstarke Methode, mit deren Hilfe Ressourcenallokationsprobleme gelöst werden können. Es können Techniken zur Leistungsregelung und zur Empfängeroptimierung für komplizierte Kommunikationsszenarien wie z. B. Ad-hoc-Systeme entwickelt werden.

Die Klasse der Interferenzfunktionen besitzt eine interessante algebraische Struktur. Bildet man z. B. das Minimum über eine Menge von Interferenzfunktionen, so erhält man wiederum eine Interferenzfunktion. Die Dienste, die ein Mehrteilnehmersystem unterstützen kann, lassen sich durch Interferenzfunktionen vollständig beschreiben, die die mathematischen Eigenschaften von allgemeinen Interferenzfunktionen besitzen.

Weiterhin können bestimmte Utility-Optimierungsaufgaben bei der Ressourcenallokation auf die Optimierung von Interferenzfunktionen zurückgeführt werden. Ein Beispiel ist die Nash-Bargaining-Lösung [22], bei der das Produkt der Utilities der einzelnen Teilnehmer maximiert wird. Diese Produktfunktion ist eine elementare mathematische Interferenzfunktion, die allerdings keine ursprünglich physikalische Interpretation erlaubt. Die auf der Basis der Nash-Bargaining-Lösung beruhende Ressourcenallokationsstrategie ist äquivalent zu der in der Nachrichtentechnik entwickelten proportional fairen Ressourcenallokationsstrategie für drahtlose Kommunikationssysteme. Eine wichtige Leistung von J. Nash (Nobelpreis für Ökonomie, 1994) bestand darin, seine Lösung axiomatisch zu charakterisieren [22]. Dieser Ansatz dürfte auch für die Ressourcenallokation von drahtlosen Kommunikationssystemen in Zukunft an Bedeutung gewinnen. Für weitere Diskussionen sei auf Abschnitt 5.4 verwiesen.

3.4 KAPAZITÄTSPLANUNG FÜR UMTS-FUNKNETZE

Die Abdeckungsfläche von UMTS-Netzen hängt von der Auslastung der Funkzellen ab. (Im Gegensatz dazu können bei GSM diese beiden Größen noch in zwei getrennten Schritten nacheinander bestimmt werden.) Die Abhängigkeit ergibt sich aus lastabhän-

Abb. 9: Axiomatische Theorie der Interferenzfunktionen

Wie in Abb. 8 bezeichnet $\mathbf{p} = (p_1, \dots, p_K) \in \mathbb{R}_+^K$ den Leistungsvektor, dessen Komponenten gleich den Sendeleistungen der Nutzer sind. In der Mobilfunkkommunikation wird angenommen, dass jede Interferenzfunktion $I : \mathbb{R}_+^K \rightarrow \mathbb{R}, K \geq 1$, die folgenden Bedingungen (Axiome) erfüllt:

- A1 $\mathbf{p} > \mathbf{0} \implies I(\mathbf{p}) > 0$
(Positivität)
- A2 $\mathbf{p} > \mathbf{0}, \mu > 1 \implies I(\mu\mathbf{p}) < \mu I(\mathbf{p})$
(Skalierbarkeit)
- A3 $\mathbf{p}^{(1)} \leq \mathbf{p}^{(2)} \implies I(\mathbf{p}^{(1)}) \leq I(\mathbf{p}^{(2)})$
(Monotonie)

Es kann gezeigt werden [30,32], dass A1-3 die Stetigkeit der Interferenzfunktion impliziert. Die Funktionen aus dem Beispiel in Abb. 8 erfüllen die Axiome A1-3, wenn die Sendebeamformer unabhängig vom Leistungsvektor gewählt werden und damit, bei Kenntnis der instantanen Kanalrealisierungen, fest sind. Dagegen können die Empfangsbeamformer sehr wohl von den Sendeleistungen abhängen. Der hier interessanteste Fall ist der optimale Empfangsbeamformer, der das zugehörige SNR für einen festen Leistungsvektor maximiert, ohne die anderen SNR-Werte zu beeinflussen. In diesem Spezialfall ist die Interferenzfunktion eine stetige konkave Funktion.

Die aus der Sicht heutiger Mobilfunknetze wichtigsten Interferenzfunktionen sind lineare Interferenzfunktionen, welche die folgende Form besitzen:

$$\tilde{I}(\mathbf{p}) = \sum_{l \in \mathcal{K}} a_l p_l + z, \quad (8)$$

wobei $\mathcal{K} = \{1, \dots, K\}$, $z > 0$ eine zur Rauschvarianz proportionale Konstante ist und $a_l \geq 0, l \in \mathcal{K}$, ein vom Kanal abhängiger Dämpfungsfaktor für die Sendeleistung des l -ten Störers. Folglich ist die unter diesem Modell resultierende Interferenzleistung am Ausgang eines jeden Empfängers eine lineare Kombination aller Sendeleistungen plus eine rauschabhängige Konstante. Für den Entwurf von Algorithmen zur Leistungskontrolle spielt die folgende Eigenschaft der Funktion \tilde{I} eine entscheidende Rolle (siehe auch Abb. 8): $\tilde{I}(e^{\mathbf{s}})$ ist eine log-konvexe Funktion des logarithmischen Leistungsvektors $\mathbf{s} = \log(\mathbf{p}), \mathbf{p} > \mathbf{0}$ [30,32].

Ein weitere wichtige Klasse der Interferenzfunktionen bekommt man, wenn man bedenkt, dass die Konstanten $a_l = a_l(u), l \in \mathcal{K}$, und $z = z(u)$ in (8) im Allge-

meinen durch die Empfängerstruktur beeinflusst werden. Hier wird die Empfängerstruktur mit Hilfe einer (abstrakten) Variable $u \in U$ dargestellt. Diese Variable gehört zu einer kompakten Menge U , die bestimmte Beschränkungen bezüglich des Empfängerdesigns widerspiegelt. Mit diesen Definitionen nehmen die Interferenzfunktionen unter optimalen Empfängern die folgende Gestalt an:

$$\tilde{I}(\mathbf{p}) = \min_{u \in U} \left(\sum_{l \in \mathcal{K}} a_l(u) p_l + z(u) \right) \quad (9)$$

Im Gegensatz zum vorherigen Beispiel ist diese Funktion nicht linear. Ein weiterer wichtiger Unterschied ergibt sich aus der Tatsache, dass $\tilde{I}(e^{\mathbf{s}})$ keine log-konvexe Funktion des logarithmischen Leistungsvektors $\mathbf{s} \in \mathbb{R}^K$ ist. Vielmehr ist diese Funktion konkav, was aus den Eigenschaften des Minimum-Operators folgt.

Die Funktionen \tilde{I} und \bar{I} erfüllen die Axiome (A1)–(A3), so dass sie Interferenzfunktionen im Sinne der axiomatischen Definition sind. Manchmal ist es jedoch von Vorteil, das Rauschen beim Entwurf von Strategien für Leistungskontrolle oder, allgemeiner, Ressourcenallokation zu vernachlässigen. Das kann zum Beispiel der Fall sein, wenn die Rauschvarianz unbekannt oder im Vergleich zur gesamten Interferenzleistung sehr klein ist. Aus diesem Grund werden in [30] Interferenzfunktionen betrachtet, welche die folgenden Axiome erfüllen:

- A1' $\mathbf{p} > \mathbf{0} \implies I(\mathbf{p}) > 0$
(Positivität)
- A2' $\mathbf{p} > \mathbf{0}, \mu > 0 \implies I(\mu\mathbf{p}) = \mu I(\mathbf{p})$
(Homogenität)
- A3' $\mathbf{p}^{(1)} \leq \mathbf{p}^{(2)} \implies I(\mathbf{p}^{(1)}) \leq I(\mathbf{p}^{(2)})$
(Monotonie)

Im Vergleich zu (A1)–(A3) wird also nur die Skalierbarkeit (A2) durch die Homogenität (A2') ersetzt. Ein weiterer Vorteil der Axiome (A1')–(A3') ist, dass sie Funktionen erfassen, welche die Charakterisierung aller in einem Mehrnutzer-Mobilfunknetz unterstützbaren Quality-of-Service-Werte ermöglichen [32]. Somit sind diese Funktionen im Sinne der Axiome (A1'–A3') Interferenzfunktionen.

Auf den ersten Blick scheinen die beiden axiomatischen Definitionen nicht vereinbar zu sein. Sie lassen sich jedoch ineinander überführen [32].

gigen Störungen zwischen Funkverbindungen und daraus, dass das Netz die Sendeleistung anhand der Qualität des empfangenen Signals regelt.

Traditionell werden daher zur Bewertung von UMTS-Netzen in der Planung Simulationen eingesetzt, die für zufällige Realisierungen von Nutzeranforderungen (Position, Mobilfunkdienst) jeweils die Auslastung der einzelnen Zellen ermitteln. Die Realisierungen werden gemäß einer vorgegebenen Verteilung der Nutzungsintensität gezogen. Aussagen über die Netzqualität lassen sich nach Analyse einer statistisch signifikanten Anzahl von Bedarfsrealisierungen und dem jeweils ermittelten Netzzustand machen. Die Simulationen sind (zeit-)aufwändig. Sie geben zudem kaum Einblick, welche Änderungen der Netzkonfiguration zu einer Verbesserung der Netzqualität führen könnten.

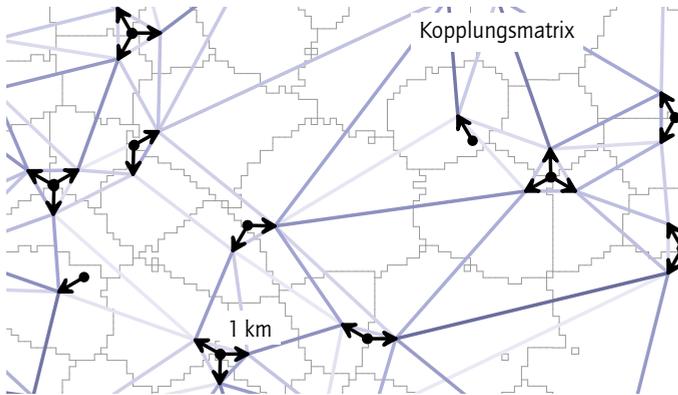
Eine geschlossene Beschreibung des Verhaltens komplexer Systeme stellt eine Herausforderung dar, ist jedoch ein wichtiger Beitrag zum Verständnis des Systems und bildet eine wesentliche Grundlage für die Planung kosteneffizienter Netze [2]. In Abb. 10 ist ein geschlossenes Systemmodell dargestellt, das zum einen eine sekundenschnelle Bewertung großer Netzausschnitte mit hunderten Zellen erlaubt. Zum anderen gibt die darin enthaltene Kopplungsmatrix Aufschluß über die Wirkung von Störungen zwischen den Zellen. Die Matrix bildet daher die Basis mehrerer mathematischer Optimierungsmodelle und automatischer Verfahren für die Netzplanung [18, 35]. Das zugrundeliegende Systemmodell muss im Laufe der Entwicklung der Systemtechnik an die veränderten Systemeigenschaften angepasst werden.

Praktisch sind automatische Netzoptimierungsverfahren von großer Bedeutung. Für einen Netzausschnitt in der Schweiz wurde das Einsparpotenzial bei den Infrastrukturkosten mit 30 % ermittelt, für den Fall, dass die Netzplanung konsequent unter Einsatz von Optimierungswerkzeugen erfolgt [12].

3.5 HARDWAREENTWICKLUNG

Die rasante Entwicklung der drahtlosen Kommunikation wurde ebenfalls durch Innovation in der Halbleiterindustrie ermöglicht. Die dort erzielten technologischen Fortschritte sind die Basis dessen, was Mooresches Gesetz genannt wird, nämlich die Verdopplung der Rechenleistung alle 18 Monate bei gleichbleibenden Kosten [26]. Mittlerweile werden Hardwareplattformen für die Empfängerimplementierung genutzt, die eine Rechenleistung von mehreren Tera-Operationen pro Sekunde erzielen. Für die nächsten Jahre dürfte sich die rasante Entwicklung im Hardwarebereich fortsetzen. Schon heute werden leistungsstarke mathematische Techniken genutzt, um Algorithmen der Nachrichtenübertragung wie z. B. optimale Kanalentzerrer oder optimale Mehrnutzerempfänger auf die leistungsstarken Hardwareplattformen abzubilden. Neben Fragestellungen für das optimale Scheduling von nachrichtentechnischen Algorithmen ist jedoch die Mathematik ebenfalls für Innovationen in der Halbleiterindustrie von zentraler Bedeutung. So werden bei der fortschreitenden Miniaturisierung in der Halbleiterindustrie die quantenphysikalischen Effekte noch stärker an Bedeutung gewinnen. Für das Schaltungsdesign

Abb. 10: Ein Systemmodell zur Leistungsbewertung von UMTS-Funknetzen unter Last



Die Kopplung der Störung zwischen den Zellen eines UMTS-Funknetzes kann mathematisch durch lineare Komplementaritätssysteme beschrieben werden. Diese erlauben die Berechnung von Zellauslastung und ggf. -überlastung.

Die ständige Anpassung der Sendeleistungen für die Verbindungen zwischen den Basisstationen und ihren Nutzern ist ein Systemmerkmal von UMTS. Am Beispiel des Downlink wird dargestellt, wie das dynamische Netzverhalten geschlossen beschrieben werden kann. Das Kopplungssystem der Zellen (z. B. [25]) beschreibt das Gleichgewicht, das sich zwischen den Sendeleistungen der einzelnen Zellen einstellt. Dieses Gleichgewicht wird maßgeblich durch den zu unterstützenden Verkehr sowie die Kopplung der Sendeleistungen zwischen den Zellen durch gegenseitige Störungen bestimmt.

Den Ausgangspunkt für die Modellierung bilden die SNR-Ungleichungen, die für jede Verbindung benennen, welches Signal-zu-Rauschleistungsverhältnis mindestens erzielt werden muss. Die Leistungssteuerung sorgt dafür, dass dieses Mindestmaß möglichst knapp erreicht wird. Mit der Annahme einer idealen Leistungssteuerung (*perfect power control*) werden aus den Ungleichungen jeweils Gleichungen. In einem zweiten Schritt können die Gleichungen je Nutzer so zusammengefaßt werden, dass sich ein Gleichungssystem ergibt, in dem nur noch die Gesamtsendeleistungen je Zelle auftreten (woraus sich die Sendeleistungen der einzelnen Verbindungen ableiten lassen):

$$p_i = C_{ii}p_i + \sum_{j \neq i} C_{ij}p_j + p_i^{(n)} + p_i^{(c)} \quad (10)$$

Diese Gleichung drückt die Abhängigkeit der Übertragungsleistung in Zelle i von den Leistungen der anderen Zellen $j \neq i$ aus. Die Nutzer bestimmen die Einträge der Kopplungsmatrix C durch ihre geographische Anordnung und den genutzten Dienste.

Die obige Grafik illustriert einen Ausschnitt aus einer solchen Matrix. Die farbliche Intensität der Verbindungslinien nimmt mit der Stärke der Kopplung zwischen zwei Zellen (dargestellt durch die Pfeile als Ausrichtung der Sektoren und ihrer Zellflächen) zu.

Der Wert $p_i^{(n)}$ gibt den Leistungsanteil je Zelle an, der zur Bedienung der Nutzer ausreichen würde, wenn es keine Störung gäbe, und $p_i^{(c)}$ bezeichnet die konstante Sendeleistung, die auf das sogenannte Pilotsignal u. ä. verwendet wird. Das Pilotsignal ist eine Art „Leuchtturm“. Mit der Zelle kann nur eine Kommunikation stattfinden, sofern der Nutzer dieses Leuchtturm sieht.

Das Kopplungssystem aller Gleichungen (10) kann mit Methoden der Numerik sehr schnell gelöst werden und erlaubt eine (hinreichend genaue) Ermittlung der für die Versorgung der Nutzer erforderlichen Sendeleistungen je Zelle und der Systemauslastung. Diese Berechnung ist jedoch nur für den Fall aussagekräftig, dass das System nicht mit der Bedienung aller Nutzer überlastet wäre.

In der Praxis reagiert ein UMTS-Netz auf Überlastung durch teilweises oder ganzes Ablehnen von Bedarfsanforderungen. Dieses Verhalten kann durch eine Modellerweiterung als idealisierte Laststeuerung (*perfect load control*) berücksichtigt werden [14, 18]. Mathematisch geht das lineare Gleichungssystem dabei in ein Komplementaritätssystem über, das sich mit vergleichbarem Aufwand lösen lässt.

müssen dann beispielsweise Resultate der Quanteninformationstheorie und der Operatoralgebra verwendet werden.

4 ZUSAMMENARBEIT STÄRKEN

Die Entwicklung in der Mobilkommunikation bringt immer bessere Systeme hervor. Was auf der Seite der Anwendung zu effizienteren Kommunikationsnetzen führt, bedeutet auf der Seite der Forschung und Entwicklung eine rapide steigende Komplexität in den Gesamtsystemen. Beherrschen lässt sich diese Komplexität nur dann, wenn die Mittel zur Beschreibung und Analyse entsprechend mitentwickelt werden. Es gibt keine andere Möglichkeit, als den Einsatz von Mathematik als Mittel der Abstraktion weiter zu forcieren!

Die Abschnitte 2 und 3 gaben einen Eindruck von den vielfältigen Anwendungen der Mathematik im Bereich des Mobilfunks. Diese kommen einerseits auf der Basis der Ausbildung eines Ingenieurs der Nachrichtentechnik zustande, die traditionell mathematische Konzepte und Methoden vermittelt. Der Ausbildungsschwerpunkt liegt dort jedoch häufig noch in der Analysis. Diese Beschränkung muss aufgehoben werden. Die Grundlagen der mathematischen Gebiete Wahrscheinlichkeits- und Graphentheorie, Kombinatorik und Optimierung sind heute wichtiges Rüstzeug für angehende Nachrichtentechniker.

Die genannten Fortschritte basieren andererseits ganz wesentlich auf der Zusammenarbeit von Nachrichtentechnikern und Mathematikern. Diese ist auf breiter Front auszubauen. In Abschnitt 5 sind wichtige Grundlagenfragen zusammengestellt für deren Erforschung die Förderung von Verbundprojekten durch die DFG und/oder das BMBF beste Voraussetzungen schaffen kann.

Im Weiteren werden Beispiele für Entwicklungen der jüngeren Zeit gegeben, deren Verstärkung gleichfalls empfohlen wird.

4.1 INTERDISZIPLINÄRE ZUSAMMENARBEIT

Zunehmend wichtig ist die interdisziplinäre Zusammenarbeit von Mathematikern und Ingenieuren. Dies kann mit der gemeinsamen Ausbildung von Studenten beginnen, wie sie zum Beispiel als Teil des Studienganges Techno-Mathematik erfolgt, und sich in interdisziplinären Graduiertenkollegs fortsetzen. Solche Graduiertenkollegs sind von besonderer Bedeutung, da hier für Studenten mit nachhaltigem wissenschaftlichen Interesse wesentliche Grundlagen für Forscherkarrieren oder für den Transfer von Methoden und Erkenntnissen in die Industrie gelegt werden. Eine breitere Anwendung dieses Instrumentes als Teil der akademischen Ausbildung ist sehr wünschenswert.

Jenseits der Graduiertenschule gewinnt auch die interdisziplinäre Grundlagenforschung zwischen Mathematik und Nachrichtentechnik an Bedeutung. Die DFG hat mit dem Forschungszentrum MATHEON, in Berlin und dem Exzellenzcluster Ultra High Speed

Mobile Information and Communication an der RWTH Aachen wichtige Impulse auch im Bereich der Mobilkommunikation gegeben.

4.2 INTERNATIONALE ZUSAMMENARBEIT

Auf der Ebene der Forschung und Entwicklung wird interdisziplinäre Zusammenarbeit mehr und mehr durch die Europäische Union gefördert. Ein Beispiel dafür ist die European Cooperation in the Field of Scientific and Technical Research, kurz COST [27]. Unter diesem Dach wurde und wird z. B. in der Aufeinanderfolge der COST-Actions 259, 273 und aktuell 2100 von Nachrichtentechnikern und Mathematikern gemeinsam an Lösungen für eine effektive Netzplanung der jeweils aktuellen Technikgenerationen gearbeitet. Im Rahmen einer Action finden für die Dauer von einigen Jahren (überlicherweise 3–4 Jahre) jeweils drei bis vier Treffen pro Jahr statt. Die Teilnehmer sind Experten aus Wissenschaft und Industrie. Die regelmäßigen Treffen helfen über anfängliche „Sprachbarrieren“ zwischen den Disziplinen hinweg und schaffen eine gemeinsame Arbeitsgrundlage. Sie erleichtern den Mathematikern den Zugang zur Forschung in der Mobilkommunikation und führen bei den Nachrichtentechnikern neue mathematische Methoden ein. Das Instrument hat sich bewährt.

Ein weiteres Beispiel sind klassische Forschungs- und Entwicklungsprojekte in den Rahmenprogrammen der Europäischen Union zur Forschungsförderung. Im 7. Rahmenprogramm wird z. B. aktuell ein Projekt gefördert, in dem Ingenieure und Mathematiker gemeinsam Mechanismen zur Selbstorganisation künftiger Mobilfunknetze erforschen [9]. Diese Mechanismen sollen Netze nach dem LTE-Standard in der Praxis deutlich besser handhabbar machen als dies für GSM- und UMTS-Netze der Fall ist. Diese Fragestellung ist von hohem wirtschaftlichen Interesse [24]. Die damit verbundenen Herausforderungen sind auch wissenschaftlich hoch interessant und umfassen Fragestellungen der Spieltheorie, des Mechanism Design und der (Online-)Optimierung. Die erfolgreiche interdisziplinäre Zusammenarbeit von Ingenieuren und Mathematikern im Rahmen von EU-Programmen ist ein gutes Argument, dieses Instrument auch in Zukunft stark zu nutzen.

4.3 TRANSFER

Die sehr große Bedeutung technischer Innovation für die deutsche Wirtschaft ist bekannt. Der Wissenstransfer aus der Forschung in die Praxis ist eine wesentliche Quelle von Innovation.

Neben dem Wissenstransfer in der Form von Personaltransfer spielt die gemeinsame Projektarbeit von wissenschaftlichen Einrichtungen und industriellen Partnern eine wichtige Rolle. Diese Zusammenarbeit wird traditionell durch das BMBF gefördert. Da beim Wissenstransfer auch zunehmend kleine und mittlere Unternehmen eine bedeutende Rolle spielen, sind die Anstrengungen des BMBF wichtig, diese Unternehmen in

ihre Förderung aufzunehmen. In Anbetracht der im Vergleich zur (Groß-)Industrie geringeren Kapitalausstattung von kleinen Unternehmen ist eine zeitnahe Auszahlung von Förderleistungen erforderlich. Hierauf stellen sich BMBF und die EU gerade ein. Diese Umstellung ist wichtig.

Die Förderung von technologie-basierten Unternehmensgründungen aus Universitäten und Forschungseinrichtungen heraus wird seit Jahren verbessert. Es besteht jedoch nach wie vor Handlungsbedarf, damit innovative Techniken häufiger und erfolgreicher den Weg in die Wirtschaft finden. So ist z. B. die Wiedereintrittsschwelle für Unternehmensgründer aus der Forschung zu hoch, wenn sie nach ein oder zwei Jahren unternehmerischer Tätigkeit ihre wissenschaftliche Laufbahn fortsetzen wollen. Gerade jungen Forschern fällt unter diesen Umständen die Entscheidung für das unternehmerische Risiko schwer.

4.4 FORSCHUNG VON NACHWUCHSTALENTEN FÖRDERN

Neue Konzepte werden häufig von jungen Wissenschaftlern erdacht und erforscht. Daher ist die Förderung einer eigenverantwortlichen Grundlagenforschung dieser Altersgruppe von besonderer Bedeutung. Die noch jungen Juniorprofessuren sind ein Instrument, das etablierte Emmy-Noether-Programm der DFG ein weiteres. Beide sollten auch im Bereich der drahtlosen Kommunikationssysteme intensiver eingesetzt werden, um Nachwuchstalente mit einer Art „Venture Capital“ für ihre (universitäre) Forschung auszustatten. Die Unterstützung von Wissenschaftlern durch Drittmitteln aus der Industrie für die industrienaher Forschung und Entwicklung kann zu diesem Zweck nur ergänzend wirken.

5 PERSPEKTIVEN

In diesem Abschnitt werden wichtige potenzielle Entwicklungsrichtungen der Kommunikations- und Informationstheorie diskutiert. Obwohl diese Fragestellungen eine lange Geschichte besitzen, konnten bisher nur kleine Fortschritte bei ihrer Lösung erzielt werden. Eine vollständige Lösung der diskutierten Probleme erlaubt auf der einen Seite, neue Anwendungen für die drahtlose Kommunikation zu erschließen. Auf der anderen Seite lässt sich ein besseres Verständnis der Natur und der gesellschaftlichen Prozesse erreichen.

5.1 INFORMATIONSTHEORIE UND DIGITALE WELT

Die in den Abschnitten 2 und 3 diskutierten Resultate der Informationstheorie haben zweifelsohne einen enormen Einfluss auf die Entwicklung der Mobilkommunikation gehabt. Wichtige Fragestellungen der Informationstheorie sind jedoch noch ungelöst. Hier sei z. B. die Kapazität von zeitkontinuierlichen Kanälen genannt. Eine Weiterentwicklung

der Informationstheorie für allgemeinere Kanalmodelle, also eine methodische Verbindung von Informations- und Warteschlangentheorie sowie von Informations- und Regelungstheorie, wäre von größter Bedeutung. Zahlreiche neue Anwendungen der drahtlosen Kommunikation würden auf dieser Basis möglich.

Die Informationstheorie zeitdiskreter Systeme hat sich in den letzten 60 Jahren gut entwickelt. Sie war einer der Ausgangspunkte der Entwicklung, die zu unserer heutigen „Digitalen Welt“ führte. Das Shannonsche Abtasttheorem bildet die theoretische Grundlage der Digitalisierung von Daten (auch von Sprache und Bildern), der digitalen Signalverarbeitung sowie der digitalen Übertragung. R. P. Feynman (Nobelpreis für Physik, 1965) würdigt die Bedeutung dieses Theorems als Basis des modernen Rechnens in [15] durch „Consideration of such a problem will bring us on to consider the famous Sampling Theorem, another baby of Claude Shannon“. Für die Fortsetzung des Trends zur stärkeren Digitalisierung ist es notwendig, neue Mechanismen für die Signalrekonstruktion zu entwickeln. Als methodische Grundlage muss hierfür in der Grundlagenforschung die Signaltheorie erweitert werden.

5.2 NETZWERKINFORMATIONSTHEORIE

Im Abschnitt 2 wurde die Bedeutung der Informationstheorien des Vielfachzugriffs- und des Verteilkanals für die Entwicklung des zellularen Mobilfunks herausgearbeitet. Es wurde festgestellt, dass die Charakterisierung der optimalen Übertragungsstrategien für diese beiden Kommunikationsszenarien maßgeblich die Standardisierung von HSDPA, HSUPA und LTE beeinflusst hat. Im Allgemeinen sind jedoch zentrale informationstheoretische Fragestellungen für Ad-hoc Systeme, Sensornetze usw. ungelöst. Es fehlt das tiefere Verständnis von Kommunikationssystemen mit vielen gleichzeitig aktiven Teilnehmern. Die Entwicklung dieser Systeme schreitet deshalb bisher vergleichsweise langsam voran. Es kann erwartet werden, dass die Lösung der relevanten informationstheoretischen Fragestellungen ganz wesentlich zum Verständnis dieser komplexen Kommunikationssysteme beiträgt. Das Ziel ist die Entwicklung einer Netzwerkinformationstheorie. Diese wäre von großer praktischer Bedeutung für viele neue Anwendungen in der klassischen drahtlosen Kommunikation, in der Automobilindustrie, im Maschinen- und Anlagenbau usw.

5.3 QUANTENINFORMATIONSTHEORIE

Die in den Abschnitten 5.1. und 5.2. aufgeführten Fragestellungen gehören zum Bereich der klassischen Informationstheorie. In jüngster Zeit hat sich aufgrund der erzielten Fortschritte in der Experimentalphysik die Quanteninformationstheorie entwickelt. Die Basis dieser Theorie bildet die Quantenphysik. Ein zentrales Resultat stellt das Kodierungstheorem für Quantenkanäle dar. Es zeigt in Analogie zur Shannonschen Informationstheorie, dass auch über Quantenkanäle Information übertragen werden kann.

Die Quanteninformationstheorie ist eine sehr junge Disziplin. Für wichtige Mehrteilnehmerkommunikationsszenarien sind die Quantenkapazitätsregionen nicht bekannt. Eine universelle Quantenkodierung, die auch eine robuste Übertragung über Quantenkanäle erlaubt, muss in Zusammenarbeit mit Experimentalphysikern entwickelt werden. Es darf erwartet werden, dass die Quanteninformationstheorie der Schlüssel zum Verständnis und zur Beherrschung von miniaturisierten Systemen darstellt. Insbesondere wird die Quanteninformationstheorie zum Verständnis der Übertragung und der Manipulation von klassischer Information unter Einfluß von quantenmechanischem Rauschen beitragen, ein Szenario welches z. B. bei der On-Chip-Kommunikation in miniaturisierten Chip-Architekturen auftreten wird. An dieser Stelle lässt sich auch beobachten, wie reine bzw. abstrakte Teilgebiete der Mathematik durch neue Anwendungsbereiche zur produktionsrelevanten Mathematik werden. So übernimmt z. B. die Geometrie der Banachräume bzw. die konvexe Geometrie in der Quanteninformationstheorie die Rolle der Kombinatorik in der klassischen Informationstheorie. Die Theorie der Operatoralgebren und die Quantenwahrscheinlichkeitstheorie bieten den geeigneten Rahmen, um die stochastische Struktur der Systeme der Quanteninformationstheorie zu beschreiben und zu untersuchen.

5.4 NEUE ANSÄTZE DER FREQUENZNUTZUNG

Die bereits erwähnten Konzepte zur Einhaltung bestimmter Fairness-Maße bei der Ressourcenallokation bilden den Ausgangspunkt einer Entwicklung an deren Ende völlig neue Konzepte der Frequenznutzung und -vergabe stehen könnten. Es ist unbestritten, dass das Frequenzspektrum eine volkswirtschaftlich wichtige Ressource ist. Die Zuteilung von Frequenzen ist ein sehr komplexes Problem und erfordert eine weltweite Koordination, zum Beispiel während der regelmäßig stattfindenden Weltradiokonferenzen. Vom Gesetzgeber müssen in der Regel eine große Vielzahl unterschiedlicher Standpunkte bei der Verteilung von Frequenzbändern berücksichtigt werden. Nur selten wird die notwendige Diskussion auf einer rationalen Ebene geführt. Durch die mathematischen Disziplinen wie z. B. Social Choice Theory/Social Welfare Theory und Mechanism Design kann eine bessere Durchdringung des Frequenzverteilungsproblems erzielt werden. Innerhalb einer axiomatischen Theorie können Fragestellungen der Fairness, Effizienz, Robustheit, Sicherheit und des gesellschaftlichen Nutzen formalisiert sowie zum Teil auch antagonistische Widersprüche aufgedeckt werden. Als sehr erfolgreiches Beispiel für ein solches Verständnis sei K. J. Arrows „Impossible Result“ (Nobelpreis für Ökonomie, 1972) aus der klassischen Social Choice Theorie genannt [4, 5]. Aufgrund der dargestellten gesellschaftlichen Relevanz ist der Entwurf einer axiomatischen Theorie für die Frequenznutzung von zentraler Bedeutung. Diese Theorie muss auf der einen Seite, die durch die Nachrichtentechnik vorgegebenen sehr komplexen Kommunikationsszenarien erfassen, und auf der anderen Seite eine Formulierung für unterschiedliche Fairness-Konzepte, Konzepte des Gesetzgebers und Mechanismen der Märkte geben.

5.5 SELBSTORGANISIERENDE NETZE

Mobilfunksysteme wie GSM und UMTS werden großflächig als Infrastruktur für mobile Sprach- und Datendienste eingesetzt. Der Aufwand, der mit dem Betrieb eines oder mehrerer solcher Netze verbunden ist, übersteigt mittlerweile den für die Betreiber akzeptablen Rahmen. Daher soll bei der Standardisierung der Nachfolgesysteme, zum Beispiel von LTE, zum einen das System grundsätzlich weniger Konfigurationsmöglichkeiten anbieten. Zum anderen sollen im Netz Mechanismen der Selbstorganisation dazu beitragen, dass manuelle Eingriffe nur noch in deutlich reduziertem Umfang notwendig sind. Die Forschung zu diesem Thema steht erst am Anfang. Die Prinzipien, nach denen eine dezentrale Regelung von sich teilweise überlappenden Parametermengen auf unterschiedlichen Zeitskalen organisiert werden kann, sind bislang unverstanden. Weder liegen mathematisch fundierte Entwurfsprinzipien vor, gemäß derer solche Regelungen entworfen werden können. Noch sind Unmöglichkeitsergebnisse bekannt, die abgrenzen, unter welchen Bedingungen eine stabile Regelung in diesem Anwendungsfeld nicht möglich ist. Antworten auf diese grundlegenden Fragen sind dringend erforderliche Beiträge für den Entwurf sogenannter selbstorganisierender Netze. Denn sie könnten die Mechanismen charakterisieren, mit denen eine hohe Automatisierung des Betriebs zukünftiger Mobilfunksysteme möglich wäre.

6 LITERATUR

- [1] K. I. Aardal, C. P. M. van Hoesel, A. M. C. A. Koster, C. Mannino und A. Sassano. Models and solution techniques for the frequency assignment problem. *4OR: Quarterly Journal of the Belgian, French and Italian Operations Research Societies*, 1(4):261–317, 2003.
- [2] R. Abiri, Z. Neyman, H.-F. Geerdes und A. Eisenblätter. Automatic network design. In M. Nawrocki, H. Aghvami und M. Dohler, Hrsg., *Understanding UMTS Radio Network Modelling, Planning and Automated Optimisation: Theory and Practice*, Kapitel 14. Wiley, 2006.
- [3] M. Aigner und G. Ziegler. *Das Buch der Beweise*. Springer-Verlag Berlin Heidelberg New York, 2002.
- [4] K. J. Arrow. A difficulty in the concept of social welfare. *Journal of Political Economy*, 58(4):328–346, 1950.
- [5] K. J. Arrow. *Social Choice & Individual Values*. Yale University Press, 1963.
- [6] GSM Association. Presseerklärung vom 16.04.2008.

- [7] D. P. Bertsekas und R. Gallager. *Data Networks*. Prentice Hall, 1992.
- [8] M. Chiang. Balancing transport and physical layers in wireless multihop networks: Jointly optimal congestion control and power control. *IEEE Journal on Selected Areas in Communications*, 23(1):104-116, Jan. 2005.
- [9] Socrates Project Consortium. Self-optimisation and self-configuration in wireless networks, 2008.
- [10] L. M. Correia, Hrsg. *Wireless Flexible Personalized Communications - COST 259: European Co-operation in Mobile Radio Research*. John Wiley & Sons, 2001. COST Action 259—Final Report.
- [11] T. M. Cover und J. A. Thomas. *Elements of Information Theory*. Wiley & Sons, 2 edition, 2006.
- [12] S. Dehghan. A new approach. *3GSM Daily 2005*, 1:44, 2005.
- [13] A. Eisenblätter, M. Grötschel und A. M. C. A. Koster. Frequency assignment and ramifications of coloring. *Discussiones Mathematicae Graph Theory*, 22:51-88, 2002.
- [14] A. Eisenblätter, H.-F. Geerdes, T. Koch, A. Martin und R. Wessälly. UMTS radio network evaluation and optimization beyond snapshots. *Math. Methods Oper. Res.*, 63:1-29, March 2006.
- [15] R. P. Feynman. *Feynman Lectures on Computation*. Penguin Book, 1999.
- [16] UMTS Forum. Presseerklärung vom 20.01.2008.
- [17] A. Gamst. Application of graph theoretical methods to GSM radio network planning. In *Proceedings of IEEE International Symposium on Circuits and Systems*, volume 2, pages 942-945, 1991.
- [18] H.-F. Geerdes. *UMTS Radio Network Planning: Mastering Cell Coupling for Capacity Optimization*. Doktorarbeit, Technische Universität Berlin, 2008.
- [19] O₂ Germany. Telefónica investiert 3,5 Milliarden Euro in Deutschland. Presseerklärung vom 11.10.2007, 2007.
- [20] M. Grötschel, L. Lovász und A. Schrijver. Relaxations of vertex packing. *J. of Combinatorial Theory, Series B*, 40:330-343, 1968.

- [21] W. K. Hale. Frequency assignment: Theory and applications. *Proceedings of the IEEE*, 68:1497–1514, 1980.
- [22] H.W. Kuhn und S. Nasar. *The Essential John Nash*. Princeton University Press, 2002.
- [23] L. Lovász. On the Shannon Capacity of a Graph. *IEEE Trans. Information Theory*, 25:1–7, 1979.
- [24] NGMN Ltd. Next generation mobile networks: Beyond HSPA and EVDO. White Paper, 2006.
- [25] L. Mendo und J. M. Hernando. On dimension reduction for the power control problem. *IEEE Trans. Comm.*, 49:243–248, 2001.
- [26] G. E. Moore. Cramming more components onto integrated circuits. *Electronics*, 38(8), 1965.
- [27] European Science Foundation: COST Office. European cooperation in the field of scientific and technical research.
- [28] J. Plehn. Applied Frequency Assignment. In *Proceedings of the IEEE Vehicular Technology Conference*. IEEE, 1994.
- [29] Jahresbericht der Bundesnetzagentur 2007. Bundesnetzagentur für Elektrizität, Gas, Telekommunikation, Post und Eisenbahnen; Bonn, 2008.
- [30] M. Schubert und H. Boche. QoS-based resource allocation and transceiver optimization. *Foundation and Trends in Communications and Information Theory*, 2(6), 2006.
- [31] C. E. Shannon. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27:379–423, 623–656, 1948.
- [32] S. Stanczak, M. Wiczanowski und H. Boche. *Fundamentals of Resource Allocation in Wireless Networks*, volume 3 of *Foundations in Signal Processing, Communications and Networking*. Springer, Berlin, 2008.
- [33] I. Emre Telatar. Capacity of Mult-antenna Gaussian Channels. *European Transactions on Telecommunications*, 10(6):585–596, November 1999.

- [34] D. Tse und P. Viswanath. *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.
- [35] U. Türke. *Efficient Methods for W-CDMA Radio Network Planning and Optimization*. Doktorarbeit, Teubner Verlag, 2007.
- [36] G. Wunder und C. Zhou. Queueing Analysis for the OFDMA Downlink: Throughput Regions, Delay and Exponential Backlog Bounds. *IEEE Transactions on Wireless Communications*, 2008. Im Druck.
- [37] G. Wunder, C. Zhou, S. Kaminski und H. E. Bakker. Throughput maximization under rate requirements for LTE OFDMA Downlink with limited feedback. *EURASIP Journal on Wireless Commun. and Networking, special issue on multicarrier*, November 2007.
- [38] M. Xiao, N. B. Schroff und E. K. P. Chong. A utility-based power control scheme in wireless cellular systems. *IEEE Transactions on Networking*, 11(2):210–221, April 2003.
- [39] C. Zhou und G. Wunder. Throughput-optimal Scheduling with Low Average Delay for Cellular Broadcast Systems. *EURASIP Journal on Advances in Signal Processing, special issue on Cross-Layer Design for the Physical, MAC, and Link Layer in Wireless Systems*, 2008. Im Druck.

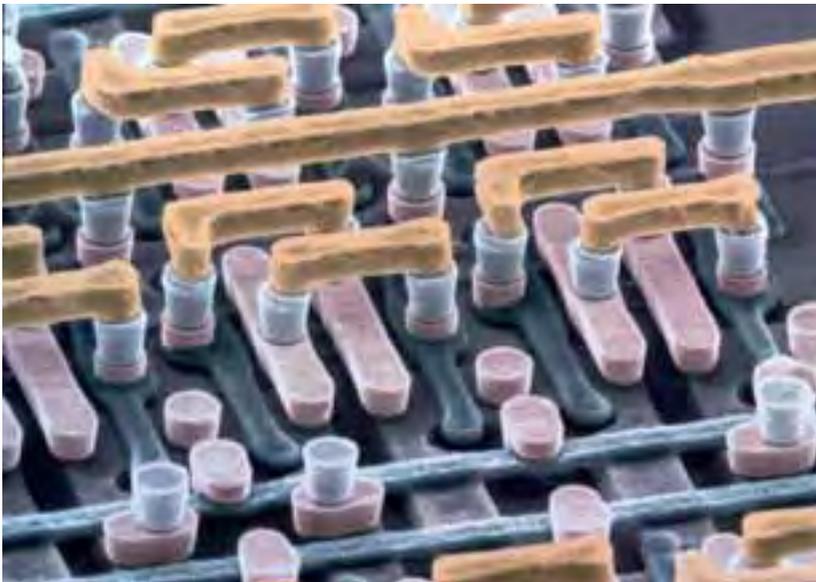
> MATHEMATIK IM CHIP-DESIGN

JÜRGEN KOEHL, BERNHARD KORTE UND JENS VYGEN

1 EXECUTIVE SUMMARY

Chips sind die wohl komplexesten Strukturen, die vom Menschen entworfen und gefertigt wurden. Auf einem kleinen Silizium-Chip von der Größe eines Fingernagels werden heute Milliarden von Transistoren untergebracht, die mit vielen Millionen Verbindungen untereinander verknüpft sind, wobei die Gesamtlänge dieser Netze mehrere Kilometer betragen kann. Abb. 1 zeigt einen kleinen Ausschnitt (~ 1 Milliardstel) eines Chips mit zweilagiger Verdrahtung im Raster-Tunnel-Mikroskop.

Abb. 1: Kleiner Ausschnitt (~ 1 Milliardstel) eines Chips im Raster-Tunnel-Mikroskop



Den Mathematiker erfüllt es mit besonderer Befriedigung, dass diese elektronischen Winzlinge ohne Methoden der Diskreten Mathematik nicht entworfen werden könnten. Es ist ein besonderes Glücksgefühl für einen Mathematiker, wenn er einen neuen mathematischen Satz bewiesen hat und damit das Wissen der Menschheit ein wenig vergrößert hat. Wenn wir einen kleinen Chip in den Händen halten und dabei wissen, dass er nur durch unsere Mathematik so möglich wurde, dann ist das ein zwar anderes, aber nicht minder starkes Glücksgefühl.

Wohl keine andere Industrie hat einen solch rasanten Anstieg in der Komplexität ihrer Produkte, in den technischen Herausforderungen und im Kosten- und Zeitdruck zu meistern wie die Chipindustrie.

- Komplexität: Technologischer Fortschritt macht es möglich, alle zwei Jahre die Anzahl der Transistoren auf einem Chip zu verdoppeln (Moore'sches Gesetz). In ähnlichem Maße wächst auch die Zahl der Bauteile, die auf einen Chip platziert werden müssen, die Zahl der elektrischen Verbindungen, die von Verdrahtungsprogrammen verlegt werden müssen, die Zahl der Logikpfade, die zur Maximierung der Taktfrequenz optimiert werden müssen und der Umfang der Logikbeschreibung, die von Syntheseprogrammen in eine optimierte Schaltung umgewandelt werden muss.
- Technische Herausforderung: Die zur Zeit fortschrittlichste Fertigung erfolgt in der 45 nm Technologie. In dieser Technologie besteht das Gate-Oxid eines Transistors aus nur noch fünf Atomlagen. Trotz eines immensen technischen Aufwands in der Herstellung sind Variationen von $+/-$ einer Atomschicht über und unter dem Gate-Oxid nicht zu vermeiden. Damit variiert die Dicke dieser Schicht um mindestens 20 %, und der Chip muss trotz dieser Variation für jeden einzelnen der Milliarden Transistoren zuverlässig und auf Dauer funktionieren. Gleichzeitig steigt die Taktfrequenz auf bis zu 5 Gigahertz, wobei die einzelnen Taktsignale bis auf wenige Pikosekunden abgeglichen werden müssen. Der Stromverbrauch darf auch bei einem neuen Chip, der wesentlich mehr Logik enthält, auf keinen Fall steigen, weil sonst die Verlustwärme nicht abgeleitet werden kann.
- Kostendruck: Mit jeder neuen Technologiegeneration wird die Funktion von zwei Chips auf einem einzigen implementiert und das bei deutlich höheren technischen Herausforderungen. Trotz der höheren Leistung erwartet der Kunde, dass er seinen neuen Computer zu ähnlichen Kosten wie das Vorgängermodell kaufen kann. Kostensteigerungen sind also nicht durchzusetzen und daher darf die Entwicklung eines Chips der doppelten Größe keinen größeren Entwicklungsaufwand als sein Vorgänger verlangen.

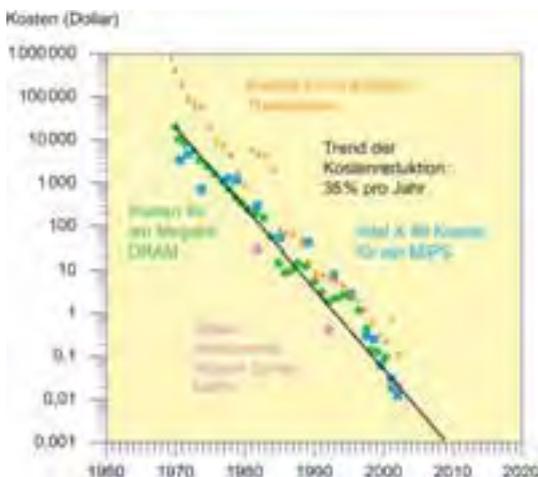
In Abb. 2 ist die Kostenreduktion in der Mikroelektronik in den letzten 60 Jahren anhand verschiedener Leistungsparameter dargestellt. Sie beträgt 35 % pro Jahr. Unseres Wissens kann keine andere Industrie auch nur annähernd solche Kostenreduktionen aufweisen.

- Zeitdruck: Die neue Spielkonsole muss vor dem nächsten Weihnachtsgeschäft auf dem Markt sein. Jedes Jahr erwarten die Kunden neue und schnellere Notebooks und vielseitigere Mobiltelefone. Trotz der exponentiell wachsenden Komplexität muss die Turnaround-Zeit für die Entwicklung eines Chips nicht nur gleich bleiben, sondern nach Möglichkeit deutlich reduziert werden.

Diese Herausforderungen können ohne die ständige Weiterentwicklung der mathematischen Entwurfsmethoden nicht gemeistert werden. Obwohl ein recht guter Stand der mathematischen Durchdringung dieses Anwendungsgebietes erreicht ist, sind ständig neue Ansätze erforderlich, die ein weites Feld für mathematische Forschung mit unmittelbaren wirtschaftlichen Anwendungen sind.

In diesem Artikel berichten wir über einige Ergebnisse einer mehr als 20-jährigen Kooperation zwischen Industrie (IBM) und Wissenschaft (Forschungsinstitut für Diskrete Mathematik der Universität Bonn). Diese Kooperation gehört nach einhelliger Meinung von Fachleuten zu den wohl längsten und auch erfolgreichsten Zusammenarbeiten zwischen Wissenschaft und Wirtschaft. Wir werden uns hier beschränken auf die Anwendung von Methoden der kombinatorischen Optimierung beim physikalischen Entwurf und bei der Optimierung des Zeitverhaltens (Timing) von Chips.

Abb. 2: Kostenreduktion in der Mikroelektronik



2 ERFOLGSGESCHICHTEN

Im Jahr 1985 sind wir zufällig auf einer Tagung mit einem Designingenieur aus der Mikroelektronik zusammengetroffen, der uns seine Probleme beim Entwurf von Chips schilderte. Obwohl um mehrere Zehnerpotenzen geringer als heute, war auch damals schon die Komplexität beim Entwurf so groß, dass manuelle Methoden oder einfache Heuristiken nicht mehr ausreichten, um eine hohe Entwurfsqualität zu erzielen. Wir realisierten sofort, dass Chipdesign ein Anwendungsgebiet der kombinatorischen Optimierung par excellence ist.

Nach einer kurzen Analyse schätzten wir, dass Methoden der kombinatorischen Optimierung das Chipdesign gegenüber den bis dato verwendeten herkömmlichen Methoden und Heuristiken um rund 30 % verbessern könnten. Ingenieure sind misstrauisch, insbesondere gegenüber Mathematikern. Zurecht? So mussten wir den Beweis an einem Testchip antreten. Wir waren mit einer Verbesserung von mehr als 30 % erfolgreich. Das war der Beginn einer außerordentlich erfolgreichen Kooperation zwischen Industrie und Wissenschaft, deren Ergebnis ein vollständiges Designsystem für den physikalischen Entwurf von Chips wurde. Die so genannten BonnTools[®] sind inzwischen in der Chipindustrie weit verbreitet. Mehr als 1 000 Chips wurden bisher mit diesen Algorithmen entworfen. Die drei nachfolgenden Beispiele beleuchten die Entwicklung in den letzten 20 Jahren.

Die diskrete Optimierung hat das Chipdesign wesentlich verbessert, ja in vielen Fällen überhaupt erst möglich gemacht. Andererseits haben aber auch neue Probleme und Fragestellungen im Chipdesign dazu geführt, Theorie, Methoden und Algorithmen in der Diskreten Mathematik weiter zu entwickeln und neue theoretische Erkenntnisse zu erreichen.

Telekommunikationschip ZORA. Die erste große Herausforderung für die diskrete Optimierung im Chipdesign war ein Telekommunikationschip, der 1987 im IBM Forschungslabor Zürich/Rüschlikon für die amerikanische Telekommunikationsfirma Rolm entworfen wurde. Heutige Chips sind um mindestens zwei bis drei Zehnerpotenzen komplexer. Dennoch war ZORA seinerzeit eine besondere Herausforderung. Der Chip hatte die Maße $12,7 \times 12,7$ mm, rund eine Million Transistoren, 15 566 Netze und 58 738 Anschlusspins. Die Zykluszeit mit 28 ns war recht unkritisch, so dass Timing-Optimierung, auch in rudimentärer Form, nicht notwendig war. Die Technologie war HPCMOS, eine der ersten CMOS-Technologien, die nach der bipolaren Transistortechnik aufkamen. Die Schwierigkeit bei diesem Chip war die Verdrahtung, die seinerzeit nur in zwei (zueinander orthogonal stehenden) Verdrahtungsebenen möglich war. Der Chip enthielt ein sehr großes Speichermakro, das nur in der Mitte platziert werden konnte (Abb. 3), während sich die Input-Output-Kontakte (I/O-pins) an der Peripherie des Chips befanden. Bis dahin waren Speicher-Makros für Daten- oder Instruktions-Caches kleiner. Sie wurden in den Ecken eines Chips platziert. Da die Verdrahtung in nur zwei Lagen nicht über das

Makro hinweg führen durfte, wurde von allen Experten und Designern eine Verdrahtung für nicht möglich erachtet.

Unser damals noch rudimentärer Routing-Algorithmus schaffte die Verdrahtung auf Anhieb mit 14,9 m gesamter Netzlänge und 91 439 Vias (Kontakte zwischen den beiden Verdrahtungsebenen) (vgl. Abb. 4).

Der Chip wurde produziert (Abb. 5: Mikroskopische Vergrößerung von ZORA). Wir konnten zum ersten Mal ein Stück Hardware bewundern, das absolut identisch mit den von unseren Algorithmen produzierten Designbildern (Abb. 3 und 4) war.

Abb. 3: Platzierung von ZORA



Abb. 4: Routing von ZORA

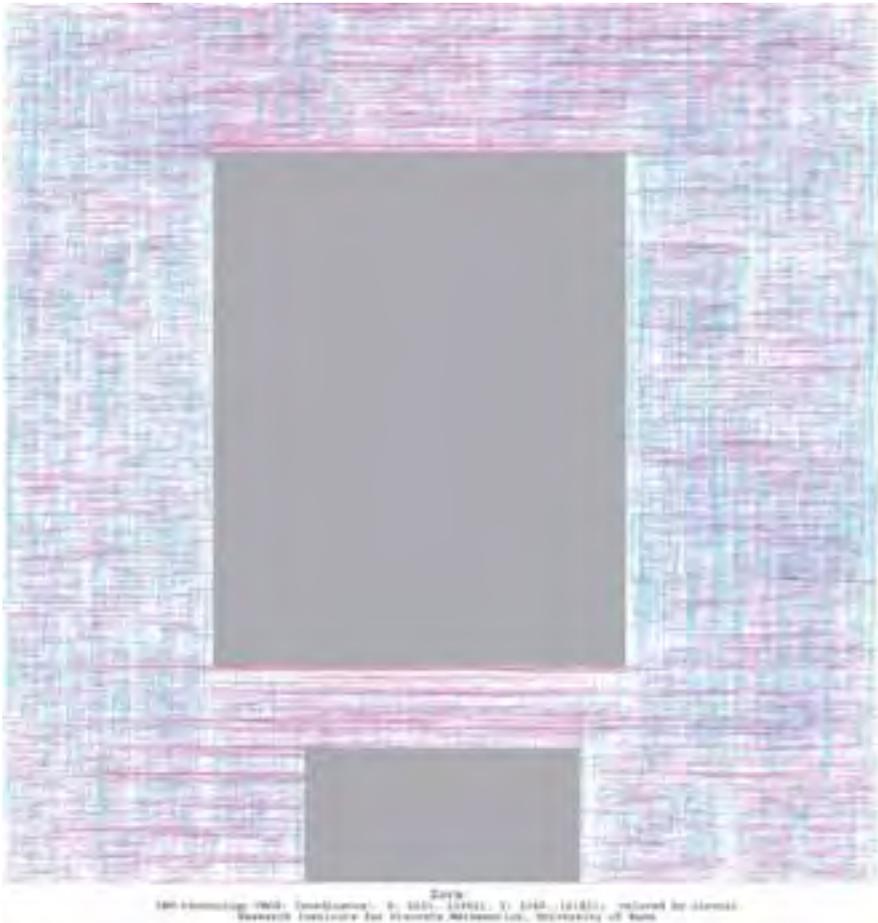
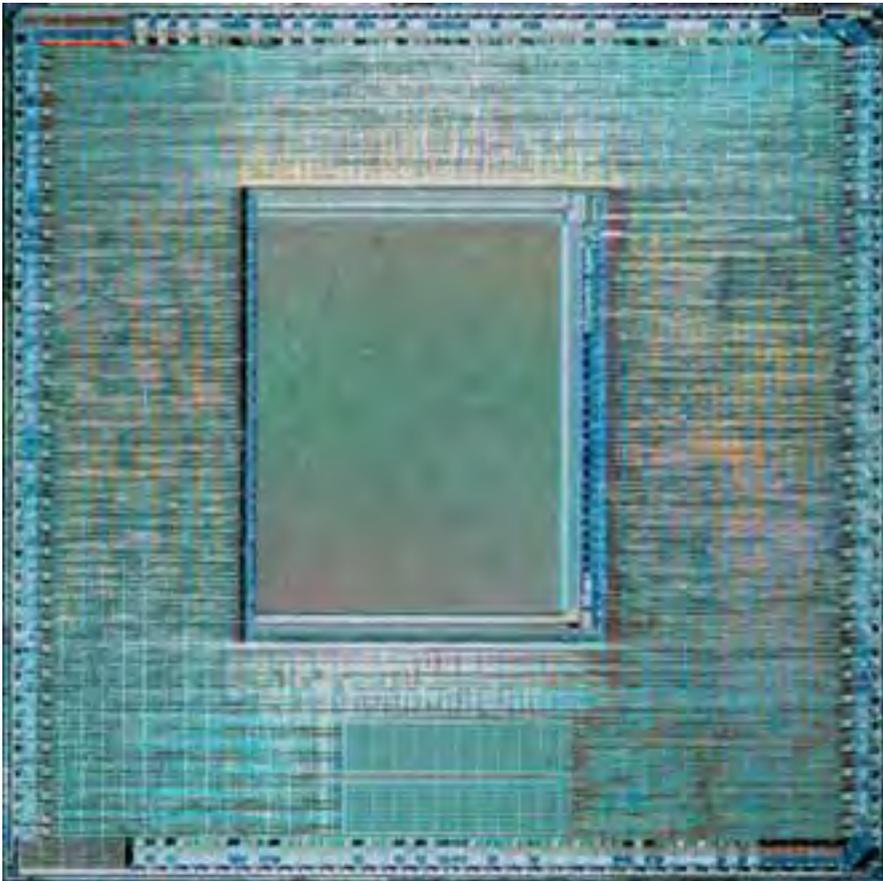


Abb. 5: ZORA in Silizium (mikroskopische Vergrößerung)



Mikroprozessor P2SC. Im Jahr 1995 entstand der POWER2-Chip in CMOS5X6S-Technologie als Mikroprozessor für die POWER-Serie der IBM. Hier wurden zum ersten Mal auf einem einzelnen Chip mehrere – sonst getrennt gefertigte – Einheiten untergebracht (Prozessor, Floatingpoint, Caches und Storage-Controller). Der Chip benötigte schon eine Timing-Optimierung, um die Zykluszeit von 7,4 ns einzuhalten. Er hatte 149 613 Netze. Seine gesamte Verdrahtungslänge betrug 166,3 m mit insgesamt rund 1,4 Millionen Vias. P2SC war übrigens der Prozessor des Rechners Deep Blue, gegen den bekanntlich Kasparow im Schach verlor.

Bei einem hierarchischen Design wird die Logik in einige Hundert Einheiten zerlegt und das Zeitverhalten jeder einzelnen Einheit global vorgegeben. Jede Einheit muss dann die vorgegebenen Laufzeiten einhalten. Abschließend werden die fertig entworfenen Einheiten eines hierarchischen Designs auf der Chipfläche platziert und miteinander verbunden (Top-Level-Verdrahtung). Ein Vorteil dieser hierarchischen Designmethode besteht darin, den Entwurfsprozess der kleinen Segmente genauer kontrollieren zu können, um insbesondere bei nachträglichen Designänderungen lokal und schnell reagieren zu können. Natürlich kann man sich durch diese Segmentierungen von einem möglichen globalen Optimum weit entfernen. Ein hierarchisches Design erfordert einen extrem hohen Entwicklungsaufwand von mehreren 100 Mannjahren, allein 20 Mannjahre werden für den physikalischen Entwurf benötigt.

Bei P2SC hatten wir zum ersten Mal die Gelegenheit, die Vorteile eines flachen Entwurfs (flat design) unter Beweis zu stellen. Beim flachen Entwurf wird keine Segmentierung in kleinere Einheiten vorgenommen, sondern die gesamte Netzliste mit allen Netzen und Circuits flach und ohne hierarchische Einschränkungen platziert. Hierzu sind dann aber Placement- und Routing-Algorithmen notwendig, die sehr wesentlich Methoden der kombinatorischen Optimierung benutzen. Unsere Ergebnisse waren verblüffend: Das Layout konnte in nur sechs Wochen neu entworfen werden. Daran waren nur zwei Mitarbeiter beteiligt. Beim flachen Entwurf wurden Netzlänge und Via-Anzahl um mehr als 20 % reduziert. Die Zykluszeit konnte nicht verbessert werden, da der kritische Pfad, d. h. der langsamste Pfad, der die Zykluszeit bestimmt, in einem Makro lag, das auch beim flachen Design nicht aufgelöst werden konnte. Allerdings wurde beim flachen Design eine wesentlich bessere Verteilung der Laufzeiten auf den einzelnen Pfaden erreicht, was zu einem wesentlich robusteren Chip führt, bei dem auch Produktionstoleranzen besser verkraftet werden können. Das Slack-Histogramm, das die Differenz von erwarteter und tatsächlicher Ankunftszeit zeigt, hatte eine deutlich bessere Verteilung.

Abb. 6 zeigt eine Platzierung vom P2SC im hierarchischen Design, während Abb. 7 eine Variante unseres flachen Entwurfs zeigt. Man sieht, dass der flache Entwurf wesentlich kompakter ist. Große Teile der Chipfläche bleiben ungenutzt. In der Fertigung kann dann eine kleinere Chip-Größe (*die size*) benutzt werden, was wesentliche Produktionskosten einspart.

Abb. 6: P2SC: Hierarchische Platzierung



Abb. 7: P2SC: Flache Platzierung



Systemcontroller U3 für den Apple Power Mac. Im Jahr 2002 entschied sich Apple für den Entwurf eines neuen Power Mac P5, der nach Aussagen seines CEO Steve Jobs der schnellste PC der Welt werden sollte. Die System-Architektur (Abb. 8) enthält als Dual-Core zwei Power PC 970 Prozessoren. Das physikalische Design des Systemcontrollers dieses Systems, der millionenfach produziert wurde, wurde um die Jahreswende 2002/2003 von einem IBM-Ingenieur in Bonn mit BonnTools[®] als flaches ASIC-Design entworfen. Der Chip hatte eine Kantenlänge von 9,3 mm in 130 nm-Cu11-Technologie. Eine Million Circuits mussten mit mehr als einer Million Netzen an vier Millionen Pins verdrahtet werden. Die Gesamtnetzlänge betrug rund 260 m auf sechs Verdrahtungsebenen.

Das besondere Problem beim Entwurf des U3 war die Zykluszeit. Hier sollte zum ersten Mal ein ASIC-Design mit einer Frequenz von mehr als einem Gigahertz getaktet werden. Zur Zykluszeit-Optimierung haben wir ein neues Verfahren BonnCycleOpt[®] eingesetzt, das im Kasten von Abb. 9 kurz beschrieben wird.

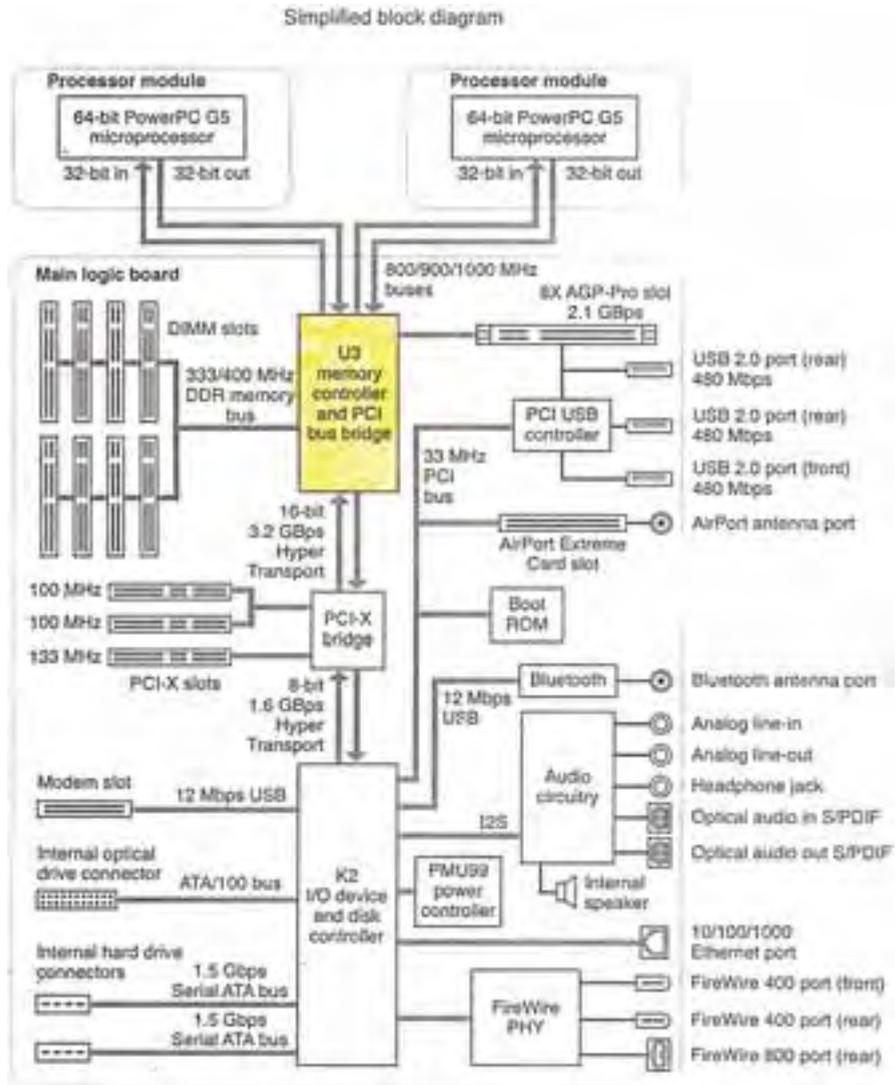
Der Erfolg von BonnCycleOpt[®] war sehr überzeugend: Wir gehen von einem fertig platzierten und verdrahteten Chip aus, bei dem bereits alle zusätzlichen Ansätze zur Laufzeitverbesserung (Gate Sizing, Buffering, Timing-Driven-Placement) ausgenutzt wurden. Der längste Pfad zwischen zwei Registern bestimmt dann die Zykluszeit bzw. Taktfrequenz des Chips. Auf die Frage, ob man die so ermittelte Bottleneck-Zykluszeit weiter verbessern kann, gibt BonnCycleOpt[®] die eindeutige Antwort: Ja, indem das Öffnen und Schließen der Register durch die optimale Berechnung von Clocksignalen für jedes einzelne Register (clock skew) bestimmt wird.

Der Chip U3 hat mehr als ein Dutzend verschiedene Clock-Netz-Gruppen. Abb. 10 zeigt ein besonders kritisches Hochgeschwindigkeits-Clock-Netz. Hier konnte durch die Anwendung von BonnCycleOpt[®], also nur mit mathematischen Methoden eine Verbesserung der Taktfrequenz um 29,6 % und damit eine Taktung mit mehr als einem Gigahertz erreicht werden [9].

Verbesserungen der Chipgeschwindigkeit bzw. Chipqualität in dieser Größenordnung sind im Allgemeinen nur durch neue Technologien möglich (zum Beispiel silicon-on-insulator, Kupfer- statt Aluminium-Verdrahtung, strained silicon). Solche technologischen Veränderungen bedingen eine vollständig andere Fertigung und damit eine neue Chip-Fertigungsstätte (silicon foundry). Der Bau einer neuen „silicon foundry“ kostet heute zwischen 3,5 und 5 Milliarden Dollar. Das ist eine ganz wesentliche Investition. Wie oben gesagt, können nur mit mathematischen Methoden Verbesserungen in derselben Größenordnung erreicht werden. Mathematische Methoden haben Null Investitionskosten, man muss nur nachdenken!

Abb. 11 zeigt den Chip U3 vor und nach Optimierung mit BonnCycleOpt[®]. Blaue und grüne Bauteile bedeuten, dass die Signale dort rechtzeitig bzw. zu früh ankommen. Gelbe, orange und rote Färbungen besagen, dass die Signale zu spät ankommen. Die zugehörigen Slack-Histogramme zeigen die Verteilung der Signalankunftszeiten an den Latches. Die Farben haben dieselbe Bedeutung. Ein direkter Vergleich der Histogramme

Abb. 8: Schaltbild der Apple Power Mac G5 Chip-Architektur

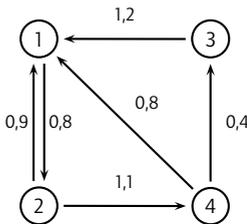


Block Diagram and Buses
Preliminary © Apple Computer, Inc. July 2, 2003

Abb. 9: Grundzüge von BonnCycleOpt®

Für Verfahren zur Optimierung der Zykluszeit wird davon ausgegangen, dass das physikalische Design, insbesondere die Platzierung, vollständig gegeben ist. Alle logischen Verknüpfungen auf einem Chip werden mit Speicherelementen (= Registern) synchronisiert. Die Register werden mit periodischen Clock-Signalen gesteuert. Der Signalfluss zwischen zwei Registern geht über mehrere Stufen (logische Schaltelemente = Circuits) und über eine Vielzahl von Netzen, die diese Circuits verbinden. Die Signale „betreten“ den Chip über primary inputs und verlassen ihn über primary outputs, die wir uns als spezielle I/O-Register vorstellen können.

Sei R die Menge der Register. Auf einem modernen Chip können das einige Hunderttausend sein. Wir konstruieren einen gerichteten Graphen $G^R = (V(G) = R, E(G))$. Zwischen zwei Registern $v, w \in R$ gibt es eine gerichtete Kante (v, w) , wenn ein Signalfluss von v nach w existiert. Den Kanten sind Delays $\delta(v, w)$ zugeordnet, das ist die längste Zeit, die ein Signal von v nach w benötigt. Diese Delays setzen sich aus den Signallaufzeiten auf Drähten und den Schaltzeiten der beteiligten Circuits zusammen.



Im obigen Bild ist ein Registergraph G^R mit 4 Knoten (ohne primary I/O's) dargestellt. Die Zahlen an den Kanten sind Signallaufzeiten (z. B. in ns). Wenn alle Register zur gleichen Zeit öffnen und schließen (zero skew), hätte dieser Schaltkreis eine Taktzeit (=Zykluszeit) T von 1,2 ns, die dem längsten Delay auf der Kante (3,1) entspricht. Wenn man allerdings die Register zu unterschiedlichen Zeiten schalten lässt, z. B.

- Register 1 zu den Zeiten $0; T; 2T; \dots$
- Register 2 zu den Zeiten $-0,1; T-0,1; 2T-0,1; \dots$
- Register 3 zu den Zeiten $-0,3; T-0,3; 2T-0,3; \dots$
- Register 4 zu den Zeiten $0,1; T+0,1; 2T+0,1; \dots$

dann lässt sich, wie man leicht im obigen Graphen

nachprüfen kann, eine Zykluszeit von $T = 0,9$ ns realisieren.

In Registergraphen mit mehreren Hunderttausend Knoten lässt sich die Schaltzeit jedes einzelnen Registers und die optimale Zykluszeit nicht durch Ausprobieren ermitteln. Hier hilft ein mathematischer Satz, der besagt, dass die optimale Zykluszeit gleich dem maximalen Durchschnittsdelay auf gerichteten Kreisen in G^R ist. Der obige Registergraph enthält drei gerichtete Kreise, nämlich

Kreis $1 \rightarrow 2 \rightarrow 1$ mit Durchschnittsdelay

$$\frac{0,8 + 0,9}{2} = 0,85$$

Kreis $1 \rightarrow 2 \rightarrow 4 \rightarrow 3 \rightarrow 1$ mit Durchschnittsdelay

$$\frac{0,8 + 1,1 + 0,4 + 1,2}{4} = 0,875$$

Kreis $1 \rightarrow 2 \rightarrow 4 \rightarrow 1$ mit Durchschnittsdelay

$$\frac{0,8 + 1,1 + 0,8}{3} = 0,9$$

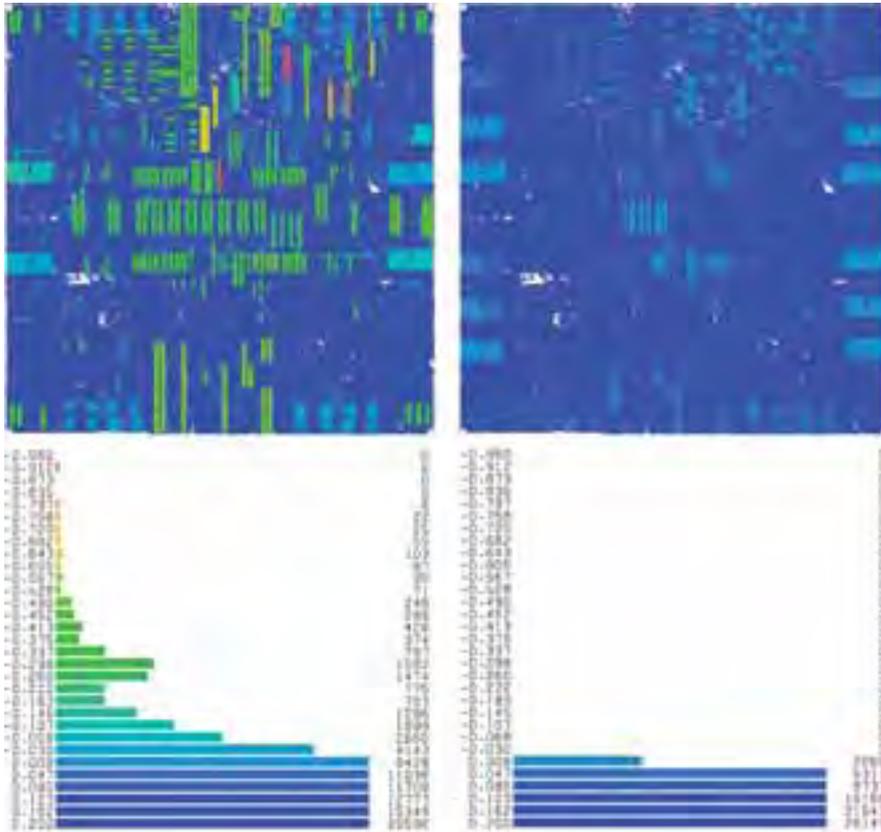
(vgl. [1, 9, 13])

Dieser Satz ist eine Grundlage für BonnCycleOpt®. Allerdings sind die praktischen Probleme zur Zykluszeitoptimierung ungleich komplexer und schwieriger als in diesem einfachen Beispiel dargestellt. Hier haben wir nur die späteste Ankunftszeit der Signale (late mode) berücksichtigt. Die Signale dürfen aber auch nicht zu früh ankommen (early mode). Schließlich geht es nicht nur darum, die Zykluszeit zu minimieren, die durch den gerichteten Kreis in G^R mit maximalem Durchschnittsdelay bestimmt wird. An den Registern in diesem Kreis sind die Schaltzeiten scharf, d. h. es besteht kein Schlupf (slack) zwischen Ankunftszeit des Signals und Schaltzeiten der Register. An allen anderen Registern kann positiver slack vorliegen. Ein erweitertes Problem besteht dann darin, nicht nur den worst slack zu minimieren, sondern die slack-Verteilung zu optimieren (slack balance problem). Das entspricht der Potentialbalancierung auf den Knoten von G^R . Für alle Probleme zur Zykluszeitoptimierung können wir streng polynomielle Verfahren angeben.

Abb. 10: Gigahertz-Clock-Netz von U3



Abb. 11: Chip U3 vor und nach Optimierung mit BonnCycleOpt®



me bzw. der Chipbilder verifiziert die oben genannte Beschleunigung der Zykluszeit um nahezu 30 %.

3 STATUS QUO DES CHIP-DESIGNS

Wie oben bereits erwähnt, beschäftigen wir uns am Forschungsinstitut für Diskrete Mathematik der Universität Bonn seit 1985 mit mathematischen Methoden und Algorithmen zum Chip-Design. Inzwischen gibt es auch kommerzielle Unternehmen, sogenannte EDA (Electronic Design Automation)-Companies, die Tools zum Chip-Design vertreiben. Die drei Marktführer (Cadence, gegr. 1988, Synopsys, gegr. 1986, und Magma, gegr.

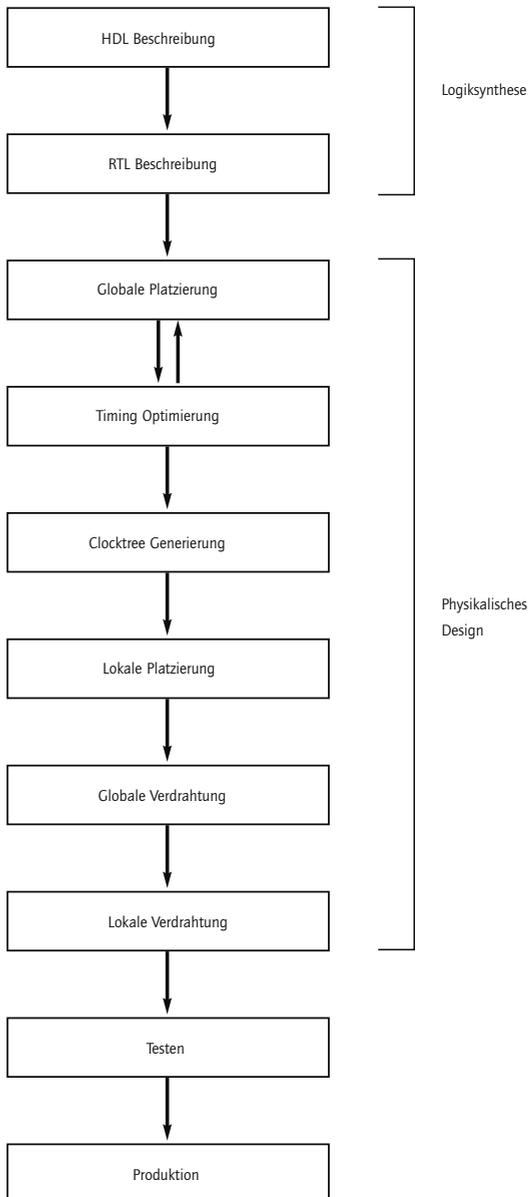
1997) sind später gegründet worden. Der jährliche Umsatz dieser Marktführer liegt bei rund 4 Milliarden Dollar. Wir wollen im Folgenden Entwicklung und Status quo des Chip-Designs anhand unserer BonnTools[®] beschreiben. Die BonnTools[®] zeichnen sich durch einen besonders hohen Anteil nichttrivialer, innovativer (diskreter) Mathematik aus. Das Ergebnis des Chip-Designs soll ein funktionierendes industrielles Produkt sein, das in sehr großen Stückzahlen produziert wird. Der Designingenieur erwartet eine vollständige Lösung seines Problems, nicht nur die Anwendung von schönen und schweren mathematischen Sätzen. Insofern müssen auch wir neben komplexen mathematischen Verfahren schließlich die eine oder die andere „quick-and-dirty“-Heuristik zusätzlich anwenden, um eine vollständige Lösung des Chip-Designs zu erreichen. Das Verhältnis von Mathematik und quick-and-dirty-Verfahren bei den BonnTools[®] liegt bei etwa 90:10. Das ist auch der große Vorteil der BonnTools[®] gegenüber kommerziellen Produkten. Eine ausführlichere Übersicht über die BonnTools[®] findet man in [10] und in [11].

In Abb. 12 haben wir den Design-Fluss von der ersten Beschreibung der Funktion eines Chips in einer Hardware Description Language (HDL) bis zur Produktion schematisch skizziert. Die HDL kann man sich wie eine höhere Programmiersprache vorstellen, die Funktion und Funktionsblöcke (noch ohne Berücksichtigung einer technischen Realisierung) formal beschreibt. In einem ersten Schritt der Logiksynthese wird diese Beschreibung auf den Register Transfer Level (RTL) heruntergebrochen. Register sind kleine Speicherelemente, die das Ergebnis der logischen Verknüpfungen in der kombinatorischen Logik zwischenspeichern. Die Laufzeit zwischen den einzelnen Registern bestimmt die Zykluszeit T bzw. die Taktfrequenz $1/T$ des Chips. In einem weiteren Schritt der Logiksynthese werden dann bereits technologieunabhängige Boolesche Ausdrücke generiert, die dann in einem nächsten Schritt in die Schaltelemente (Circuits) der entsprechenden Technologie abgebildet werden. Wenn man von gewissen Modifikationen der Logik bei der Timing-Optimierung absieht, ist die eigentliche Logiksynthese nicht Gegenstand der BonnTools[®]. Logikoptimierung ist ein extrem schwieriges, aber noch weit offenes Feld, für das ein effizienter mathematischer Ansatz sehr wichtig wäre, aber bisher noch nicht bekannt ist.

BonnTools[®] bearbeiten den gesamten Komplex des physikalischen Designs, der sich dann anschließt. Die beiden ersten Stufen im physikalischen Design, nämlich globale Platzierung und Timing-Optimierung werden mehrfach iterativ durchlaufen. Sie sind heutzutage ein Kernstück und wohl auch der schwierigste Teil des physikalischen Designs. Diese Iteration wird auch unter dem Namen „Timing Closure“ zusammengefasst. In diesem Teil des Design-Flusses wird die Platzierung bis auf sehr kleine Veränderungen und Verschiebungen festgelegt, die optimale Transistorgröße der Circuits bestimmt, der Slack an den einzelnen Registern optimiert, die Topologie von Netzen (Repeater Trees) festgelegt, die Verstärkung der Signale durch Inverter oder Buffer (Repeater) optimiert und der V_t -Level (threshold voltage) bestimmt.

Anschließend wird dann der Clock Tree gebaut, wobei die berechneten Slacks an den Registern durch genau festgelegte Ankunftszeiten der Clock-Signale berücksichtigt

Abb. 12: Vereinfachte schematische Darstellung des Design-Flusses



werden müssen. Moderne Chips können mehrere hundert verschiedene Clocks haben und damit Wälder von Clock Trees.

In der anschließenden lokalen Platzierung wird eine überlappungsfreie Anordnung aller Circuits durch möglichst geringe Verschiebungen erreicht. Sowohl bei der globalen als auch bei der lokalen Platzierung muss die Verdrahtbarkeit des Chips beachtet werden. Die in den einzelnen Circuits durch logische Verknüpfung entstehenden Signale werden von deren Outputpin an die Inputpins von mehreren anderen Circuits verteilt. Ein moderner Chip kann mehr als 10 Millionen solcher Netze haben, die in bis zu 10 rektilinearen Verdrahtungslagen knotendisjunkt eingebettet werden müssen. Das geschieht im globalen und lokalen Routing.

Sehr vereinfachend kann man das Problem des physikalischen Designs wie folgt beschreiben: Gegeben sind mehrere Millionen kleiner Legosteine (Platzhalter für die Circuits), die gleiche Höhe aber unterschiedliche Breite haben, sowie einige größere Legosteine (Speicherelemente, Makros, Cores), die alle auf einer Chipfläche von 1 bis 3 cm² disjunkt platziert werden müssen. Diese Legosteine sprechen miteinander über Netze. Der Outputpin eines Legosteins schickt sein Signal über ein Netz an mehrere andere Legosteine. Diese Netze müssen in einem dreidimensionalen Gittergraphen knotendisjunkt verlegt werden. Hierbei müssen insbesondere die Signalflüsse zwischen den Registern in der geforderten Zykluszeit möglich sein. Es gibt noch zahlreiche zusätzliche technische Nebenbedingungen (zum Beispiel Noise, Coupling), auf die hier nicht eingegangen werden kann.

In die verschiedenen Komponenten der BonnTools[®] gehen eine Vielzahl unterschiedlicher kombinatorischer Optimierungsprobleme ein, auf die hier ebenfalls nicht im Detail eingegangen werden kann. Zur Demonstration der mathematischen Vielfalt und Tiefe dieser Anwendungen im Chip-Design seien hier die wichtigsten aufgelistet: kürzeste Wege, Netzwerk-Design, Steiner-Bäume, maximale Flüsse, Transportprobleme, kostenminimale Flüsse, Mehrgüterflüsse, disjunkte Wege und Bäume, diskrete Time-Cost Trade-off Probleme, Minimum-Mean-Cycle, parametrische kürzeste Wege, minimale aufspannende Bäume, Binpacking, Knapsackprobleme, Travelling Salesman Probleme, Huffman-Codes, Facility Location.

Wie nicht anders zu erwarten, werden diese kombinatorischen Optimierungsprobleme nicht in ihrer lehrbuchmäßigen Reinform angewendet, sondern fast immer in für die Anwendung notwendigen Modifikationen, deren Richtigkeit und Effizienz auch bewiesen werden muss.

Wie oben bereits erwähnt, ist modernes Chip-Design ohne die wesentliche Anwendung der diskreten Mathematik nicht mehr möglich. Andererseits haben die Anwendungen im Chip-Design neue methodische und theoretische Fragestellungen gebracht, die zu einer Weiterentwicklung von Theorie und Methoden beigetragen haben. Aufgrund der Vielzahl neuer Probleme, die gerade aktuell bei neuen Technologien auftreten, wird diese gegenseitige Befruchtung weiter zunehmen und noch lange anhalten.

Nachfolgend werden wir einen kurzen Überblick über die einzelnen Stufen des physikalischen Designs geben:

3.1 PLATZIERUNG

Der Platzierungsalgorithmus BonnPlace[®] besteht im Wesentlichen aus zwei Komponenten, der globalen Platzierung und der lokalen Platzierung. Die globale Platzierung endet mit einem nicht zulässigen Ergebnis, bei dem die einzelnen Objekte noch nicht überlappungsfrei platziert sind. Die Auslastung der verschiedenen Regionen des Chips ist aber so, dass eine überlappungsfreie Platzierung durch kleine lokale Verschiebungen erreicht werden kann, die im lokalen Platzierungsalgorithmus vorgenommen werden. Daher wird die lokale Platzierung auch Legalisierung oder Detailed Placement genannt.

Für die globale Platzierung gibt es drei verschiedene Ansätze:

- Simulated Annealing
- min-cut Verfahren
- analytische Platzierung

Das Simulated Annealing ist eine Heuristik, die einem physikalischen Abkühlprozess nachempfunden wurde und die wir zu den Quick-and-Dirty Verfahren zählen, da sie weder eine Motivierung durch das spezielle Anwendungsproblem (hier: Chip-Design) noch durch einen besonderen mathematischen Ansatz hat. Chips mit 10 Millionen beweglichen Objekten hätten bei einem Simulated Annealing Verfahren selbst auf den schnellsten Computern astronomische Rechenzeiten und darüber hinaus schlechte Resultate. Diese Kritik gilt auch für andere ähnliche Heuristiken (zum Beispiel genetische Algorithmen, neuronale Netzwerke u. a.), die sich aus anderen Wissenschaftsgebieten eine interessante Philosophie und vielleicht auch noch eine Metatheorie ausgeborgt haben, die aber für den Anwendungsfall des Chip-Designs nicht brauchbar ist.

Min-cut Verfahren waren bis vor etwa 10 bis 15 Jahren populär. Hier wird durch eine Austauschheuristik die Anzahl der Netze, die über eine vorab definierte Schnittlinie gehen, minimiert. Hauptkritik an diesem Verfahren, das bei gewissen Chip-Strukturen brauchbare Ergebnisse liefert, besteht darin, dass die Zielfunktion hinsichtlich eines eindimensionalen Objekts (Schnittlinie) minimiert wird, obwohl der Chip selbst ein zweidimensionales Objekt ist. Durch Minimierung der Netzübergänge über den Schnitt können auf den beiden Seiten des Schnitts Platzierungen entstehen, die dann schlecht oder gar nicht verdrahtet werden können.

Analytische Platzierungsverfahren minimieren eine lineare oder quadratische Schätzfunktion der Netzlänge, die ein nicht-disjunktes Ergebnis liefert, dessen Überlappungsfreiheit dann in einem zweiten Verfahren (zum Beispiel Multisection) erreicht wird. BonnPlace[®] gehört zu dieser Kategorie.

BonnPlace[®] minimiert die Summe der gewichteten quadratischen Längen aller Netze, wobei als Schätzung für die Netzlänge jedes einzelne Netz als Clique modelliert wird (nachfolgend quadratische Platzierung genannt). Brenner und Vygen [5] haben gezeigt, dass das Cliquenmodell die topologieunabhängige Approximation eines Steinerbaums

minimaler Länge ist. Für die Wahl einer quadratischen Zielfunktion gibt es mehrere triftige Gründe. Eine quadratische Platzierung führt im Allgemeinen zu eindeutigen Positionen für die beweglichen Objekte, sie ist stabil, d. h. nahezu invariant gegenüber kleinen Änderungen der Netzliste [15], und sie ist mit konjugierten Gradienten-Verfahren mit unvollständiger Cholesky-Zerlegung sehr schnell und parallel zu lösen. Eine quadratische Platzierung für 5 Millionen bewegliche Objekte kann in wenigen Minuten Rechenzeit erzeugt werden. Ein weiteres Argument für den quadratischen Ansatz liegt darin, dass die Signalverzögerung quadratisch mit der Länge der Drähte wächst.

Die quadratische Platzierung hat üblicherweise in ihren ersten Schritten sehr viele Überlappungen, die nicht durch lokale Korrekturen behoben werden können. Wir müssen daher sicherstellen, dass Teile des Chips nicht überlastet sind, d. h. der Flächenbedarf für die beweglichen Objekte in dieser Region darf nicht größer sein als die zur Verfügung stehende Fläche selbst. Ein zweiter Algorithmus, den wir Multisection nennen, trägt dafür Sorge. Die Grundidee besteht darin, eine Region des Chips in mehrere Teile aufzuteilen und dabei die beweglichen Objekte diesen Teilregionen so zuzuordnen, dass die Kapazitätsbedingungen eingehalten werden und die Gesamtbewegung der Objekte minimiert werden soll, d. h. die durch die quadratische Platzierung erreichte Positionierung soll so wenig wie möglich verändert werden. Bei diesem Zuordnungsproblem von Objekten zu Regionen haben wir eine Millionenzahl von Objekten, aber im Allgemeinen nur wenige Regionen (vier, neun o. ä.). Dennoch ist das (0,1)-Zuordnungsproblem NP-schwer. Wir können hier aber die nicht-ganzzahlige Zuordnung wählen, also ein Hitchcock-Transportproblem, weil Vygen [14] gezeigt hat, dass jede fraktionale Lösung nur $k-1$ nicht-ganzzahlige Zuordnungen hat, wobei k die (kleine) Anzahl von Regionen ist. Brenner [4] hat für dieses Problem einen neuen Algorithmus mit einer Laufzeit von $O(nk^2(\log n + k \log k))$ angegeben, der darüber hinaus in der Praxis außerordentlich schnell ist.

In Abb. 13 sehen wir das Ergebnis einer Multisection auf neun Gebiete. Die Farben entsprechen den Zuordnungen zu den Gebieten. Rote Objekte werden der linken oberen Region zugeordnet, gelbe Objekte der mittleren oberen Region, usw. Die Zuordnung ist optimal bezüglich der l_1 -Distanz. In Abb. 14 zeigen wir die ersten sechs Schritte der globalen Platzierung von BonnPlace[®]. Das globale Verfahren terminiert, wenn die Höhe einer Region der Höhe der Circuits entspricht. Die dann noch fehlende Überlappungsfreiheit wird durch die Legalisierung (lokale Platzierung) erreicht.

Auf die Legalisierung wird hier nur sehr vereinfachend mit wenigen Worten eingegangen. Nach Abschluss der globalen Platzierung gibt es Regionen auf dem Chip, die überlastet sind, d. h. mehr Zellen enthalten als dort platziert werden können. Sie sind die Angebotsknoten. Regionen, die noch zusätzlichen Platz haben, werden Nachfrageknoten. Zwischen den Knoten werden Kanten generiert, auf denen ein Fluss fließen kann. Die Überlappungsfreiheit, d. h. der Ausgleich zwischen Angebots- und Nachfrageknoten wird über kostenminimale Flüsse erreicht. Die Bestimmung der Regionen, deren Größe und die Bestimmung der Flusskanten ist sehr komplex. Wir gehen darauf hier

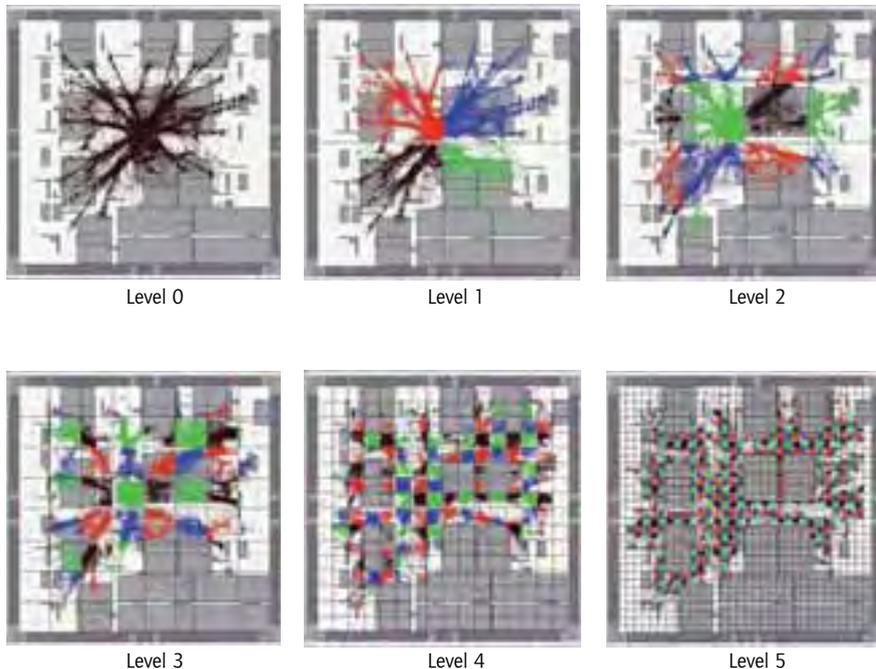
Abb. 13: Multisection mit $3 \times 3 = 9$ Gebieten

nicht ein, sondern zeigen in Abb. 15 den „Flussgraphen“, der Legalisierung eines kleinen Teils eines Chips. Angebotsregionen sind rot, Nachfrageregionen sind grün gezeichnet. Die blauen Kanten repräsentieren den kostenminimalen Fluss. Die Breite der Kanten ist proportional zur Menge des Flusses. Die Legalisierung wird im Design Flow mehrfach eingesetzt, zum Beispiel auch nach der Generierung des Clock Tree oder nach Veränderung der Circuitgrößen in der Timing-Optimierung.

3.2 TIMING OPTIMIERUNG

Vor 20 Jahren spielte die Timing-Optimierung im physikalischen Design eines Chips so gut wie keine Rolle. Der Chip wurde platziert und verdrahtet, anschließend wurde zur Kontrolle das Timing berechnet. Die geforderte Zykluszeit wurde stets erreicht und damit

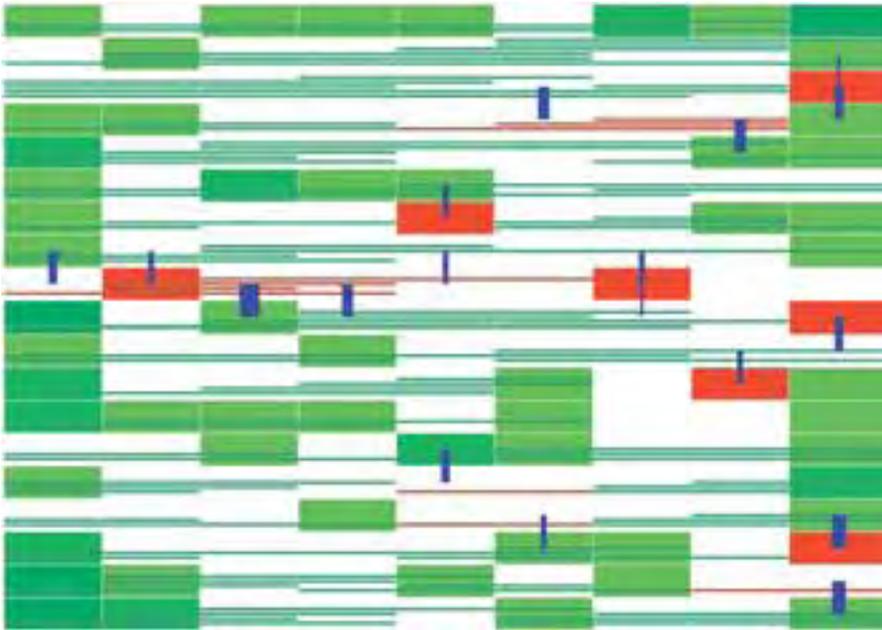
Abb. 14: Die ersten sechs Schritte der globalen Platzierung mit Aufteilung auf 1, 4, 16, 64, 256 und 1024 Gebiete



war das physikalische Design abgeschlossen. Damals hatte man aber auch Zykluszeiten von 20 Nanosekunden und mehr, was Taktfrequenzen von 50 MHz entspricht. Heute takteten schnelle Prozessoren mit einer Frequenz von 4 GHz, was einer Zykluszeit von nur 250 Pikosekunden entspricht. Hier müssen viele technologische Tricks und mathematische Verfahren angewendet werden, um ein solches Ziel zu erreichen.

Das Gesamtgebiet der Timing-Optimierung, auch Timing-Closure genannt, besteht aus einer Vielzahl von Algorithmen und Methoden. Hierbei können neue Circuits (Verstärker = Buffer, Repeater, Inverter) eingefügt werden oder Circuits hinsichtlich ihrer Größe verändert werden. Dabei wechseln sich diese Verfahren mit der lokalen Platzierung iterativ ab. Wir wollen hier nur einige wesentliche Komponenten der Timing-Closure erwähnen: Fan-out-Tree (Repeater-Tree) Konstruktion, Fan-in-Tree Optimierung, Gate-Sizing (Transistor-Sizing), Clock-Skew-Scheduling (Zykluszeitoptimierung) und Slack-Balancing, Clock-Tree-Konstruktion.

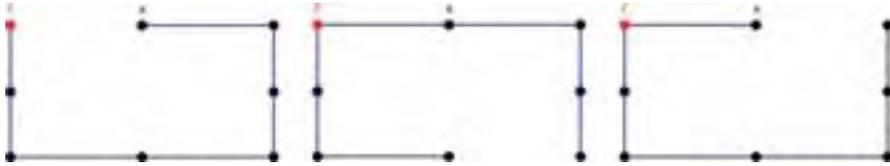
Abb. 15: Legalisierung auf einem kleinen Teil eines Chips



Repeater-Trees sind ein gutes Beispiel, um zu demonstrieren, wie neue Technologien auch neue Anforderungen an die Design-Methoden und damit an die mathematischen Algorithmen stellen. Früher war die Verzögerung von Signalen durch die aktiven Schaltelemente (Circuits, Transistoren) dominiert, während die Signalverzögerung über die Verbindungen (Drähte) zweitrangig war. In sehr modernen Technologien spielen die letztgenannten Verzögerungen die Hauptrolle. Man geht davon aus, dass in zukünftigen Technologien mehr als die Hälfte aller Circuits nicht für die eigentlichen Schaltvorgänge, sondern nur für die Überbrückung der Distanzen (als Verstärker = Repeater) benutzt werden müssen.

Bei der Konstruktion von Repeater-Trees geht es darum, ein Ausgangssignal eines Circuits (Quelle) an die Eingänge von mehreren anderen Circuits (Senken) optimal zu verteilen. Hierbei muss einerseits die Topologie des Baumes optimal sein, andererseits müssen, wenn zu lange Distanzen überbrückt werden, Verstärker (Repeater) eingefügt werden. In Abb. 16 sehen wir drei Steiner-Bäume, die die Quelle r mit sieben Senken verbindet. Alle drei Bäume haben dieselbe minimale Länge. Wenn nun der Weg von r zur Senke s zeitkritisch ist, dann ist der linke Baum die schlechteste Realisierung, weil

Abb. 16: Drei verschiedene Baum-Topologien minimaler Länge

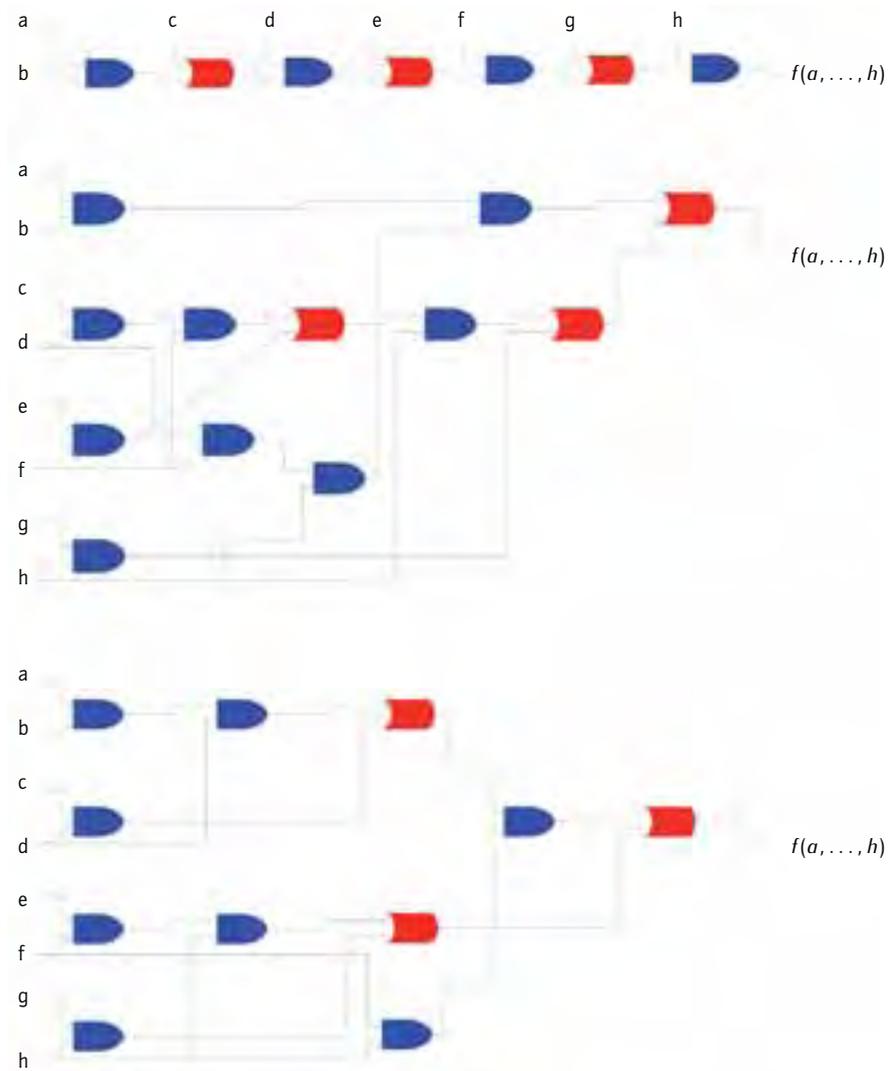


die kapazitive Last von r nach s am größten ist. Beim mittleren Baum ist zwar der direkte Weg von r nach s am kürzesten, das Circuit s muss aber noch die kapazitiven Lasten der folgenden Kanten und Knoten des Baumes treiben, was zu einer Signalverschlechterung führt. Der rechte Baum ist in diesem Fall als Repeater-Tree optimal. Wie schon gesagt, müssen neben der Topologiegenerierung noch Wege durch Repeater überbrückt werden. Einzelheiten des sehr komplizierten Verfahrens zur Konstruktion von Repeater-Trees findet man bei Bartoschek et al. [2].

Bei der Konstruktion von Fan-in-Trees hat man die umgekehrte Situation. Hier müssen verschiedene Signale zu einem einzigen Signal kombiniert werden, was durch eine logische Funktion (Boolescher Ausdruck) vorgeschrieben ist. Ein solcher Fan-in-Tree kann aber unterschiedlich gestaltet sein und insbesondere eine unterschiedliche Tiefe haben. Allgemein kann man sagen, dass je kürzer ein solcher Tree ist, desto schneller werden die Signale propagiert. Im Grunde geht es bei der Fan-in-Tree-Optimierung daher um Optimierung der Logik (Logiksynthese), die hier aber nicht vor der physikalischen Realisierung betrieben wird, sondern bereits unter Kenntnis des physikalischen Layouts. In Abb. 17 sehen wir drei logisch äquivalente Realisierungen eines Booleschen Ausdrucks, der nur aus logischen UND-Gattern (rot) und logischen ODER-Gattern (blau) besteht. Dieser Ausdruck hat acht Boolesche Variable (a bis h). Der Leser kann sich davon überzeugen, dass diese drei Realisierungen logisch äquivalent sind, wobei die Bäume mit kürzerer Tiefe eine schnellere Signalpropagierung ermöglichen. Die Fan-in-Tree-Optimierung geschieht im Wesentlichen mittels dynamischer Programmierung. In [16] findet man weitere theoretische und praktische Resultate dieses Verfahrens. Die Fan-in-Tree-Optimierung wird sehr selektiv am Ende der Timing-Closure-Iteration eingesetzt und zwar auch nur bei zeitkritischen Pfaden. Hier können dann aber Zeitgewinne von 100 Pikosekunden und mehr einen großen Gewinn beim Erreichen der Timing-Closure bedeuten.

Das Gate- oder Transistor-Sizing ist ein klassisches Verfahren bei der Timing-Optimierung. In einer Technologie-Bibliothek ist jedes Circuit in verschiedenen Ausprägungen vorhanden. Ein großes Circuit treibt ein Signal besonders schnell, es hat aber einen hohen Stromverbrauch und auch eine große Kapazität. Kleinere Circuits schalten lang-

Abb. 17: Drei logisch äquivalente Realisierungen eines Schaltkreises aus UND- und ODER-Gattern



samer bei geringerem Stromverbrauch und kleinerer Kapazität. Leider produziert das Gate-Sizing gegenläufige Effekte, die in der Optimierung berücksichtigt werden müssen. Ein großes Circuit beschleunigt zwar die Signalpropagierung. Es hat aber eine große kapazitive Last, so dass es zu einer Verzögerung bei seinen Vorgängern kommt, die ja diese gesamte nachfolgende Last treiben müssen. Insofern ist ein Gate-Sizing, das sowohl vorwärts als auch rückwärts (getriebene Last) orientiert ist, notwendig. Klassische Verfahren des Gate-Sizings beruhen auf konvexer/geometrischer Optimierung [3, 6, 7], wobei angenommen wird, dass alle auftretenden Funktionen als Posynome approximiert werden können. Leider haben diese Verfahren den Nachteil, dass sie nur sehr kleine Instanzen bearbeiten können und damit praktisch unbrauchbar sind. Wir benutzen ein iteratives Verfahren, das zur Größenbestimmung des Circuits auch berücksichtigt, wie schnell die Spannung eines Signals von 0 nach 1 bzw. umgekehrt wechselt (slew). Dieses Verfahren ist auch für große praktische Instanzen sehr schnell. Man kann beweisen, dass unter der Annahme, die Delay-Funktion sei ein Posynom und die Circuits sind kontinuierlich skalierbar (wie auch beim geometric programming gefordert), dieses iterative Verfahren die minimale Signalverzögerung liefert [8].

Das Problem, an jedem Register die Clock-Signale individuell und optimal zu steuern (clock skew scheduling) haben wir schon oben in Abschnitt 2 (Systemcontroller U3, vgl. auch Abb. 9) kurz erwähnt. Hier können wir nur noch kurz nachtragen, dass wir hierfür seit Neuem ein Verfahren zur Balancierung der Skews an allen Registern benutzen, das formal eine Potentialbalancierung auf dem Knoten eines gerichteten Graphen ist. Held [8] konnte ein erstes stark polynomielles Verfahren für das Potentialbalancierungsproblem in nicht-kreisfreien Graphen angeben.

Es gibt noch eine Reihe weiterer Tuningparameter für das Timingverhalten (z. B. plane assignment, V_T -Level), auf deren Darstellung hier aus Platzgründen verzichtet wird. Nachdem alle Register der Timing-Optimierung gezogen wurden und „die letzten Pikosekunden ausgeknautscht wurden“, bleibt nur noch die Konstruktion des Clock Trees, eines Baums, der die Clock-Steuerung an allen Registern vornimmt. Moderne Chips haben mehrere Hundert verschiedene Clock Trees mit einer Million und mehr Registern. Bei einer Taktfrequenz im Gigahertz-Bereich muss man etwa 20–30 % der Zykluszeit als Sicherheitsintervall für Fertigungstoleranzen reservieren, so dass nur noch einige Hundert Pikosekunden übrig bleiben. Das bedeutet wiederum, dass bei der Optimierung der Ankunftszeiten der Clocksignale an den Registern (clock skew) schon wenige Pikosekunden Bedeutung haben können.

Wir konstruieren den Clock Tree bottom-up, aus einzelnen Teilbäumen zusammengesetzt. Zur Clock Tree-Konstruktion benötigen wir eine Menge von Senken (Register) mit einem Zeitintervall für jede Senke und eine Menge von Quellen (Wurzeln von Teilbäumen). Ferner benötigen wir einen (abstrakten) logisch korrekten Clock Tree, der alle Register mit Clock-Signalen versorgt. Um die Ankunftszeit des Clock-Signals in dem vorgegebenen Zeitintervall zu gewährleisten, wird üblicherweise eine Balancierung über die Drahtlänge vorgenommen, das heißt, die exakte Signalverzögerung wird gegebenenfalls

durch Umwege in der Verdrahtung erreicht. Dieser Ansatz hat den Nachteil, dass erhebliche Verdrahtungsressourcen nur für diese Balancierung benötigt werden, und dass ein genau vorgeschriebenes Drahtlayout für den Clock Tree bei einem bereits verdrahteten Chip schwer zu realisieren ist. Wir benutzen daher einen anderen Ansatz [9], bei dem die Verdrahtung kürzestmöglich ist und bei dem die Balancierung der Verzögerungen durch die Konstruktion der Topologie des Clock Trees und durch entsprechendes Sizing der Repeater erreicht wird. Dieses Verfahren ähnelt der oben beschriebenen Konstruktion von Repeater-Trees. Die Signalverstärkung erfolgt durch Inverter, wobei zwei hintereinander geschaltete Inverter einem Repeater (buffer) entsprechen. In Abb. 18 sehen wir verschiedene Stufen einer Clock-Tree-Konstruktion mit BonnClock[®]. Die farbigen Flächen bedeuten Gebiete, in denen Inverter (aktuelle Senken) platziert werden können. Die Farben entsprechen Ankunftszeiten innerhalb des Clock Trees: Blau steht für Signale nahe der Quelle, Grün, Gelb und Rot für spätere Ankunftszeiten. Während der bottom-up-Konstruktion bewegen sich die farbigen Flächen langsam in Richtung Quelle, die bei diesem Beispiel etwa in der Mitte des Chips angeordnet ist.

Wie gesagt, wird der Clock-Tree bottom up konstruiert. Fast 90 % des gesamten Stromverbrauchs eines Clock Trees verursacht bereits die unterste Stufe. Insofern ist hier ein gutes Clustering der Senken von großer Bedeutung. Wir lösen dieses Problem mit einem modifizierten Facility-Location-Ansatz, für den Maßberg und Vygen [12] einen ersten Approximationsalgorithmus mit konstantem Approximationsfaktor angegeben haben. In Abb. 10 haben wir bereits einen Gigahertz-Clock-Tree für den Chip U3 gezeigt.

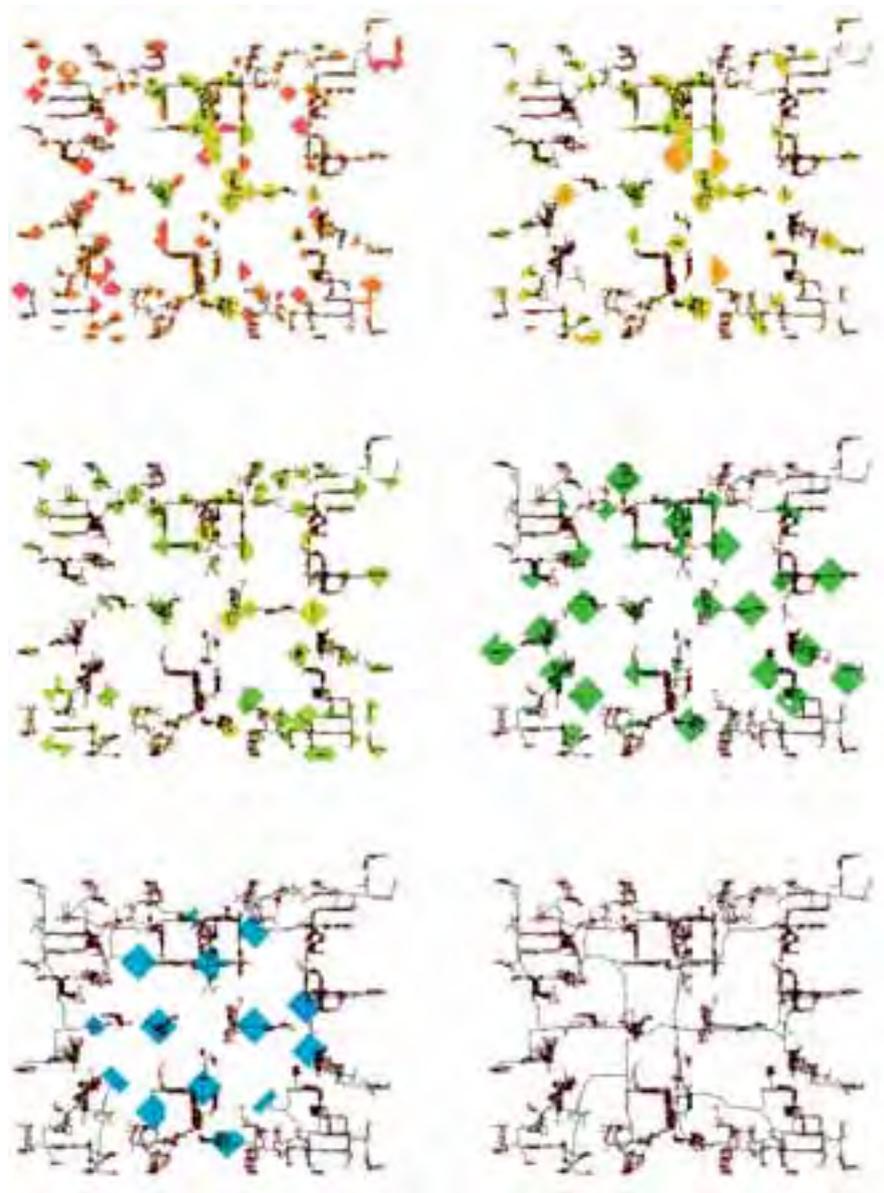
3.3 ROUTING

Das kantendisjunkte Verlegen der Netze eines Chips in einem dreidimensionalen Gittergraphen geschieht im Routing. Hier sind die Datenvolumina gigantisch. Bei einem sehr großen Chip sind Zehnmillionen Netze in einem Gittergraphen zu verdrahten. Dieser Graph hat mehrere Hundertmilliarden Knoten und Kanten. Für ein flaches Routing, das heißt ohne eine hierarchische Zerlegung in Teile, ist das eine sehr große Herausforderung, die unser Routingalgorithmus BonnRoute[®] bei mehreren Tausend Anwendungen in Designcentern hervorragend gemeistert hat.

BonnRoute besteht aus zwei Hauptteilen, nämlich einem globalen und einem lokalen (detaillierten) Routing. Das globale Routing bestimmt das Gebiet des Chips, in dem jedes Netz verlegt werden soll. Hierdurch wird die Suche im lokalen Routing stark eingeschränkt. Der Global Routing Algorithmus arbeitet auf einem wesentlich kleineren Graphen, in dem nämlich einige hundert benachbarte Knoten eines Gittergraphen zu einem Superknoten kontrahiert werden. Die Kanten zwischen den Superknoten haben dann eine Kapazität, die der Anzahl der Kanten des Gittergraphen entspricht, die in den kontrahierten Knoten münden.

Das Global Routing-Problem besteht dann darin, Steinerbäume für die einzelnen Netze in diesem kontrahierten Gittergraphen kantendisjunkt zu packen. Eine fraktionale

Abb. 18: Verschiedene Stufen der Clock-Tree-Konstruktion bei BonnClock®



Relaxierung dieses Problems kann mit einer Modifikation des Mehrgüter-Fluß-Problems gelöst werden. Das zugehörige LP Problem können wir dann um weitere Nebenbedingungen ergänzen, die zum Beispiel Timing, Coupling, Yield oder Stromverbrauch berücksichtigen. Hierdurch können dann auch diese Zielfunktionen bei der Optimierung berücksichtigt werden. Der Global Router von BonnRoute ist der erste Algorithmus mit einer beweisbaren Gütegarantie, der auch Nebenbedingungen für Timing, Coupling, Noise, Yield und Stromverbrauch berücksichtigt [11]. Zur detaillierten Lösung des relaxierten LP's werden dann die klassischen Techniken des Randomized Roundings benutzt. Der Global Router bestimmt für eine gegebene Platzierung ein Global Routing, soweit ein solches existiert. Andernfalls liefert er ein Zertifikat, dass das gegebene Placement nicht verdrahtbar ist.

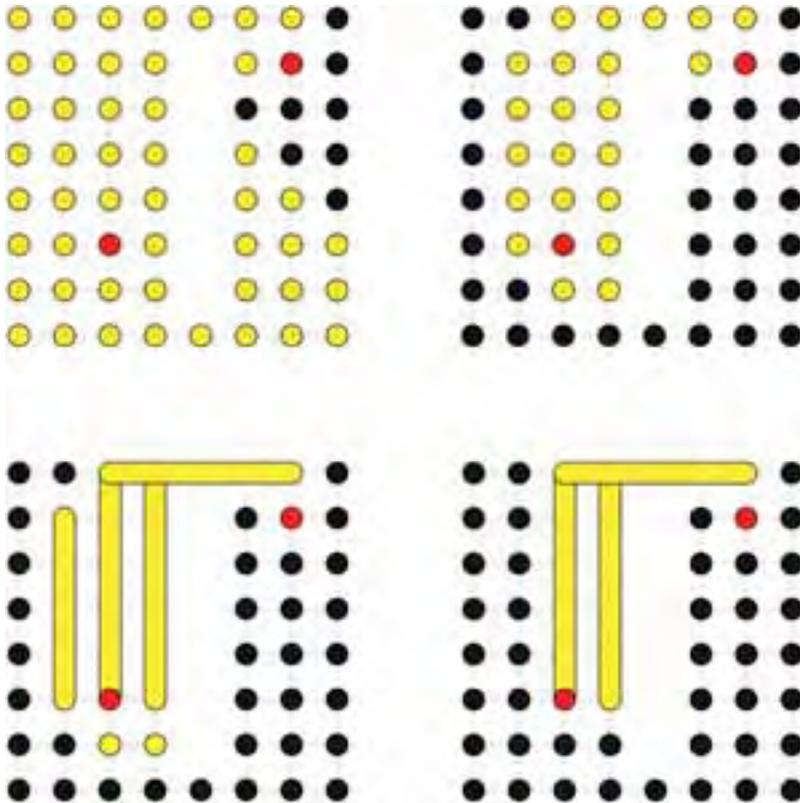
Der detaillierte Router verlegt dann die einzelnen Netze in dem ursprünglichen Gittergraphen, wobei er die Informationen des globalen Routers benutzt. Die Kanten des globalen Routinggraphs, in denen die Steinerbäume der einzelnen Netze bereits verlegt wurden, definieren dann Korridore im Gittergraphen für die detaillierte Verlegung der Netze. Hierdurch wird der Suchraum im Gittergraphen wesentlich eingeschränkt. So wird es erst möglich, viele Millionen Netze in vernünftiger Rechenzeit zu verlegen.

Die Steinerbäume der einzelnen Netze werden sequentiell zum Beispiel in der Reihenfolge ihrer Zeitkritikalität verdrahtet, wobei die einzelnen Steinerbäume wiederum auch sequentiell aus Zwei-Punkt-Verbindungen aufgebaut werden. Diese Verbindungen werden dann durch kürzeste Wege realisiert. Hierfür bietet sich natürlich der klassische Markierungsalgorithmus von Dijkstra an. Leider ist eine Lehrbuch-Implementierung dieses Algorithmus für so große Probleminstanzen ungeeignet. Die theoretische schnellste Implementierung des Dijkstra-Algorithmus mit Fibonacci-Heaps hat eine Laufzeit von $O(m + n \log n)$ wobei m und n die Anzahl der Knoten und Kanten des Graphen sind. Dies ist für unsere Anwendung viel zu langsam.

Zur Beschleunigung benutzen wir im Wesentlichen zwei Modifikationen. Eine erste Beschleunigung besteht in einer zielgerichteten Suche. Hierbei wird für jeden Knoten eine untere Schranke seines kürzesten Weges zur Senke als Knotenpotential berechnet. Mit diesen Knotenpotentialen wird die Kostenfunktion der Kanten modifiziert. Hierdurch wird erreicht, dass die Markierungsstrategie des Dijkstra-Algorithmus sich nicht in Wellen um die Quelle bewegt, sondern zielstrebig von Quelle in Richtung Senke durchgeführt wird. Die beiden oberen Bilder in Abb. 19 zeigen die markierten Knoten (gelb) für den Dijkstra-Algorithmus ohne (links) und mit (rechts) Future Cost. Es ist ein kürzester Weg von dem roten Knoten links unten zu dem roten Knoten rechts oben zu finden. Der klassische Dijkstra-Algorithmus markiert hierfür fünfzig Knoten (gelb), während der zielgerichtete Algorithmus nur 24 Knoten markiert.

Eine ganz wesentliche Beschleunigung des Dijkstra Algorithmus erreichen wir durch die Markierung von Kantenzügen (Intervallen) anstelle von einzelnen Knoten. Für die Chipverdrahtung ist der Gittergraph nicht vollständig. Er enthält verschiedene Blockaden und auch Löcher. Nun kann man einen vollständigen Kantenzug bis zum nächsten

Abb. 19: Dijkstra Algorithmus mit vier verschiedenen Markierungsverfahren



Hindernis (Blockade oder Loch) in einem Schritt markieren. Dann wächst die Laufzeit nicht mehr mit der Anzahl der Knoten sondern mit der Anzahl der Hindernisse, die aber ungleich kleiner ist als die Anzahl der Knoten. Die beiden unteren Bilder in Abb. 18 zeigen die Markierung eines intervall-basierten Dijkstra-Algorithmus ohne (links) und mit (rechts) Zielorientierung. Hier sind dann nur noch sieben bzw. vier Markierungen vorzunehmen. Da die Laufzeit des Dijkstra Algorithmus proportional zu den durchgeführten Markierungen ist, wird die drastische Einsparung durch intervall-basierte Markierungen augenfällig.

Neue Chip-Technologien arbeiten mit sogenannten gridless Bibliotheken. Hier liegen die Anschlusspunkte der Circuits nicht mehr auf Gitterpunkten. Wir können hier nicht auf Details eingehen, sondern nur anmerken, dass die Datenstrukturen von Bonn-

Route so gewählt wurden, dass sowohl gitter-basierte als auch gridless Circuits verdrahtet werden können.

4 STÄRKEN/SCHWÄCHEN-ANALYSE, HERAUSFORDERUNGEN

Noch immer benötigt das Design von Prozessoren mit einer Taktfrequenz von mehreren GHz den zehnfachen Ingenieuraufwand eines ASICs von einem GHz. Diese Effizienzlücke kann nur durch leistungsfähigere Algorithmen geschlossen werden, die die technischen Herausforderungen der Nanometer Technologien berücksichtigen können:

- Variabilität: Das Layout muss so beschaffen sein, dass selbst bei Produktionsvariationen der elektrischen Parameter von 30 % ein korrektes Zeitverhalten sichergestellt ist. Im Prozessorendesign wird das üblicherweise durch ein sehr symmetrisches, meist von Hand erzeugtes Layout sichergestellt, das aber keineswegs eine global optimale Lösung darstellt. Ein Optimierungsansatz muss die statistische Variation der elektrischen Parameter und ihre Korrelation berücksichtigen und für alle Situationen ein gültiges Layout liefern.
- Zuverlässigkeit: Die gleichen Faktoren, die eine Erhöhung der Variation erzeugen, führen auch zu einer größeren Ausfallwahrscheinlichkeit der Transistoren und der Verdrahtung. Diese Risiken werden durch einen hohen Designaufwand in DFM (Design for Manufacturability) minimiert. Ein Chip ist aber nur dann funktionsfähig, wenn jeder einzelne der über eine Milliarde Transistoren fehlerfrei arbeitet. Ein bekannter Lösungsansatz für Speicherbausteine sind redundante Speicherzellen. In der Zukunft müssen auch Logikbausteine und deren Verdrahtung redundante Strukturen aufweisen. So ist BonnRoute® heute schon in der Lage, lokale „Loops“ einzubauen, um die Robustheit von Vias deutlich zu erhöhen. In Zukunft muss die Verdrahtung auch in großem Maße durch redundante Verbindungen gegen Unterbrechungen der Leiterbahnen geschützt werden. Auch redundante Logik ist ein weiterer logischer Schritt zur Erhöhung der Zuverlässigkeit. Aus Gründen der Testbarkeit wird redundante Logik heute vermieden. Wir werden hier also einen Paradigmenwechsel erleben, der völlig neue Ansätze benötigt.
- Verlustleistung: Die Reduktion der Verlustleistung tritt immer mehr in den Vordergrund und zeigt, dass die reine Erhöhung der Schaltfrequenz zu einem sehr ungünstigen Verhältnis von Rechenleistung und Stromverbrauch geführt hat. Es ist viel sinnvoller, die Rechenleistung pro Watt zu optimieren. Die in der Kooperation entwickelten Entwurfsprogramme stellen durch Timingoptimierung der Logik sicher, dass die angestrebte Schaltfrequenz mit einer möglichst geringen Verlustleistung erreicht wird. Dabei wird sowohl die bei einem Schaltvorgang verbrauchte ‚dynamische‘ Energie als auch der immer stärker dominierende passive Energieverbrauch durch Leckströme berücksichtigt. Wie bereits erwähnt, ist

der Energieverbrauch zum großen Teil durch die Architektur und Logikbeschreibung gegeben. Optimierungsmethoden müssen daher schon auf der Verhaltensbeschreibung ansetzen und schnell Vergleiche von verschiedenen Implementierungen erlauben.

5 VISIONEN UND HANDLUNGSEMPFEHLUNGEN

Viele Probleme der zukünftigen Technologien und Produkte können nur durch die gleichzeitige Optimierung des Layouts und der Logik erreicht werden. Die Verlustleistung ist zum Beispiel zu 80 % durch die Logikstruktur der Schaltung vorgegeben und die möglichen Optimierungen der verbleibenden 20 % sind im Wesentlichen heute schon erreicht. Auch muss die Architektur der Logik schon frühzeitig die Möglichkeiten des Layouts berücksichtigen: So ist es einfach, die Anzahl der Ein- und Ausgänge eines Crossbars mit jeder neuen Technologie zu verdoppeln, aber die dadurch entstandene Vervielfachung der Verbindungen ist ab einer bestimmten Größe nicht mehr realisierbar. Solche Probleme frühzeitig zu erkennen, verlangt eine enge Zusammenarbeit zwischen Logikdesign und Layout und die Möglichkeit, eine gegebene Logikbeschreibung effizient auf die Realisierbarkeit zu überprüfen und Probleme aufzuzeigen. Diese Überlegungen zeigen auch, wie wichtig es ist, verschiedene Architekturen schnell und in hohem Maße automatisiert zu implementieren, um genaue Vergleiche der endgültigen Leistungsmerkmale zu haben und sich nicht auf grobe Abschätzungen verlassen zu müssen.

Auf der anderen Seite müssen die physikalischen Herausforderungen im Layout berücksichtigt werden, um überhaupt noch Strukturen produzieren zu können, die deutlich kleiner sind als die Wellenlänge des verwendeten Lichts. Der gesamte Lithographie- und Herstellungsprozess wird heute mathematisch modelliert und optimiert (virtual fab). Dieses Modell wird dann benutzt, um Strukturen auf dem Chip gezielt zu verändern und die unvermeidbaren Abweichungen im Herstellungsprozess zu kompensieren.

Die Komplexität des Entwurfsprozesses und der Technologien haben einen Punkt erreicht, in dem nur die holistische Betrachtung von der Logikstruktur bis zum Herstellungsprozess ein erfolgreiches Produkt garantieren kann.

Das ultimative Ziel eines ‚silicon compilers‘ der aus jeder gegebenen Logikbeschreibung automatisch eine Schaltung erzeugt, die mit einer Taktfrequenz von mehreren GHz betrieben werden kann, wird wohl so bald nicht erreichbar sein. Trotzdem muss die Automatisierung des Schaltungsentwurfs weiter vorangetrieben werden. Ein wichtiges Erfolgsrezept unserer Kooperation ist die enge Zusammenarbeit auf echten Produktdesigns in den neuesten verfügbaren Technologien, die heute 45 nm Strukturbreiten haben, und auch die frühzeitige Diskussion der Herausforderungen zukünftiger Technologien bis zu 22 nm. Eine erfolgreiche Kooperation ist nur durch offenen Informationsaustausch und gemeinsame Arbeit an herausfordernden Projekten möglich.

6 LITERATUR

- [1] Albrecht, C., Korte, B., Schietke, J., und Vygen, J. [2002]: Cycle time and slack optimization for VLSI-chips. *Discrete Applied Mathematics* 123 (2002), 103–127.
- [2] Bartoschek, C., Held, S., Rautenbach, D., und Vygen, J. [2006]: Efficient generation of short and fast repeater tree topologies. *Proceedings of the International Symposium on Physical Design* (2006), 120–127.
- [3] Boyd, S., Kim, S.-J., Patil, D. und Horowitz, M. [2005]: Digital circuit optimization via geometric programmierung. *Operations Research* 53 (2005), 899–932.
- [4] Brenner, U. [2008]: A faster polynomial algorithm for the unbalanced Hitchcock transportation problem. *Operations Research Letters* 36 (2008), 408–413.
- [5] Brenner, U., und Vygen, J. [2001]: Worst-case ratios of networks in the rectilinear plane. *Networks* 38 (2001), 126–139.
- [6] Chen, C.-P., Chu, C.C.N., und Wong, D.F. [1999]: Fast and exact simultaneous gate and wire sizing by Lagrangian relaxation. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 18 (1999), 1014–1025.
- [7] Fishburn, J., und Dunlop, A. [1985]: TILOS: A posynomial programming approach to transistor sizing. *Proceedings of the IEEE International Conference on Computer-Aided Design* (1985), 326–328.
- [8] Held, S. [2008]: Timing-Closure in Chip Design. Ph.D. Thesis, University of Bonn 2008.
- [9] Held, S., Korte, B., Maßberg, J., Ringe, M. und Vygen, J. [2003]: Clock scheduling and clocktree construction for high performance ASICs. *Proceedings of the IEEE International Conference on Computer-Aided Design* (2003), 232–239.
- [10] Held, S., Korte, B., Rautenbach, D., und Vygen, J. [2008]: Combinatorial Optimization in VLSI-Design, to appear in: Chvatal, V., Sbihi, N. (eds.) *Combinatorial Optimization: Methods and Applications*. IOS Press.
- [11] Korte, B., Rautenbach, D., und Vygen, J. [2007]: BonnTools: Mathematical innovation for layout and timing closure of systems on a chip. *Proceedings of the IEEE* 95 (2007), 555–572.

- [12] Maßberg, J., und Vygen, J. [2008]: Approximation algorithms for a facility location problem with service capacities. *ACM Transactions on Algorithms*, 4 (2008), to appear. (Preliminary version in APPROX 2005).
- [13] Szymanski, T. [1992]: Computing optimal clock schedules. *Proceedings of the ACM/IEEE Design Automation Conference*, 1992, 399–404.
- [14] Vygen, J. [2005]: Geometric quadrisection in linear time, with application to VLSI placement. *Discrete Optimization 2* (2005), 362–390.
- [15] Vygen, J. [2007]: New theoretical results on quadratic placement. *Integration, the VLSI Journal* 40 (2007), 305–314.
- [16] Werber, J., Rautenbach, D., und Szegedy, C., [2007]: Timing optimization by restructuring long combinatorial paths. *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design* (2007), 536–543.

MATERIALIEN UND MECHANIK

> CHANCEN UND VISIONEN DER MODERNEN MECHANIK

WOLFGANG EHLERS UND PETER WRIGGERS

1 EXECUTIVE SUMMARY

Die Gesellschaft des 21. Jahrhunderts wird ihre Anforderungen an Lebensqualität und Umweltstandards stetig erhöhen. Dies betrifft auch die Anforderungen an Qualität und Vorhersagegenauigkeit bei der Berechnung komplexer Probleme und umfasst insbesondere das Gesamtdesign von Produkten unserer unmittelbaren Umgebung wie Architektur- und Ingenieurbauwerken, aber auch von Industrieprodukten, die wir in unserem täglichen Leben einsetzen.

Anforderungen an Design, Konstruktion und Materialfestigkeit von Bauwerken machen es erforderlich, den Status einer häufig immer noch empirisch dominierten Materialbeschreibung zu überwinden und dieser ein simulationsbasiertes Design neuer Werkstoffe mit maßgeschneiderten Hightech-Eigenschaften entgegenzusetzen. Was für Bauwerke gilt, gilt in einer industriell geprägten Gesellschaft wie der unsrigen in besonderem Maße auch für industriell gefertigten Produkte, durch deren Qualität der Industriestandort Deutschland und seine Gesellschaft nicht nur großes Ansehen in der Welt genießen, sondern mit denen wir auch unser hohes Bruttosozialprodukt und unseren Lebensstandard sichern.

Neben Forschungen in der Materialwissenschaft, die zu einer Verbesserung und Weiterentwicklung von Materialien und deren Festigkeitseigenschaften beitragen, kann eine ganzheitliche Qualitätssteigerung von Produkten nur erreicht werden, wenn die Lösung zunehmend komplexer werdender Ingenieurprobleme durch neue und weiterentwickelte Strategien der Modellierungs- und Simulationsmethoden vorangetrieben wird. Erste Entwicklungen neuer Simulationswerkzeuge strahlen bereits heute in große Teile unserer Gesellschaft aus, die weit über die üblichen Ingenieur Anwendungen hinausgehen. Dies betrifft in besonderem Maße auch die Medizintechnik und weite Bereiche der in einer alternden Gesellschaft immer bedeutender werdender Lebenswissenschaften.

Was kann die Mechanik für die Wissenschaft, die Wirtschaft und für die Gesellschaft des 21. Jahrhunderts tun? Die Verwendung von Hightech-Materialien und die systematische Untersuchung skalenübergreifender Problemstellungen schließen heute ebenso wie die Modellbildung multiphysikalischer Zusammenhänge Formulierungen ein, bei denen

gekoppelte Gleichungssysteme aus mehreren Gebieten der Physik zu lösen sind. Um Prozesse zeitlich und räumlich genauer abbilden zu können, werden außerdem zunehmend sogenannte Multi-Skalen-Methoden in Anwendung gebracht. Diese ermöglichen durch „Zoomen“ in kleinere Orts- und Zeitskalen eine detaillierte Analyse der dort vorhandenen Strukturen und tragen somit zu einem besseren Verständnis des Materialverhaltens bei.

Die Simulation geometrisch dreidimensionaler Probleme mit mehrskaligen und multiphysikalischen Eigenschaften führt in der Regel auf sehr große Systeme gekoppelter partieller Differentialgleichungen, die nur numerisch mit dem Einsatz modernster Computer gelöst werden können. Hier ist ein Zusammenwirken von Wissenschaftlern aus den Bereichen Mechanik und Mathematik unbedingt erforderlich. Grundsätzlich ist diese Erkenntnis nicht neu, sondern geht bereits auf die Tradition der Gesellschaft für Angewandte Mathematik und Mechanik (GAMM) seit ihrer Gründung am 21. September 1922 in Leipzig zurück. Diese Tradition gilt es zu stärken und zum Wohle unserer Gesellschaft und ihrer Wissenschaft und Wirtschaft zu gestalten. Durch das Zusammenwirken von Mechanik und Mathematik wird es uns gelingen, die bereits entstandenen Simulationswerkzeuge weiterzuentwickeln und sie gemeinsam mit Wissenschaftlern aus der Informatik und den Informationswissenschaften zu einem integrativen Werkzeug der Simulationstechnologie auszubauen, vgl. dazu auch den Exzellenzcluster „Simulation Technology“ (SimTech) an der Universität Stuttgart (www.simtech.uni-stuttgart.de).

2 ERFOLGSGESCHICHTEN

Auf dem Weg zu einer genaueren Analyse und einem besseren Verständnis von Materialien und Strukturen ist noch ein weiter Weg zu gehen. Durch den Einsatz moderner Computer und schneller Algorithmen aus dem Bereich der Numerischen Mathematik sind jedoch in den letzten Jahren Voraussetzungen geschaffen worden, die eine neue Ära in der Mechanik eingeleitet haben. Die Behandlung von mechanischen Problemstellungen im Rahmen einer komplexen mechanischen Modellbildung, der Entwicklung zugehöriger mathematischer Modelle und ihrer numerischen Lösung wird heute häufig unter dem Begriff *Computational Mechanics* zusammengefasst.

Die in diesem Rahmen entwickelten Werkzeuge werden Motor für viele technische Entwicklungen sein. Einige Beispiele beleuchten die Anwendungsgebiete sowie die Chancen und Visionen der modernen Mechanik.

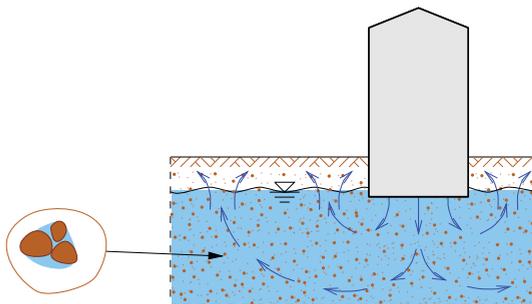
2.1 KONTINUUMSMECHANIK VON MEHRKOMONENTEN-MEHRPHASEN-MATERIALIEN (MULTI-FIELD AND MULTI-PHYSICS MATERIALS)

Gründung von Bauwerken – ein Jahrtausende altes Problem der Baukunst. Häufig ist es ausreichend, die Gründung von Bauwerken als Festkörperproblem im Sinne der Belastung des Untergrunds durch ein Fundament zu behandeln. Allerdings ergeben sich zu-

nehmend Problemstellungen, bei denen diese Vorgehensweise nicht mehr genügt. Nehmen wir an, ein Bauwerk mit einer Fundamentplatte soll in wassergesättigtem Boden gegründet werden, so genügt es nicht, die Bauwerkslast über die Fundamentplatte auf den Baugrund abzusetzen und diesen wie einen reinen Festkörper zu behandeln. Stattdessen muss ein Konsolidationsproblem berechnet werden, bei dem nach dem Aufbringen der äußeren Last diese zunächst allein durch das Porenwasser übernommen wird. Allerdings bewirkt die Belastung des wassergesättigten Bodens einen Druckgradienten im Porenwasser, der wiederum zu einem Strömungsprozess führt. Das Wasser wird aus dem Porenraum des Bodens gedrängt, der sich in der Folge verformt und nun über den Zusammenhang von Verformung und Spannung in der Lage ist, mit der äußeren Belastung durch die Fundamentplatte einen Gleichgewichtszustand zu bilden.

Offensichtlich muss hier ein gekoppeltes Problem der Festkörper-Fluid-Interaktion berechnet werden, d. h. die Eigenschaften von Festkörpern und Fluiden treffen aufeinander und beeinflussen sich gegenseitig. Im Sinne der mathematischen Beschreibung dieses Problems im Rahmen der Theorie Poröser Medien (TPM) [1], einer kontinuumsmechanischen Methode zur Beschreibung interagierender Körper, deren Mikrostrukturen auf dem gemeinsam eingenommenen Gebiet homogenisiert werden, entsteht ein komplexes, nichtlineares System gekoppelter, partieller Differentialgleichungen. Dieses Gleichungssystem kann in der Regel nur numerisch zum Beispiel mit Hilfe der Finite-Elemente-Methode (FEM) gelöst werden. Moderne Rechner sind zunehmend in der Lage, große und größte Gleichungssysteme zu beherrschen. Damit diese Rechner optimal genutzt werden können, sind die Methoden der Mechanik und der Angewandten Mathematik erforderlich, um gekoppelte Probleme dieser Art zu analysieren und im Rechner erfolgreich zu behandeln.

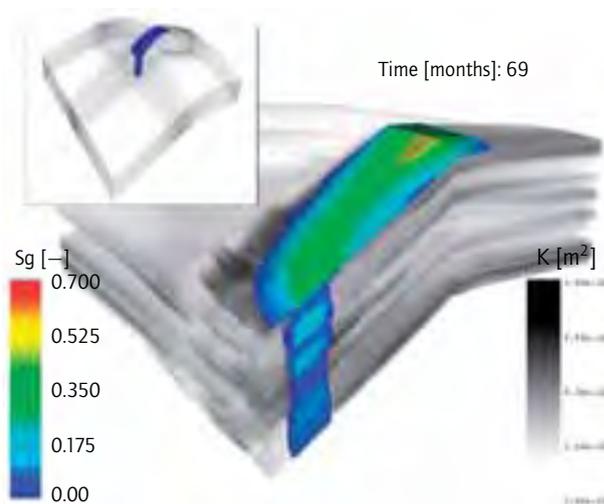
Abb. 1: Schematische Darstellung des Konsolidationsproblems



CO₂-Sequestrierung – eine Herausforderung an die Mechanik umweltrelevanter Systeme. Vor dem Hintergrund der globalen Erwärmung gewinnt die Speicherung umweltrelevanter Gase zunehmend an Bedeutung. Vor allem Kraftwerke, die mit fossilen Brennstoffen betrieben werden, sondern große Mengen an Kohlendioxid (CO₂) ab. Moderne Rauchgasreinigungsanlagen sind in der Lage, CO₂ abzuscheiden und es damit einer Speicherung zugänglich zu machen. Basierend auf dem Gedanken, dass die fossilen Brennstoffe, die im Verbrennungsprozess CO₂ erzeugen, in Georeservoirien gespeichert waren, liegt es nahe, abgeschiedenes CO₂ in Georeservoirien zu sequestrieren. Dazu wählt man wasserführende Bodenschichten (Aquifere), die sich unter einem undurchlässigen Deckgebirge befinden.

Die Berechnung der CO₂-Sequestrierung stellt wie das Konsolidationsproblem ein gekoppeltes Problem dar. Neben dem Bodengerüst, das den Aquifer trägt, und dem Wasser des Aquifers kommt jedoch zusätzlich CO₂ ins Spiel, das in überkritischem Zustand unter hohem Druck in den Aquifer eingepreßt wird. Dadurch verformen sich sowohl das Bodengerüst als auch das Deckgebirge. Sinkt der Fluiddruck, dann kann in der Folge eine Phasentransformation einsetzen und das CO₂ vom überkritischen in den gasförmigen Zustand übergehen. Neben dem gekoppelten Problem der Festkörper-Fluid-Interaktion (Festkörperdeformation vs. Filterströmung) kommen hier weitere physikalisch-chemisch motivierte Kopplungen ins Spiel. Es entsteht ein hochgradig komplexes System stark

Abb. 2: Numerische Berechnung der CO₂-Sequestrierung. Aufsteigen des eingepressten CO₂ bis unter das undurchlässige Deckgebirge (Quelle: www.simtech.uni-stuttgart.de).



gekoppelter partieller Differentialgleichungen, das im Rahmen der Finite-Elemente-Methode numerisch gelöst werden kann. Die Größe des numerischen Problems setzt jedoch in der Regel zusätzlich adaptive Algorithmen (Orts- und Zeitadaptivität) auf massiv parallelen Systemen voraus, vgl. dazu auch [1, 2].

Herausforderung Biomechanik. Die Biomechanik, und insbesondere die Biomechanik harter und weicher Gewebe, ist eine relativ junge, aufstrebende Disziplin der modernen Kontinuumsmechanik. Die Vision der beteiligten Wissenschaftler ist die Schaffung eines „Menschmodells“, bei dem skalenübergreifende Informationen zur Verfügung stehen, die von der Grobskala des „Crashdummys“ über die Feinskala der Gewebe und Organe bis zur Mikroskala auf Zellebene reichen. Ein solches Modell könnte helfen, Unfallfolgen genauer zu analysieren oder den Verlauf von Operationen virtuell vor auszuplanen. Aber auch die Abschätzung von Alterungsprozessen des Bewegungsapparats oder die Vorausplanung patientengenaue Implantate als Ersatz für verschlissener Gelenke oder Teile der Wirbelsäule sind vorstellbar.

Im Rahmen der modernen Kontinuumsmechanik Poröser Medien [1] können bereits heute Teile des menschlichen Körpers simuliert werden. Betrachtet man die Wirbelsäule als den wesentlichen Teil unseres Stütz- und Bewegungsapparats, so lässt sich sowohl das mechanische Verhalten der einzelnen Wirbelkörper als auch das der einzelnen Bandscheiben beschreiben, vgl. [3]. Insbesondere stellt die mechanische Modellierung der Bandscheibe mit ihrem gallertartigen Gewebe aus dem Nucleus pulposus (Innenbereich) und dem Anulus fibrosus (Außenring) eine große Herausforderung dar, da das Material nicht nur inhomogen aufgebaut ist, sondern sich auch anisotrop verformt. Zusätzlich müssen elektrochemisch getriebene Schwell- und Schrumpfprozesse berücksichtigt werden, durch die die Bandscheibe als ein avaskuläres Gewebe mit Nährstoffen versorgt wird. Diesen Effekt nehmen wir z. B. dadurch wahr, dass unser Körper morgens etwa 2 cm größer ist als abends. Während des Tages ist die Wirbelsäule belastet. Im Nucleus entsteht ein Innendruck, der dazu beiträgt, einen Teil der interstitiellen Flüssigkeit aus dem Gewebe auszupressen. Dies erzeugt ein chemisches Ungleichgewicht, das sich während der nächtlichen Ruhephase ausgleicht, indem die ausgepresste Menge der interstitiellen Flüssigkeit wieder angesaugt wird.

Neben den gekoppelten Gleichungen zur Beschreibung des komplexen mechanischen Verhaltens des Feststoffmaterials aus Kollagenfasern und Proteoglykanen mit anhaftenden elektrischen Ladungen und des Verhaltens der interstitiellen Flüssigkeit, in der sich positiv und negativ geladene Ionen befinden, müssen Diffusionsgleichungen zur Beschreibung der Schwell- und Schrumpfprozesse berücksichtigt werden, d. h. das mechanische Verhalten und das elektrochemische Verhalten sind gekoppelt und können nicht isoliert betrachtet werden. Wie bei den zuvor diskutierten Beispielen entsteht auch hier ein hochgradig komplexes System stark gekoppelter partieller Differentialgleichungen, das im Rahmen der Finite-Elemente-Methode numerisch gelöst werden kann. Die Größe des numerischen Problems verlangt auch hier in der Regel den Einsatz adaptiver Algorithmen (Orts- und Zeitadaptivität) auf massiv parallelen Systemen.

Abb. 3: Links: Aufnahme eines Patienten und automatische Generierung individueller Daten (Fiktion).
Rechts: Numerische Berechnung der Flexion der Lendenwirbelsäule und Darstellung des dadurch induzierten Druckaufbaus in den einzelnen Bandscheiben (Quelle: <http://www.simtech.uni-stuttgart.de>)



2.2 SIMULATIONSTECHNIKEN ZUR BESCHREIBUNG VON HETEROGENEN MATERIALIEN (MULTI-SCALE MODELLING)

Für ein vertieftes Verständnis des konstitutiven Verhaltens von Materialien, die sich aus unterschiedlichen Ingredienzien auf verschiedenen Längenskalen zusammensetzen – wie z. B. Beton, ein Werkstoff, der aus Zement, Sand und Kies besteht – verwendet man klassischerweise experimentelle Techniken. Basierend auf den gewonnenen Testergebnissen werden dann Materialmodelle entwickelt, die das Verhalten auf den unterschiedlichen Längenskalen verschmieren. Man kann heute jedoch auch numerische Simulationstechniken einsetzen, um das Verhalten auf den unterschiedlichen Längenskalen zu berücksichtigen, was dann zu Multi-Skalen-Methoden führt. Diese Technik ist neu und stellt eine echte Herausforderung dar, denn Simulationen, die zu quantitativen Resultaten führen, müssen dreidimensional durchgeführt werden.

Innerhalb einer Multi-Skalen-Simulation werden unterschiedliche dreidimensionale mechanische Modelle auf jeder Skale eingesetzt, um das Materialverhalten für die geometrischen und konstitutiven Eigenheiten der jeweiligen Skale zu berücksichtigen. Dazu werden häufig sogenannte Repräsentative Volumenelemente (RVE) verwendet, die das Materialverhalten auf der betrachteten Skale mit hinreichender Genauigkeit charakterisieren können. Die RVE werden dann mechanischen Belastungen ausgesetzt, die aus der höheren Skale bekannt sind. Dies führt auf eine Materialantwort, die nach einem Mit-

telungsprozess wieder an die höhere Skale zurückgegeben wird. Diesen Prozess nennt man Homogenisierung, siehe auch [4, 5].

Eine numerische Multi-Skalen-Analyse berücksichtigt die komplexe dreidimensionale Geometrie der Mikrostruktur auf der betrachteten Skale und erfordert daher ein aufwendiges numerisches Modell, das häufig auf Gleichungssysteme mit mehr als einer Million Unbekannten führt. Daher werden neben schnellen Computern auch schnelle Gleichungslöser aus der modernen Numerischen Mathematik benötigt, wenn die Methode im Ingenieurbereich eingesetzt werden soll. Damit hat man dann ein schlagkräftiges Werkzeug zur Verfügung, das die internen mechanischen Eigenschaften eines heterogenen Materials auflösen kann und so neue Einsichten ermöglicht, die in dieser Form einer experimentellen Analyse oft nicht zugänglich sind.

Multi-Skalen-Analyse von Beton. Die oben diskutierte neue Methodik zur Analyse von heterogenen Materialien soll beispielhaft auf den Werkstoff Beton angewandt werden. Dieses komplexe Material ist in Abbildung 4 mit seinen verschiedenen Skalen dargestellt. Von links nach rechts findet man die Makroskale des Bauwerks, die im Meterbereich liegt. Dann folgt die Mesoskale, in der die Zuschläge des Betons (Kies) und der Mörtel zu finden sind. Darunter liegt die Mikroskale, in der die Bestandteile des Mörtels, z. B. der Zementstein, auftreten. Die Nanoskale wird hier nicht aufgelöst, obwohl sie für einige Prozesse im Beton von Bedeutung ist.

Von den in Abb. 4 beschriebenen Skalen soll im Folgenden die Mikroskale herausgehoben betrachtet werden. Insbesondere wird der Zementstein betrachtet. Dazu ist sowohl die Kenntnis der mikrostrukturellen Geometrie als auch der Verteilung der unterschiedlichen Materialphasen erforderlich. Beim Zementstein wird das geometrische Modell mittels CT-Aufnahmen des Zementsteins bei einer Auflösung von einem Mikrometer bestimmt, siehe Abb. 5.

Auf dieser Ebene sind keine detaillierten Aussagen über die innere Struktur der einzelnen Hydratisationsprodukte im Zementstein möglich. Es kann aber zwischen hydratisierten und unhydratisierten Bereichen sowie Mikroporen unterschieden werden.

Verschiedene Simulationen sind nun möglich:

1. Eine Homogenisierung der auf der Mikroskale gewonnenen Spannungen und Dehnungen, die dann zu einem Materialmodell auf der Mesoskale führt. Diese Homogenisierungen lassen sich sowohl für den elastischen Teil als auch für inelastischen Materialantworten bestimmen. Um zu quantitativ verwertbaren Ergebnissen zu kommen, sind unterschiedliche RVE einzubeziehen, da die mikrostrukturelle Geometrie der RVE zufällig verteilt ist. Die entsprechenden Berechnungen sind sehr aufwendig, da bis zu 10 000 RVE im Rahmen einer Monte-Carlo-Methode zu berücksichtigen sind.
2. Eine echte Multi-Skalen-Berechnung, bei der Informationen zwischen der Mikrostruktur und der Mesostruktur ausgetauscht werden. Damit ist eine Mi-

Abb. 4: Unterschiedliche Skalen, die bei der Materialbeschreibung von Beton zu berücksichtigen sind

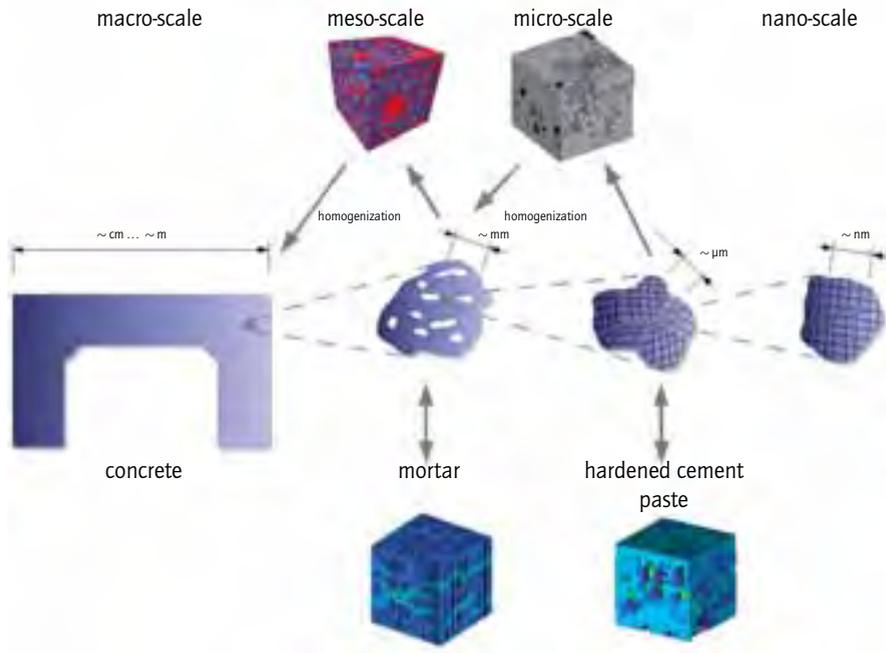
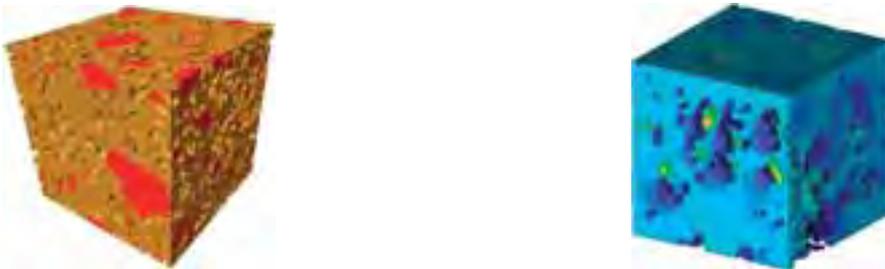


Abb. 5: Mikrostruktur von Zementstein und lokale Spannungen in der Mikrostruktur [6]



kostruktursimulation z. B. an jedem Integrationspunkt der Mesoskala durchzuführen, was in der Regel die heutige Computerkapazität im industriellen Bereich übersteigt. Dennoch wird es in der Zukunft mit steigender Rechnerleistung möglich sein, diese Simulationen erfolgreich durchzuführen. Dies erfordert die Entwicklung von adaptiven Methoden, mit denen die Bereiche spezifiziert werden können, die eine mehrskalige Berechnung erfordern.

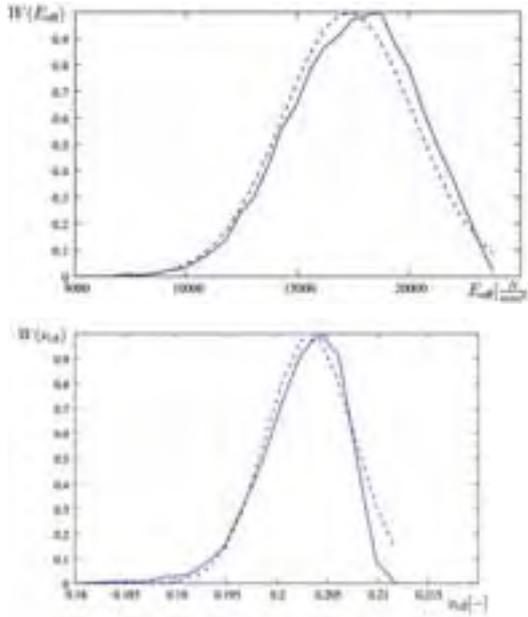
Homogenisierung von Zementstein. Die unter Punkt 1 erwähnte Methode der Homogenisierung soll am Beispiel des Zementsteins genauer erläutert werden.

Basierend auf den in Abb. 4 angegebenen CT-Scans wird ein dreidimensionales Finite-Elemente-Modell entwickelt, mit dem ein RVE mit einer Kantenlänge von 64 μm diskretisiert wird. Dies führt auf Finite-Elemente-Netze mit etwa 300 000 Elementen und etwa 820 000 Unbekannten. Da die elastischen Eigenschaften der hydratisierten und unhydratisierten Bereiche aus Mikroindentationsversuchen bekannt sind, können diese in dem Modell der Mikroskala direkt zur Homogenisierung verwendet werden. Eine Berechnung führt nach Auswertung zu den gemittelten oder effektiven Werten für den Elastizitätsmodul und die Querkontraktion des Zementsteins.

Die Verwendung von 8 200 unterschiedlichen RVE im Rahmen einer Monte-Carlo-Simulation erfordert die Verarbeitung von großen Datenmengen und die Berechnung von 8 200 Mikrostrukturen mit jeweils etwa 800 000 Unbekannten. Dies lässt sich parallel auf PC-Clustern durchführen, wobei jede einzelne Rechnung etwa sechs Minuten auf heutigen PC erfordert. Die Ergebnisse für die effektiven Werte des Elastizitätsmoduls und der Querkontraktion sind in Abb. 6 zusammen mit der entsprechenden Gauss-Verteilung angegeben. Diese Resultate wurden mit am Institut für Baustoffkunde an der RWTH Aachen durchgeführten Experimenten verglichen. Dabei konnte gezeigt werden, dass die durch die numerische Simulation erhaltenen effektiven Materialkennwerte sehr genau sind (Abweichung ungefähr 2 %). Damit ist die vorgestellte Strategie schon heute sehr gut für Vorhersagen von Materialkennwerten einsetzbar, siehe [6].

Frostschädigung von Zementstein. Die Methodik der Homogenisierung kann jetzt auch auf Materialien ausgeweitet werden, die inelastisch reagieren. Als Beispiel soll hier die Frostschädigung des Zementsteins betrachtet werden, die infolge einer Wasserfüllung der Poren auftreten kann, siehe auch [7]. Das mechanische Verhalten des untersuchten Zementsteins lässt sich durch spröde Schädigung und bleibende Verformungen infolge von Mikrorissen charakterisieren. Das mikrostrukturelle Materialmodell muss anhand von experimentellen Ergebnissen verifiziert werden. Daraufhin kann die Frostschädigung, also eine Schädigung der Mikrostruktur infolge der Volumenzunahme wassergefüllter Poren, numerisch simuliert werden. Eine statistische Auswertung für verschiedene Temperaturen und Feuchtegrade und anschließende Homogenisierung führt schließlich zu einem mikrostrukturell begründeten, effektiven Materialmodell, das in Makrostruktursimulationen verwendet werden kann.

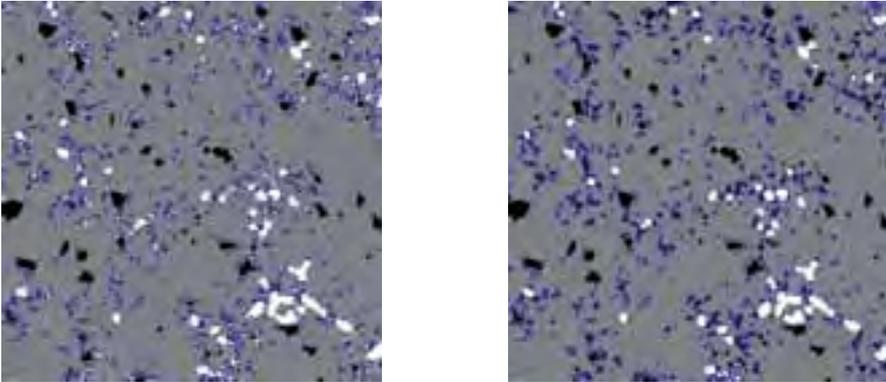
Abb. 6: Wahrscheinlichkeitsdichte für den effektiven Elastizitätsmodul und die effektive Querkontraktionszahl (Gauss-Verteilung gestrichelte Linie) [6]



Für die numerische Simulation von Frostschädigung auf der Mikroebene wird ein numerisches Modell für den Gefrierprozess von Wasser entwickelt, das die Volumenzunahme von 9% beschreibt. Offensichtlich ist der Gefrierprozess temperaturabhängig, und es muss daher ein mikrostrukturelles Temperaturfeld durch eine thermomechanisch gekoppelte Simulation bestimmt werden. Allerdings ist die Mikrostruktur so klein, dass ein etwaiger Temperaturgradient vernachlässigt werden kann. Damit ist die Temperatur in der Mikrostruktur durch einen externen Parameter kontrollierbar. Diese hängt vom aktuellen Lastschritt ab. Entsprechend ist keine thermomechanische Kopplung auf Mikroebene erforderlich, und es kann eine exklusive mechanische Analyse durchgeführt werden.

In Experimenten tritt bei thermischen Belastungszyklen eine Absorption von Wasser auf, die eine kontinuierliche Zunahme des Feuchtegehalts bewirkt. Die Wasserabsorption wird hier nicht simuliert, vielmehr wird eine fast gesättigte Mikrostruktur verwendet, die numerisch unter Verwendung der medialen Achsen generiert wurde. In Abb. 7 sind Schnitte von Mikrostrukturen mit verschiedenen Feuchtegraden w_h dargestellt. Wie zu erwarten ist, bedeckt das blau dargestellte Wasser zunächst den Rand einer Pore und

Abb. 7: Schnitt durch den wassergefüllten Zementstein ($256 \times 256 \mu\text{m}$) für die Feuchtegrade $w_h = 0,70$ und $w_h = 0,80$



danach erst das innere Volumen. In den Abbildungen sind die unhydratisierten Klinkerphasen in schwarz, die Hydratisationsprodukte in grau und die Poren in weiß dargestellt.

Eine Volumenzunahme der wassergefüllten Poren findet statt, sobald die Temperatur unter eine bestimmte Gefriertemperatur θ_f fällt, die vom aktuellen Radius der Pore abhängt. In den CT-basierten Mikrostrukturen variieren die Porenradien von $0.5 \mu\text{m}$ bis $2.0 \mu\text{m}$. Infolge der limitierten Auflösung können kleinere Poren, die einen Radius bis zu 10 nm erreichen, nicht aufgelöst werden. Damit vernachlässigt eine Simulation mittels finiter Elemente den Einfluss kleinerer Poren. Für die numerische Analyse werden die wassergefüllten Poren durch ein linear elastisches Material inklusive Temperaturdehnung beschrieben.

Basierend auf derartigen Konstitutivbeziehungen kann dann die Schädigung innerhalb eines RVE bestimmt werden. Dies ist in Abb. 8 für eine Abkühlung des RVE dargestellt, wobei die helleren Zonen die Bereiche der Schädigung angeben. Basierend auf diesen Ergebnissen kann dann mittels einer Monte-Carlo-Simulation von vielen unterschiedlichen RVE ein effektives Schädigungsgesetz durch Homogenisierung erhalten werden, das die Frostschädigung auf der nächst höheren Skale beschreibt.

Multi-Skalen-Methoden für dreidimensionalen Rissfortschritt. Viele Materialien enthalten Mikrorisse, die sich unter ständig variierenden Belastungen erweitern und so zu Makrorissen führen können. Die Kenntnis des Risswachstums ist wesentlich für die Gewinnung neuer Erkenntnisse im Bereich der Dauerfestigkeit von Materialien. Durch Mikro-Makro-Simulationen unter dynamischen Lastwechseln kann hier ein Fortschritt erzielt werden, der Ingenieuren bei der besseren Auslegung von Konstruktionen hilft. Diese

Abb. 8: Schädigung eines Repräsentativen Volumenelements (RVE) durch Frostangriff in der Mikrostruktur [7]



Berechnungen sind aber immer noch extrem zeitintensiv, da sie auf der Analyse von Finite-Elemente-Diskretisierungen beruhen, die mehrere Millionen an Unbekannten besitzen. Erfolgversprechend sind hier neue Ansätze im Rahmen der *eXtended Finite Element Method* (XFEM). Diese Methodik erlaubt es, den Einfluss von Rissen durch zusätzliche Freiheitsgrade zu berücksichtigen, so dass bei einem Rissfortschritt das vorhandene Finite-Elemente-Netz nicht verändert werden muss.

In Abb. 9 sind neun zufällig verteilte, elliptische Risse in einem Teilvolumen eines Festkörpers dargestellt, das aus einem linear elastischen Material besteht. Dieser Teilkörper ist in vertikaler Richtung einer Zugkraft ausgesetzt, was zu den in Abb. 9 gezeigten Konturen der Normalspannung in vertikaler Richtung führt.

Die Kopplung der XFEM mit einer Multi-Skalen-Analyse ermöglicht dann die Analyse feinerer Strukturen, in denen Makro- und Mikrorisse aufgelöst werden können, wie es in Abb. 10 für einen geraden und gekrümmten Riss dargestellt ist.

Die numerische Simulation der Rissformationen, die insgesamt zu einem Risswachstum des Makrorisses führen können, wurde mit einer Projektionsmethode innerhalb einer Multi-Skalen-Analyse berechnet. Der Makrowürfel wird in vertikaler Richtung gedehnt. Die Farbkonturen stellen die Normalspannungen in vertikaler Richtung dar. Mit diesen Simulationen können sowohl Rissverstärkungen als auch -abschirmungen untersucht und so die Auswirkungen von Mikrorissen auf das Risswachstum in einer dreidimensionalen Struktur geklärt werden.

Abb. 9: Konturen der vertikalen Normalspannung in einem RVE mit neun Mikrorissen und Rissfront in einer Multi-Skalen-Analyse

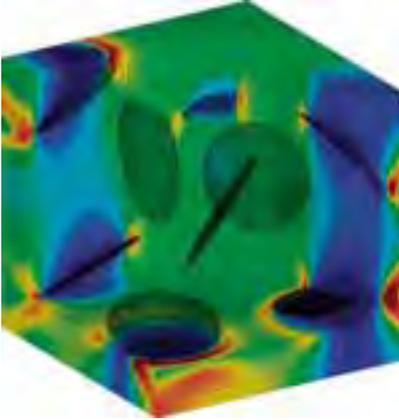
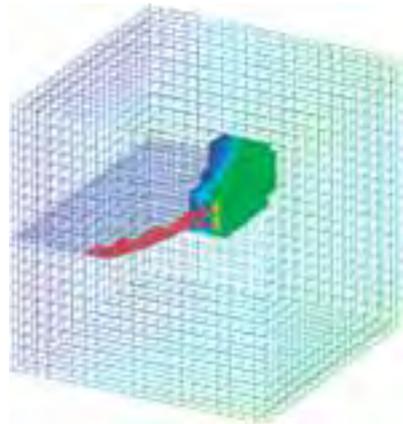
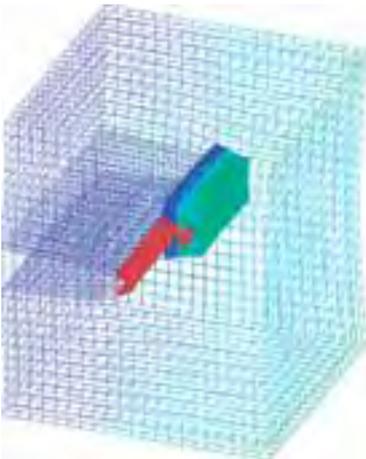


Abb. 10: Rissfront mit Makro- und Mikroriss in einer Multi-Skalen-Analyse



3 ÜBERSICHT UND STATUS QUO

3.1 FAKTEN

Kontinuumsmechanik von Mehrkomponenten-Mehrphasenmaterialien (Multi-field and Multi-physics Materials). Die Kontinuumsmechanik beschreibt das Verhalten von Materialien (Festkörpern, Fluiden und Gasen) auf der Makroskala. Dies trifft auch auf die Theorie Poröser Medien (TPM) und die ihr zugrunde liegende Mischungstheorie (MT) zu. Während die Mischungstheorie als kontinuumsmechanisches Modellierungswerkzeug für mischbare und reaktive Stoffe entwickelt wurde, stellt die Theorie Poröser Medien ein Werkzeug dar, mit dem nicht-mischbare Materialien (z. B. Festkörpergerüst und Porenfluide) im Sinne überlagerter Kontinua mit inneren Wechselwirkungen beschrieben werden, d. h. jeder Teilkörper, der auf dem gemeinsamen Gesamtgebiet nur einen Teil einnimmt, wird über das gesamte Gebiet „verschmiert“, so dass alle am Gesamtkörper beteiligten Teilkörper gleichzeitig und gemeinsam auf dem gesamten Bereich definiert sind. Dies setzt eine tatsächliche oder eine gedachte (virtuelle) Homogenisierung der Teilkörper voraus. Während man in der TPM in der Regel von einer bereits durchgeführten Homogenisierung ausgeht, bietet eine Mehrskalenmodellierung die Möglichkeit, den Homogenisierungsprozess nicht nur virtuell zu betrachten, sondern ihn real im Rahmen einer numerischen Untersuchung abzubilden. Dies kann große Vorteile haben.

Heterogene Materialien oder heterogen aus verschiedenen Stoffen aufgebaute, nicht-mischbare Materialien besitzen eine Mikrostruktur, die einen großen Einfluss auf die Abbildung makroskopischer Prozesse hat. Dies soll an einem einfachen Beispiel verdeutlicht werden: Betrachtet man zum Beispiel einen Schwamm, der aus einem offenporigen Schaum besteht, so wird dieser bei gleicher Porosität (Porenvolumen pro Gesamtvolumen) völlig andere Durchlässigkeitseigenschaften besitzen; und zwar je nachdem, ob er großporig oder feinporig strukturiert ist. Auf der Makroskala wird man für beide Schwämme im Labor einen Durchlässigkeitsbeiwert ermitteln, der zur ingenieurmäßigen Berechnung von gekoppelten Deformations- und Strömungsprozessen erforderlich ist. Die Mikrostruktur der Schwämme beeinflusst aber nicht nur die Durchlässigkeitseigenschaften, sondern sie beeinflusst auch die Materialeigenschaften des Schwammes selbst. Gehen wir für dieses Beispiel einmal davon aus, ein offenporiger Schaum mit 90-prozentiger Porosität (Abb. 11) und viskoelastischen Materialeigenschaften wird gestaucht, so treten auf der Makroskala große Deformationen in Deformationsrichtung auf, während die Deformationen auf der Mikroskala vor allem durch lokales Biegen und Knicken geprägt sind, vgl. Abb. 12.

Im Druck-Stauchungs-Diagramm (Abb. 12) erkennt man zunächst einen Anfangsbereich, der durch Biegung auf der Mikroskala geprägt ist. Anschließend folgt ein Bereich, der durch lokales Knicken gekennzeichnet ist, d. h. auf der Mikroskala knicken die Schaumstoffstege aus, und die Schaumstofflagen „stapeln“ sich übereinander. An diesen Bereich schließt sich der Verdichtungsbereich an. Hier verhärtet sich das Material zunehmend, bis sich schließlich trotz weiterer Laststeigerung keine weitere Verfor-

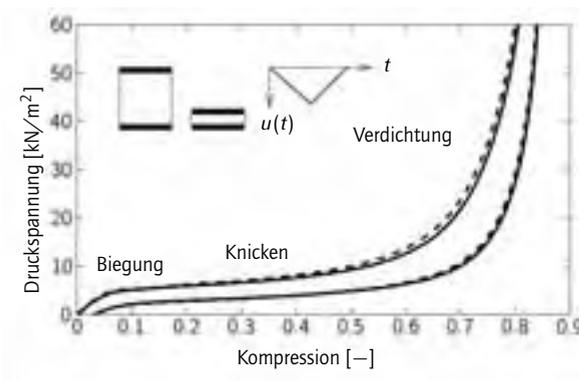
mung einstellen kann (vertikale Tangente). Der Grund für dieses makroskopisch auffällige Verhalten ist wiederum auf der Mikroskala zu finden. Polyurethan ist materiell inkompressibel, d. h. das Material reagiert auf einen allseitigen Druck nicht mit einer Volumenreduktion. Während der Polyurethanschaum zunächst weich erscheint (Biege- und Knickbereich), verhärtet sich der Schaum durch Schließen der Poren zunehmend, bis alle Schaumstofflagen „gestapelt“ sind; der Porenraum ist geschlossen. Der zunächst stark kompressible Schaum ist inkompressibel geworden; das Verhalten auf der Mikroskala entspricht demjenigen auf der Makroskala.

Das Beispiel des offenporigen Polyurethanschaums zeigt, dass auf der Makroskala jede Änderung der Porosität und jede Änderung der Porengeometrie (großporig vs. kleinporig) ein anderes makroskopisches Ausgangsmaterial erzeugt, für das zur Bestimmung

Abb. 11: Probekörper aus Polyurethanschaum (links) mit offenporiger Mikrostruktur (rechts)



Abb. 12: Druckversuch an einem Polyurethanwürfel, Druckspannung über Kompression [Längenänderung pro Ausgangslänge]. Die unterschiedlichen Versuchskurven weisen auf viskoelastisches Materialverhalten hin (oben: Belastungskurve; unten: Entlastungskurve).



der makroskopischen Materialparameter jeweils neue Versuche und Versuchsauswertungen erforderlich sind. Da wir in unserem Gedankenmodell jedoch immer davon ausgehen wollen, dass Polyurethan mit jeweils denselben intrinsischen Materialeigenschaften verwendet wird, d. h. die Materialeigenschaften auf der Mikroskala sind in jedem Fall dieselben, können die makroskopischen Materialeigenschaften im Rahmen eines numerischen Homogenisierungsprozesses über geeignete RVE gewonnen werden. Dies setzt die Kenntnis der inneren Geometrie des Materials voraus, also die Kenntnis der Geometriedaten auf der Mikroskala. Damit entfällt der unter 3.1 genannte Optimierungsprozess zur Generierung der Materialparameter auf der Makroebene; allerdings entstehen durch den Homogenisierungsprozess mindestens gleichwertige numerische Kosten.

Neben der Berechnung von Materialeigenschaften mittels numerischer Rechnungen auf der Mikroskala sind bei Vorhandensein ausreichender Rechnerkapazitäten auch sogenannte FEM²-Rechnungen denkbar. Bei dieser Vorgehensweise löst man ein makroskopisches Anfangs-Randwertproblem (ARWP), ohne ein makroskopisches Stoffgesetz zu verwenden. Im Rahmen der Finite-Elemente-Methode bedeutet dies, dass man an den Gauss-Punkten der finiten Elemente kein Stoffgesetz auswertet, sondern stattdessen eine weitere FE-Rechnung im Sinne eines RVE einbettet, dessen homogenisiertes Ergebnis die Auswertung eines makroskopischen Stoffgesetzes ersetzt. Obwohl diese Vorgehensweise auf den ersten Blick zur Lösung eines einzelnen ARWP zu aufwendig erscheinen mag, hat sie jedoch einen entscheidenden Vorteil. Es entsteht ein numerisches Labor, mit dem man im Sinne eines Optimierungsprozesses beliebig viele Mikrostrukturen in die Makrostruktur einbetten kann, um auf diese Weise die optimale Lösung eines komplexen Problems zu generieren oder um die optimale Zusammensetzung eines inhomogenen Materials im Sinne so genannter „taylorred materials“ zu finden.

Abb. 13: Druckversuch an einem Polyurethanwürfel. Von links nach rechts: Biegebereich, Knickbereich und Verdichtungsbereich.



Viele weitere Beispiele, die die zukunftsweisenden Möglichkeiten der Kontinuumsmechanik von Mehrkomponenten-Mehrphasenmaterialien (Multi-field and Multi-physics Materials) sowie von Simulationstechniken zur Beschreibung von heterogenen Materialien (Multi-scale Modelling) zeigen, sind denkbar und könnten hier angeführt werden.

Alle gezeigten Beispiele machen deutlich, dass Probleme mit multiphysikalischen Eigenschaften immer auf ein System stark gekoppelter partieller Differentialgleichungen führen, das generell monolithisch gelöst werden muss. Hier sind jedoch auch Strategien denkbar, bei denen der Kopplungsgrad des Gesamtsystems mit dem Ziel geprüft wird, ob es sinnvoll ist, eine Operatorsplit-Methode zur Reduktion des Gesamtaufwands bei der Lösung des Gleichungssystems einzusetzen, ohne die Konvergenzeigenschaften des Gesamtsystems negativ zu beeinflussen. Im Rahmen der FEM setzen volumetrisch gekoppelte Festkörper-Fluid-Probleme zur Erfüllung von Stabilitätskriterien wie der Ladysenskaja-Babuška-Brezzi-Bedingung (LBB-Bedingung) darüber hinaus in der Regel die Verwendung von Taylor-Hood-Elementen voraus, bei denen ein quadratischer Ansatz für die Festkörperverschiebung mit einem linearen Ansatz für den Porendruck und für weitere Freiheitsgrade kombiniert wird. Dies führt insbesondere bei geometrisch dreidimensionalen Problemen zu sehr großen Gleichungssystemen, deren Lösung auch heute noch eine große Herausforderung darstellt.

Die Analyse komplexen Materialverhaltens stellt die Modellierer aber vor weitere Probleme. Die in den komplizierten Stoffgesetzen enthaltenen Materialparameter entziehen sich häufig einer direkten Bestimmung, d. h. sie lassen sich nicht direkt sondern nur indirekt ermitteln. Dazu sind Optimierungsstrategien und die Methoden des Inversen Rechnens erforderlich. Alternativ zu dieser Vorgehensweise können Mehrskalenanalysen genutzt werden. Dies wird im Folgenden weiter ausgeführt.

Simulationstechniken zur Beschreibung von heterogenen Materialien (Multi-scale Modelling). Die Berücksichtigung von kleinen Skalen in einem numerischen Simulationsmodell führt zu sehr aufwendigen Diskretisierungen, da für eine Vorhersage von quantitativen Ergebnissen immer eine dreidimensionale Mikrostruktur zu wählen ist. Dies führt zu hochdimensionalen Gleichungssystemen, die nur mittels spezieller mathematischer Verfahren zu lösen sind. Hierzu gehören iterative Gleichungslöser, die jedoch an die speziellen Problemstellungen angepasst werden müssen, um optimale Simulationszeiten zu erzielen. Durch die Größe der Gleichungssysteme sind häufig parallele Multigrid-Verfahren einzusetzen, wobei für anwendende Ingenieure algebraische Multigrid-Verfahren wegen ihrer Unabhängigkeit von speziellen Netzstrukturen vorteilhaft sind.

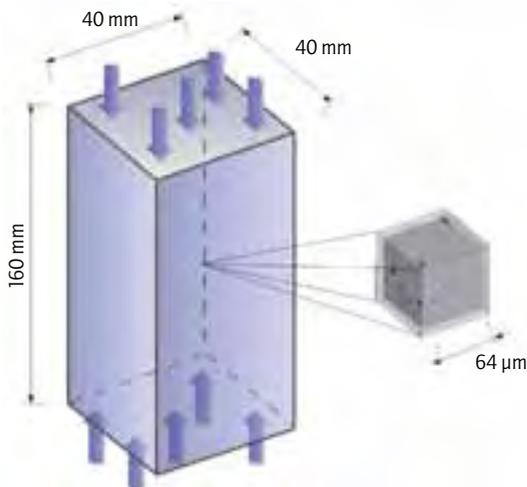
Die in den Substrukturen der RVE vorkommenden lokalen Verzerrungen sind häufig sehr groß, so dass robuste finite Elemente zum Einsatz kommen müssen. Hier besteht auch heute noch die Notwendigkeit, neue robuste Elemente speziell im Bereich der nicht-linearen Verformungen zu entwickeln.

Ebenso wie bei multiphysikalischen Problemen sind auch bei den Multi-Skalen-Analysen oft noch Parameteridentifikationen von nicht messbaren konstitutiven Eigen-

schaften auf der Mikrostruktur vorzuschalten. Dies führt auf inverse Probleme, bei denen große Gleichungssysteme gelöst werden müssen, was dann wegen vieler Funktionsauswertungen innerhalb der zugrunde liegenden Optimierungsstrategien zu großem Rechenaufwand führt. Auch hier sind parallele Berechnungsverfahren unabdingbar, wenn die Lösungen in für Ingenieure vertretbaren Rechenzeiten bestimmt werden sollen.

Der Nachteil und die damit verbundene Herausforderung dieser Methodik sind die extrem hohen Anforderungen an die Rechnerleistung. Dies folgt aus der Dimension des Mikrostrukturmodells, das auf Gleichungssysteme mit etwa 1 Million Freiheitsgraden führt. Diese Mikrosysteme sind an allen Integrationspunkten des Makromodells zu analysieren, wobei anzunehmen ist, dass die innere Struktur der Mikromodelle zufällig verteilt ist und somit kein Mikrosystem dem anderen gleicht. Da das Makromodell in der Regel mehr als eintausend Integrationspunkte aufweist, folgt, dass über tausend Mikrosysteme in jedem Rechenschritt der Analyse des Probenkörpers auszuwerten sind. Da diese in der Regel bei nichtlinearem Materialverhalten eingesetzt werden, bedeutet dies auch eine nichtlineare Simulation der Mikrosysteme. Wenn man sich jetzt vorstellt, dass das Makrosystem bei einer praxisrelevanten Berechnung nicht aus eintausend finiten Elementen sondern aus einer Million finiter Elemente besteht, dann ist klar, dass zur Zeit selbst der Einsatz modernster Rechnerarchitekturen nicht ausreicht, derartige Simulationen in vertretbaren Zeiträumen zu erledigen. Hier besteht also großer Forschungsbedarf,

Abb. 14: Zweiskalige Simulation eines Probenkörpers (siehe [7])



der sich auf paralleles Rechnen, Gleichungslösung und die Entwicklung robuster finiter Elemente für automatisierte nichtlineare Berechnungen bezieht.

3.2 ZUSAMMENFASSUNG

Die vorangegangenen Ausführungen haben gezeigt, dass die moderne Mechanik dazu beitragen kann, komplexe Probleme der Materialmodellierung unter Beachtung multiphysikalischer Eigenschaften auf allen Skalen zu beschreiben. Die moderne Mechanik ist auch in der Lage, große Anfangs-Randwertproblemen zu formulieren und einer numerischen Lösung zuzuführen. Die dabei auftretenden Probleme sind angesprochen worden und beziehen sich hauptsächlich auf Fragen der Analyse der Gleichungssysteme und der Konstruktion geeigneter Gleichungslöser einschließlich der erforderlichen Vorkonditionierer. Darüber hinaus kommt dem Problem der Parameteroptimierung und des damit verbundenen Inversen Rechnens eine zunehmend größere Bedeutung zu. Dies macht noch einmal klar, dass die Probleme der modernen Mechanik nicht von Mechanikern allein bewältigt werden können, sondern dass dies nur im Zusammenspiel mit Mathematikern zielführend möglich ist. Mechaniker und Mathematiker müssen daher auf dem von Ludwig Prandtl und Richard von Mises bei der Gründung der GAMM beschrittenen Weg weiter voranschreiten, um die Probleme der Zukunft gemeinsam zu lösen.

4 STÄRKEN/SCHWÄCHEN-ANALYSE, HERAUSFORDERUNGEN

Im vorangegangenen Abschnitt wurde auf der Basis von nachvollziehbaren Fakten erläutert, welches Potential die moderne Mechanik insbesondere im Zusammenwirken mit der Mathematik für die Herausforderungen der Zukunft besitzt. Dies soll hier vertieft werden. Dabei werden auch Stärken und Schwächen angesprochen und die Herausforderungen der Zukunft genannt.

4.1 AUSGANGSBASIS UND VISIONEN

Heute gibt es äußerst leistungsfähige Rechner, auf denen Mehrkomponentenmaterialien mit multiphysikalischen Eigenschaften ebenso analysiert werden können wie Mehrskalprobleme. Dies war noch vor 20 Jahren völlig anders. Die Mischungstheorie und die Theorie Poröser Medien sind zwar schon lange Elemente der erweiterten Kontinuumsmechanik (die Mischungstheorie seit ca. 1960, die TPM seit ca. 1980); allerdings sind neben den Fundamentalgleichungen (Kinematik und Bilanzgleichungen superponierter Kontinua) anfänglich kaum nennenswerte Materialmodelle entwickelt worden, die einer numerischen Simulation hätten zugeführt werden können. Eine sinnvolle numerische Analyse derart komplexer Materialmodelle war einfach nicht möglich. Dasselbe gilt für Mehrskalmodelle, bei denen die physikalischen Eigenschaften auf mehreren Skalen

numerisch analysiert und mittels verschiedener Homogenisierungsverfahren verglichen und in Bezug zueinander gesetzt werden.

In der modernen Mechanik nehmen beide Elemente, die Kontinuumsmechanik von Mehrfeldproblemen und von multiphysikalischen Problemen ebenso wie die Analyse von Mehrskalenproblemen, eine wesentliche Stellung ein. Diese Elemente, die wie oben beschrieben leicht kombiniert werden können, besitzen ein gewaltiges Potential für die vor uns liegenden Aufgaben. Eine konsequente Durchbildung komplexer Probleme mit den Methoden der TPM kombiniert mit Mehrskalenanalysen bietet die Möglichkeit, zukünftig Aufgaben zu lösen, die heute unlösbar erscheinen. Dies soll im folgenden an einem Beispiel verdeutlicht werden.

Der Mensch war immer schon daran interessiert zu wissen, was ihn „im innersten zusammenhält“. Insofern ist es nicht verwunderlich, dass gerade die Biomechanik eine sehr hohe Aufmerksamkeit genießt, und dies nicht nur unter Fachkollegen sondern in der gesamten Gesellschaft. Durchdringt man visionär die zukünftigen Möglichkeiten der Biomechanik und vergleicht sie mit dem, was heute machbar und möglich ist, so stellt man schnell fest, dass heute lediglich Detailprobleme innerhalb der Biomechanik behandelt werden. So gibt es Arbeitsgruppen, die den Bewegungsapparat des Menschen mit den Methoden der Starrkörperdynamik untersuchen. Diese Arbeitsgruppen sind insbesondere im Bereich der Sportwissenschaften oder der Sportmedizin angesiedelt. Auf der anderen Seite gibt es Arbeitsgruppen, die Teile des menschlichen Körpers wie Knochen, Knorpel, innere Organe oder das vaskuläre System untersuchen und numerisch beschreiben. Diese Arbeitsgruppen gehören in der Regel dem Bereich „Kontinuumsmechanik“ an und sind überwiegend an Bauingenieur- und Maschinenbauakademien angesiedelt. Gelingt es, die in den unterschiedlichen Bereichen vorhandenen Expertisen zu bündeln und zu einem gemeinsamen, mehrskalig strukturierten Modell zusammenzufassen, dann

Abb. 15: Crashversuch mit Dummy (Quelle: <http://www.euroncap.com>)



sind wir auf dem Weg, ein „Menschmodell“ zu generieren, das den menschlichen Körper in der Gesamtheit seiner mechanischen Eigenschaften beschreibt. Dieses Modell könnte auf der Grobskala des Bewegungsapparats die Methoden der Starrkörpermechanik nutzen, auf der Mesoskala Knochen, Knorpel und Organe unter Berücksichtigung des vaskulären Systems mit den Methoden der Kontinuumsmechanik und insbesondere der TPM beschreiben und auf der Mikroskala zellmechanische Untersuchungen anstellen. Alle Skalen könnten durch Mehrskalenanalysen verbunden werden.

Mit einem Menschmodell könnte man zum Beispiel Unfallfolgen analysieren, indem man auf der Grobskala mit Hilfe eines Crashdummys eine erste Untersuchung startet und diese auf der Mesoskala an solchen Stellen fortsetzt, an denen bestimmte kritische Belastungen überschritten werden. Numerische Untersuchungen dieser Art könnten auch genutzt werden, um bei der Konstruktion der äußeren Hülle von Fahrzeugen mögliche Unfallfolgen mit Fußgängern, Fahrradfahrern und anderen Verkehrsteilnehmern zu mindern. Bereits heute lassen sich patientenspezifische Implantate an Hüfte, Knie oder Wirbelsäule herstellen, bei denen Formoptimierungen zur Minimierung von Luxationswahrscheinlichkeiten berücksichtigt werden. Darüber hinaus könnte man zukünftig in Kooperation mit Chirurgen operative Eingriffe am Computer planen und virtuell durchführen. Dazu müsste ein im Rechner hinterlegtes Menschmodell mit individuellen Patientendaten abgeglichen werden, vgl. Abb. 3. Die Vision des computerorientierten Menschmodells wird auch von den Wissenschaftlern des Exzellenzclusters Simulation Technology getragen (<http://www.simtech.uni-stuttgart.de>).

Sollen solche Rechnungen aber in der Medizintechnik operationsbegleitend durchgeführt werden, dann sind komplexe Simulationen und Visualisierungen in Echtzeit erforderlich. Dies ist auf heutigen Rechnern nicht möglich.

4.2 WIE KOMMT DIE MATHEMATIK INS SPIEL?

Komplexe kontinuumsmechanische Probleme führen immer auf ein System gekoppelter partieller Differentialgleichungen. Ein solches System kann nur in Einzelfällen analytisch gelöst werden, d. h. es sind Näherungslösungen erforderlich. Mit der Methode der Endlichen Elemente steht dazu ein geeignetes Werkzeug zur Verfügung. Die Numerische und Computerorientierte Mathematik ist am Ball, wenn multiskalige und multiphysikalische numerische Probleme behandelt werden sollen. Dies schließt die Quantifizierung der in den Modellen vorhandenen Unsicherheiten genauso ein wie eine selbstadaptive Wahl von Skalen und hinterlegten Physiken, um dynamische und gekoppelte Prozesse sowohl realer als auch virtueller Natur zu behandeln.

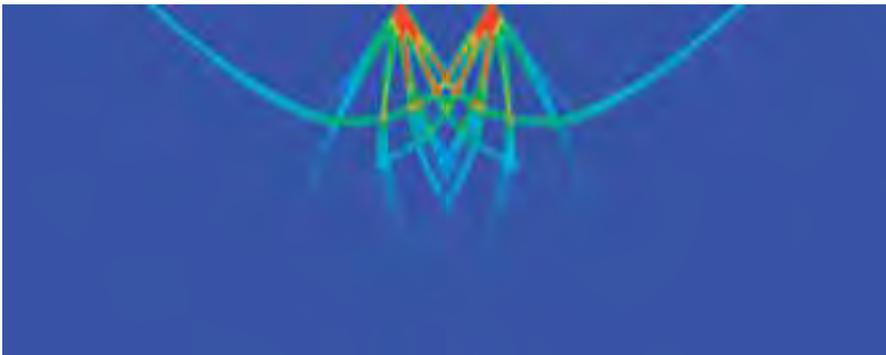
Reale Probleme sind räumlich dreidimensional. Dies führt gerade bei der Behandlung von gekoppelten Problemen im Rahmen der FEM zu einer gewaltigen und nahezu unüberschaubaren Menge an Freiheitsgraden, denen entsprechende Gleichungen gegenüberstehen. Große und größte Probleme können auch heute noch nicht auf Einprozessormaschinen sondern nur auf massiv parallelen Systemen gelöst werden. Außerdem

sind adaptive Strategien erforderlich; und zwar sowohl im Zeitbereich als auch im Raum-bereich, vgl. [1, 2]. Zusätzlich müssen Gebietszerlegungsmethoden angewandt werden, um das Gesamtproblem zu segmentieren und es so einer massiv parallelen Berechnung zuzuführen. Darüber hinaus müssen schnellere iterative Löser mit geeigneten Vorkonditionierern gefunden werden, um große Probleme in akzeptabler Zeit zu behandeln.

Komplexe mechanische Probleme basieren in der Regel auf komplexen Materialmodellen, bei denen die erforderlichen Materialparameter nicht direkt, sondern nur mittels Inversen Rechnens und mit Hilfe von Optimierungsverfahren bestimmt werden können. Dies ist zum Beispiel der Fall, wenn Scherbandlokalisierungen [8] behandelt werden sollen, die im Rahmen eines Standardkontinuums (Boltzmann-Kontinuum) häufig auf ein mathematisch schlecht gestelltes Problem führen. Dies trifft zum Beispiel auf das Grundbruchproblem zu (Abb. 16), bei dem ein elastisch-plastisch deformierbarer Baugrund infolge der Belastung durch ein Bauwerk beginnt, Scherbänder auszubilden, die dann zu einer Instabilität der Konstruktion bis hin zum Kollaps führen können. Ohne das elastisch-plastische Stoffgesetz zu verändern, kann das schlecht gestellte Problem nur durch Einsatz eines erweiterten Kontinuums in ein gut gestelltes Problem verwandelt werden. Geht man das Grundbruchproblem zum Beispiel nicht mit einem Boltzmann- sondern mit einem Cosserat-Kontinuum an, so erhält man zwar ein gut gestelltes Problem, jedoch um den Preis, die zusätzlich auftretenden Materialparameter (Cosserat-Parameter) bestimmen zu müssen.

Das Cosserat-Kontinuum [1] ist eine Erweiterung des Boltzmann-Kontinuums um rotatorische Freiheitsgrade, denen zusätzliche Momentenspannungen zugeordnet sind. Diese Freiheitsgrade und damit auch die über ein Stoffgesetz mit diesen Freiheitsgraden gekoppelten Momentenspannungen werden allerdings erst bei Auftreten der Scher-

Abb. 16: Scherbandbildung mit Grundbruch (FE-Rechnung). Regularisierung des schlecht gestellten Problems mit Hilfe eines um rotatorische Freiheitsgrade erweiterten Kontinuums (Cosserat-Kontinuum).



bandlokalisierung aktiviert [9]. Dies hat zur Folge, dass nur die Materialparameter des Boltzmann-Kontinuums mit den üblichen Optimierungsverfahren bestimmt werden können. Die Cosserat-Parameter müssen durch Inverses Rechnen, d. h. durch Invertieren des gestellten ARWP bestimmt werden, vgl. z. B. [9]. Hier wird die Lösung des Problems, charakterisiert durch die Scherbandinitiierung sowie die Scherbandrichtung und die Scherbanddicke, vorgegeben, um das Gesamtproblem bei gegebenen Materialparametern des Boltzmann-Kontinuums nach den gesuchten Cosserat-Parametern aufzulösen.

Alle hier angesprochenen Probleme, wie im übrigen auch viele weitere, können nur im Zusammenwirken von Ingenieuren mit Mathematikern erfolgreich behandelt werden. Neben diesem unmittelbaren Zusammenhang komplexer Probleme und mathematischer Lösungsansätze müssen riesige Datenmengen verarbeitet und visualisiert werden. Hier sind neben der Mathematik auch die Informatik und die Informationsverarbeitung gefragt.

4.3 STÄRKEN, SCHWÄCHEN UND HERAUSFORDERUNGEN

Aus heutiger Sicht lässt sich feststellen, daß die moderne Mechanik den Herausforderungen der Zukunft gewachsen ist. Allerdings werden die Modelle zur Erfassung realistischer Aufgaben immer komplexer. Dies führt zwangsläufig zu immer größeren Problemen, die mit den Methoden der Computerorientierten Mechanik und der Numerischen Mathematik behandelt werden müssen.

Die zu den weiter oben angeführten Anwendungen und Beispielen gehörenden Berechnungen benötigen schnelle iterative Löser, die für große dreidimensionale Systeme geeignet sind. Die zur Diskretisierung verwendeten Elemente müssen bei nichtlinearen Problemstellungen robust sein und die erforderliche Stabilität besitzen. Dieses Gebiet ist im Bereich der linearen Theorie zu großen Teilen erforscht, aber offen, wenn es sich um nichtlineare Simulationen handelt. Weiterhin sind adaptive Verfahren in der Kopplung mit mehrskaligen Berechnungen zu entwickeln, vgl. z. B. [10], um Berechnungen, wie in Abb. 14 gezeigt, für ingenieurpraktische Anwendungen durchführen zu können. Hier ist auf jeder Stufe eine optimale Lösung in Bezug auf die Rechenzeit (Kosten) anzustreben. Somit sind seitens der Ingenieure und der Mathematiker noch erhebliche Anstrengungen erforderlich, um Modellbildung, Validierung und theoretische Untermauerung voranzutreiben.

Im Gegensatz zu den derzeit verwendeten Modellen, die fast ausnahmslos deterministisch orientiert sind, sind in der existierenden Wirklichkeit unserer Welt jedoch auch stochastische Elemente vorhanden, die alle Aussagen, die wir mit deterministischen Methoden treffen, nur im Rahmen gewisser Unsicherheiten real erscheinen lassen. Die Einbeziehung von Unsicherheiten in unsere Modelle ist daher in vielen Fällen unbedingt erforderlich. Betrachten wir noch einmal das Beispiel der Berechnung der CO₂-Sequestrierung (Abb. 2), dann ist völlig klar, dass die zur Verfügung stehenden Daten des Sequestrierungsgebiets nur punktuell zur Verfügung stehen. Die zwischen den

Messstellen liegenden Gebiete werden datentechnisch nicht erfasst, stattdessen werden Annahmen über die Datenlage getroffen. In solchen Fällen müssen statistische Methoden benutzt werden, um mit unsicheren Daten umzugehen, die in unseren Modellen zu Unsicherheiten in den Geometriedaten und Materialparametern führen.

In der Zukunft werden am Rechner virtuelle Welten geschaffen, die als Prognosewerkzeug für reales Handeln eingesetzt werden. Die heutigen und zukünftigen Herausforderungen für die Realisierung derartiger Berechnungen und Visualisierungen liegen ebenso in der Verkürzung von Entwicklungs- und Rechenzeiten wie in der Förderung des Nachwuchses im Bereich der Ingenieurwissenschaften und der Mathematik. Schließlich kann es nicht nur „User“ geben, irgend jemand muss auch zukünftig in der Lage sein, komplexe Modelle zu verstehen und weiterzuentwickeln, um mit ihnen komplexe Berechnungen durchzuführen.

Nur so kann in der Zukunft eine numerische Simulationsumgebung geschaffen und ständig gepflegt und erweitert werden, die zuverlässig Vorhersagen von komplexen Ingenieurmodellen erlaubt.

5 HANDLUNGSEMPFEHLUNGEN

Die moderne Mechanik wird ihren Beitrag für die Aufgaben der Zukunft leisten. Insbesondere im Zusammenspiel mit der Mathematik und der Informatik wird die Mechanik diesen Aufgaben gewachsen sein. Visionäre Lösungsansätze gibt es in ausreichender Menge, zum Beispiel das unter 4.1 beschriebene „Menschmodell“, siehe auch www.simtech.uni-stuttgart.de. Die Verwirklichung dieser Visionen braucht jedoch nicht nur Zeit, sondern auch engagierte und leistungsbereite Wissenschaftler. Insofern ist es erforderlich, junge Menschen für die Mathematik und für die Mechanik zu begeistern, ihre Leistungsbereitschaft zu entfachen und mit ihnen die Zukunft zu gestalten. Dazu sind auch Werbemaßnahmen an Schulen und anderen Einrichtungen erforderlich, an denen man junge Menschen erreicht, um ihnen die Tätigkeitsfelder von Mechanik und Mathematik und deren wissenschaftliche, industrielle, wirtschaftliche und nicht zuletzt deren gesellschaftliche Dimension nahe zu bringen. Die GAMM und ihre Fachausschüsse werden das Nötige dazu beitragen.

6 LITERATUR

- [1] W. Ehlers: Foundations of Multiphasic and Porous Materials. In W. Ehlers and J. Bluhm (eds.): *Porous Media: Theory, Experiments and Numerical Applications*, Springer, Berlin 2002, pp. 3–86.
- [2] C. Wieners, M. Ammann, T. Graf and W. Ehlers: Parallel Krylov methods and the application to 3-d simulations of a triphasic porous media model in soil mechanics. *Computational Mechanics* 36 (2005), 409–420.

- [3] W. Ehlers, N. Karajan, B. Markert: A porous media model describing the inhomogeneous behaviour of the human intervertebral disc. *Materials Science and Materials Technology (Materialwissenschaft und Werkstofftechnik)* **37** (2006), 546–551.
- [4] T. I. Zohdi and P. Wriggers: *Introduction to Computational Micromechanics*. Springer, Berlin 2005.
- [5] S. Löhnert and P. Wriggers: Effective behaviour of elastic heterogeneous thin structures at finite deformations. *Computational Mechanics* **41** (2008), 595–606.
- [6] M. Hain and P. Wriggers: On the numerical homogenization of hardened cement paste. *Computational Mechanics* **42** (2008), 197–212.
- [7] M. Hain and P. Wriggers: Computational homogenization of micro-structural damage due to frost in hardened cement paste, *Finite Elements in Analysis and Design* **44** (2008), 233–244.
- [8] W. Ehlers, T. Graf, and M. Ammann: Deformation and localization analysis in partially saturated soil. *Computer Methods in Applied Mechanics and Engineering* **193** (2004), 2885–2910.
- [9] W. Ehlers and B. Scholz: An inverse algorithm for the identification and the sensitivity analysis of the parameters governing micropolar elasto-plastic granular material. *Archive of Applied Mechanics* **77** (2007), 911–931.
- [10] I. Temizer and P. Wriggers: An adaptive method for homogenization in orthotropic nonlinear elasticity. *Computer Methods in Applied Mechanics and Engineering* **196** (2007), 3409–3423.

> MATHEMATIK FÜR WERKZEUGMASCHINEN UND FABRIKAUTOMATISIERUNG

BEREND DENKENA, DIETMAR HÖMBERG UND ECKART UHLMANN

1 EXECUTIVE SUMMARY

Der Artikel beschreibt Stand und Perspektive des Einsatzes mathematischer Methoden bei der Simulation und Auslegung von Werkzeugmaschinen und der Fabrikautomatisierung. Erfolgreiche Beispiele für den Einsatz der Mathematik finden sich vor allem bei der Modellierung von Werkzeugmaschinen von Mehrkörpersystemen und der Simulation einzelner Produktionsprozesse.

Die Autoren diskutieren wichtige aktuelle mathematische Konzepte für die Produktionstechnik. Sie zeigen, dass eine durchgreifende mathematische Modellierung notwendig ist für die immer wichtiger werdende Beschreibung ganzer Prozessketten. Für die numerische Simulation sind adaptive Algorithmen häufig ein unverzichtbares Hilfsmittel. Diese führen zu einer drastischen Reduktion der Anzahl der Unbekannten ohne den Approximationsfehler zu vergrößern, wodurch bei komplexen Werkstückgeometrien eine numerische Simulation häufig überhaupt erst möglich wird. Das ultimative Ziel von Modellierung und Simulation ist in der Regel die Berechnung optimaler Prozessparameter. Mathematisch bedeutet das die Lösung eines so genannten Optimalsteuerungsproblems. Besonders erfolgreich lässt sich die Optimalsteuerung in Kombination mit einer maschinenseitigen Prozesssteuerung einsetzen.

Große Herausforderungen für die Zusammenarbeit von Produktionstechnik und Mathematik liegen in der Untersuchung der Wechselwirkungen zwischen Prozess, Maschine und Werkstück. Am Beispiel der Fräsbearbeitung und der Lasermaterialbearbeitung wird diese Problematik näher erläutert.

Durch eine stetig wachsende Modellvielfalt bei gleichzeitig wachsendem Druck zu kürzeren Entwicklungszeiten gibt es aus industrieller Sicht einen großen Bedarf an der Entwicklung von Tools für eine automatisierte Neuplanung oder Rekonfiguration komplexer Produktionsanlagen. Dazu bedarf es der Entwicklung eines mehrskaligen mathematischen Standardmodells für die Fabrikautomatisierung. Dieses könnte in einem interdisziplinären Schwerpunktprogramm von Produktionstechnikern, Informatikern und Mathematikern gemeinsam in Angriff genommen werden.

2 ERFOLGSGESCHICHTEN

2.1 MATHEMATIK FÜR DIE ENTWICKLUNG NEUER MASCHINENKONZEPTE: PARALLELKINEMATIKEN

In den letzten Jahren hat sich die Werkzeugmaschinenbranche durch den Kosten- und Qualitätsdruck ihrer Kunden verstärkt mit der Suche nach neuen Maschinenkonzepten beschäftigt. Maschinen mit Parallel- und Hybridkinematik kam dabei eine besondere Bedeutung zu. Gegenüber konventionellen Maschinen können Parallelkinematiken Vorteile in der Steifigkeit und Dynamik besitzen. Weiterhin erfolgt keine Superposition der in den Achsen entstehenden Fehler. Durch eine Verwendung von Gleichteilen reduzieren sich die Fertigungskosten.

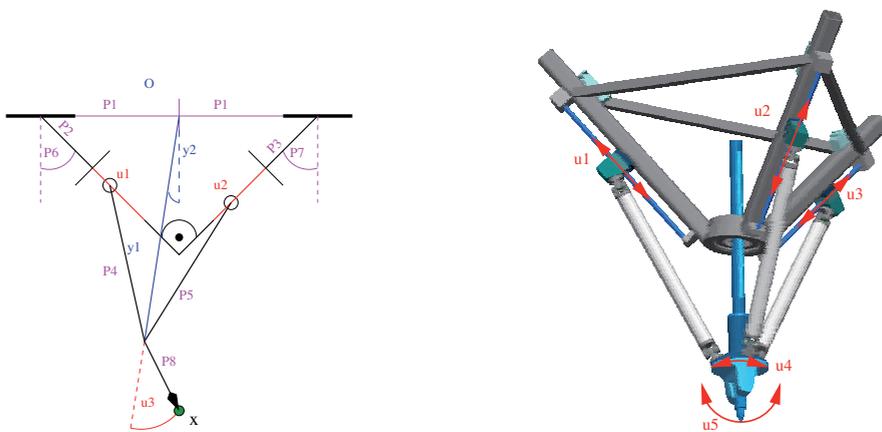
Trotz dieser Vorteile herrscht nach wie vor Zurückhaltung hinsichtlich ihres industriellen Einsatzes. Ein grundlegendes Hemmnis besteht in den Schwierigkeiten bei der Beherrschung der Komplexität des gekoppelten kinematischen Positionier- und Bewegungsverhaltens. Geometrische Fehler (aufgrund von Fertigungsungenauigkeiten der Maschinenelemente bzw. Montagefehlern), Gravitations- und Trägheitseinflüsse sowie thermisch bedingte Deformationen der Strukturelemente bewirken Abweichungen zwischen dem theoretischen, durch die Konstruktion vorgegebenen kinematischen Übertragungsverhalten und dem tatsächlichen, siehe [18]. Die Identifikation, rechnerische Abbildung und Kompensation dieser Störeffekte ist Aufgabe der Kalibrierung.

Im Rahmen des Forschungsprojektes *Kalibrierverfahren für hybride Parallelkinematiken*¹ wurde ein allgemeingültiges Verfahren zur parametrischen Kalibrierung von Parallel- und Hybridkinematiken entwickelt (vgl. Abb. 1). Die Grundlage bildet ein Modell der zu kalibrierenden Maschine, welches alle auf den Endeffektor einwirkenden Fehlereinflüsse berücksichtigen soll. Dazu wurde eine allgemeine Vorgehensweise zur kinematischen Modellbildung entwickelt, siehe [10]. Das Verfahren stützt sich auf homogene Transformationen und die Denavit-Hartenberg-Konvention, die entsprechend den Anforderungen für die Modellbildung hybrider Parallelkinematiken erweitert wurde. Die standardisierte Modellbildung wurde automatisiert und in den Kalibrierungsprozess integriert. Das entwickelte Verfahren wurde in die Kalibrierungssoftware integriert und ist allgemeingültig auf parallele oder serielle Strukturen anwendbar [8].

2.2 MATHEMATIK FÜR DIE AUSLEGUNG VON WERKZEUGMASCHINEN: ZERSPANPROZESSE

Die mathematische Modellierung und Simulation von Zerspanprozessen mit geometrisch bestimmter Schneide wie Hobeln, Bohren und Fräsen ist ein aktuelles Forschungsgebiet. Neben der Betrachtung der Spanentstehung stehen dabei auch die Beanspru-

¹ Gefördert im Rahmen des DFG-Schwerpunktprogramms 1099 „Fertigungsmaschinen mit Parallelkinematiken“ (Laufzeit: 2000–2006).

Abb. 1: Zwei- und dreidimensionale Variante der hybriden Parallelkinematik *Georg V.*


chung des Werkzeugs und Werkstück-Randzonenanalysen im Fokus der Untersuchungen. Die überwiegende Mehrheit der Untersuchungen, die auf numerische Berechnungen mithilfe der Finite-Element-Methode zurückgreifen, modellieren den Spanbildungsprozess mit den drei Elementen Werkzeug, Werkstück und Span. Besondere Aufmerksamkeit wird dabei den in der primären und den sekundären Scherzonen ablaufenden Verformungsprozessen gewidmet. Die Untersuchungen umfassen zum einen Analysen zur Spanabtrennung und zum Spanablauf, zum anderen die Simulation des Spanbruchs durch geeignete Bruchkriterien.

In den letzten Jahren trat verstärkt die Berücksichtigung des Werkstückwerkstoffes bei der Zerspansimulation in den Vordergrund. Dabei wurden für unterschiedliche Werkstoffe Materialgesetze erarbeitet, die neben der Abhängigkeit der Fließspannung von der Dehnung auch die Dehnrates und das Temperaturverhalten berücksichtigen.

Am Institut für Werkzeugmaschinen und Fabrikbetrieb der TU Berlin (IWF) wurde dazu das Fließverhalten von hochlegierten Stählen, zum Beispiel Nickelbasislegierungen, analysiert und modelliert, unter Berücksichtigung der während der Beanspruchung des Materials auftretenden Schädigungsmechanismen [23].

Abbildung 2 zeigt die Spanabhebung und eine Spanbildungssimulation. Bereits kleinste Änderungen an den Formelementen des Schneidkeils, wie die Modellierung einer ungefassten Schneidkante, führten zu einer merklichen Veränderung der spezifischen Schnittkraft. Die Schneidkeilgeometrie konnte so an den zu bearbeitenden Werkstoff angepasst werden, was zu einer deutlichen Verschleißreduzierung am Werkzeug führte [20, 22, 24].

Abb. 2: Simulation der Spanbildung beim Außenlängs-Runddrehen

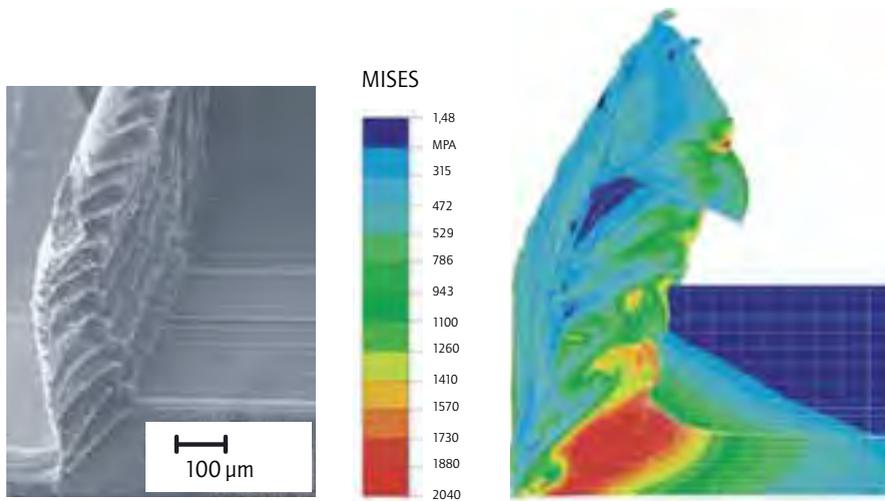
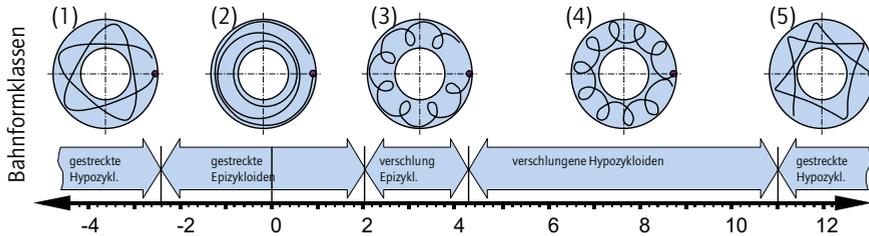


Abb. 3: Planschleifen mit Lappkinematik



Abb. 4: Bahnformklassen in Abhängigkeit vom Drehzahlverhältnis



Im Vergleich zur Zerspannsimulation mit geometrisch bestimmter Schneide ist die der unbestimmten Schneide aufgrund der hohen Komplexität der Wirkkinematik, dem Vorliegen von Abtrennvorgängen im mikroskopischen Bereich, der statistischen Natur der Schneidkörper sowie der ständigen Umspülung des Prozesses mit großen Mengen Kühlschmierstoff eine besondere Herausforderung bezüglich des Prozessverständnisses und der Prozessoptimierung.

Eine zusätzliche Schwierigkeit beim Planschleifen mit Läppkinematik ist, dass viele Werkstücke gleichzeitig zwischen zwei Schleifscheiben bewegt werden. Dabei sind die Bauteile in Werkstückhaltern fixiert, die wie Planeten in einem Planetengetriebe zwischen zwei Stiftkränzen rotieren (Abb. 3). Eine Simulation zur Vorhersage des Verschleißverhaltens der kostenintensiven Werkzeuge ist für den wirtschaftlichen Einsatz unverzichtbar.

Am IWF wurden dazu die Relativbewegungen zwischen den Wirkpartnern untersucht (siehe Abb. 4), damit konnte eine Berechnung der Aufenthaltswahrscheinlichkeit der Bauteile an unterschiedlichen radialen Abschnitten des Werkzeugs und damit auch die Berechnung der Schleifscheibenbelastung erfolgen.

Um einen inhomogenen Schleifscheibenverschleiß zu verhindern, wurde basierend auf Simulationsrechnungen die Schleifscheibenauslegung optimiert [2]. Des Weiteren konnte gezeigt werden, wie sich durch eine geeignete Wahl der kinematischen Prozessparameter das Zerspannungsvolumen ohne experimentellen Aufwand maximieren lässt [21].

3 MATHEMATISCHE KONZEPTE IN DER PRODUKTIONSTECHNIK

3.1 MODELLIERUNG VON PROZESSKETTEN

In den vergangenen Jahren sind große Fortschritte in der Simulation einzelner Prozesse und deren Effekt auf das Werkstück gemacht worden. Das kontinuumsmechanische

Materialverhalten lässt sich mittlerweile auch mit einer Reihe kommerzieller Softwarepakete simulieren, die im Wesentlichen auf der Finite-Element-Methode basieren. Auch für die Beschreibung der Dynamik von Werkzeugmaschinen sind kommerzielle Tools für die Aufstellung und Lösung der entsprechenden Bewegungsgleichungen verfügbar.

Es zeigt sich allerdings, dass die Aussagekraft der Simulationsresultate begrenzt ist, wenn es darum geht, auf der Basis der Resultate Verbesserungen an der Maschinenstruktur vorzunehmen. Genauso stößt diese Vorgehensweise an ihre Grenzen, wenn durch Kopplung kommerzieller Softwarepakete die Wechselwirkung verschiedener Effekte, beispielsweise zwischen Maschinenstruktur und Prozess beschrieben werden soll. Häufig sind die verschiedenen Pakete nur über graphische Nutzeroberflächen zu steuern, was eine zeitlich inkrementelle Kopplung verhindert. Selbst wenn die Pakete über Skriptsprachen und klar definierte Schnittstellen verfügen, die eine solche inkrementelle Kopplung ermöglichen, führt dieses Vorgehen häufig zu inakzeptabel kleinen Zeitschrittweiten für eine stabile numerische Berechnung.

Eine verlässliche Simulation der Wechselwirkungen zwischen Maschine, Prozess und Werkstück oder sogar einer ganzen Prozesskette erfordert eine durchgreifende Modellierung des gesamten Prozesses. Von besonderer Bedeutung ist dabei die Definition von Schnittstellen zwischen den verschiedenen wechselwirkenden Prozess(-ketten-)komponenten, die mathematisch als Rand- bzw. Anfangsbedingungen für die relevanten Feldgleichungen beschrieben werden können. Zur Illustration wird das in Abb. 5 für die gekoppelte Modellierung des Systems Fräser-Werkstück gezeigt. Der Einfachheit halber wird hier die Maschinendynamik nur als 2-Massen-Schwinger beschrieben (weitere Informationen, siehe [16]).

3.2 ADAPTIVE NUMERIK

Eine effiziente Gesamtsystemanalyse von Produktionsprozessen erfordert den Einsatz moderner Algorithmen der Numerischen Mathematik. Die numerische Behandlung von partiellen Differentialgleichungen, heute auch Scientific Computing genannt, steht in dem ständigen Bemühen nach optimalen und zuverlässigen Verfahren vor der Aufgabe der effektiven Netzgenerierung und der schnellen und ggfs. parallelen Lösung auftretender hoch-dimensionaler diskreter Probleme.

Aus Effizienzgründen sollte ein Gitter so grob wie möglich sein, andererseits sollte es so fein wie nötig sein, um die gewünschte Genauigkeit bei der Berechnung erzielen zu können. Um beiden Anforderungen gerecht zu werden, wurden in den letzten Jahren adaptive Algorithmen entwickelt. Diese generieren eine Folge von Netzen und dazugehörigen Finite-Elemente-Räumen, die nicht gleichmäßig verfeinern, sondern nur dort, wo ein Fehlerindikator anzeigt, dass der Fehler zu groß ist.

Wenngleich viele mathematische Fragen im Kontext adaptiver Algorithmen noch offen sind, so ist es doch der allgemein anerkannte Stand der Technik, dass adaptive Algorithmen gegenüber anderen uniformen Verfeinerungsmethoden vorzuziehen sind.

Abb. 5: Gekoppeltes Modell für das System Fräser-Werkstück

Mehrkörpermodell Maschine (MKS):

- Bewegungsgleichung Fräser

$$\ddot{q} + D\dot{q} + Kq = \frac{1}{m} F$$

- Kraft auf Fräser

$$F = - \sum_{j=1}^{N_z} g(\phi_j) \underbrace{O(\phi_j)}_{F^j} \hat{F}$$

- Schnittkraft \hat{F} in Abhängigkeit von Spanungsdicke h

$$\hat{F} = a_p \hat{K} h$$

Feldgleichungen Werkstück (PDEs):

- Impulsbilanz

$$\begin{aligned} \rho_0 u_{tt} &= \nabla \cdot \sigma, \\ \sigma &= \lambda(\varepsilon : I)I + 2\mu\varepsilon - 3K\alpha(T - T_0), \\ \varepsilon &= \frac{1}{2}(\nabla u + \nabla^T u), \end{aligned}$$

- Energiebilanz

$$\rho_0 c_v T_t = \kappa \nabla T - 3K\alpha T_0 \nabla \cdot u_t$$

Randbedingung auf Kontaktfläche $\Gamma(t)$

- Impulsbilanz

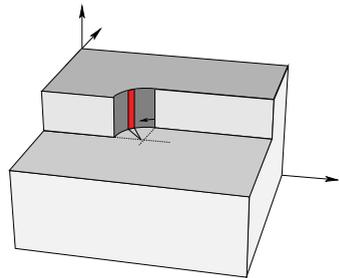
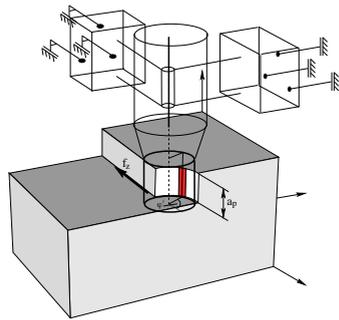
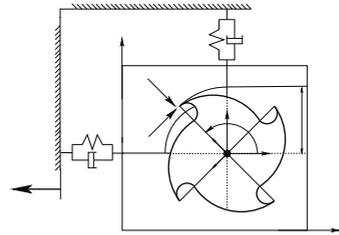
$$-\sigma \cdot n = \frac{F^j}{|\Gamma(t)|}$$

- Energiebilanz

$$-\kappa \partial_n T = W(F^j, v_T, h)$$

- Kopplung zwischen MKS und PDEs durch Spannungsdicke h

$$h = \left[\underbrace{R(t) - R(t - \tau)}_{\text{Vorschub}} + \underbrace{q(t) - q(t - \tau)}_{\text{Maschinendef.}} - \underbrace{u(t) - u(t - \tau)}_{\text{Werkstückdef.}} \right] e_r^j$$



Adaptive Algorithmen verringern die Zahl der Unbekannten dramatisch und führen auf ein deutlich kleineres, auf das Wesentliche reduziertes, diskretes Problem, das dann effektiv gelöst werden muss. Daneben aber wird eine Netzhierarchie bereitgestellt, die innerhalb der schnellen Lösung einen schnellen Informationsaustausch erlaubt und so Lösungsverfahren optimaler Komplexität überhaupt erst möglich macht [3, 15].

Abb. 6 zeigt zwei Zeitpunkte einer adaptiven Simulation der Laseroberflächenhärtung. Das Laserlicht besitzt typischerweise eine Eindringtiefe von weniger als 1 mm, daher muss die Randschicht stark verfeinert werden, um die metallurgischen Effekte räumlich auflösen zu können, die durch die Erwärmung hervorgerufen werden. Eine räumlich und zeitlich adaptive Simulation erlaubt es, das Gitter nur in der Nähe des Laserfokus fein aufzulösen. Das obere Bild zeigt jeweils den Temperaturverlauf und das untere das Gitter. Man sieht, wie die Verfeinerung mit dem Laserfokus mitwandert.

Auch die numerische Simulation der Wechselwirkung von Werkzeug und Werkstück bei Zerspanungsprozessen erfordert den Einsatz leistungsfähiger numerischer Näherungsmethoden. Typischerweise kommt es zu zeitlich sich verändernden elastischen und plastischen Bereichen und aufgrund von Reibung und Verformung zusätzlich zu einer Überlagerung thermischer Effekte. Abb. 7 zeigt die adaptive numerische Simulation des NC-Formschleifens von Freiformflächen [25]. Dabei wird das System Spindel-Schleifscheibe (oben links) durch ein FE-Modell (oben rechts) diskretisiert. Dazu kommt eine Kontaktbedingung zum Werkstück. Das untere Bild zeigt die adaptive Vernetzung des Kontaktbereichs sowie die Größe des Fehlers.

Abb. 6: Temperaturverteilung und adaptives Gitter bei der Laserhärtung zu zwei verschiedenen Zeiten

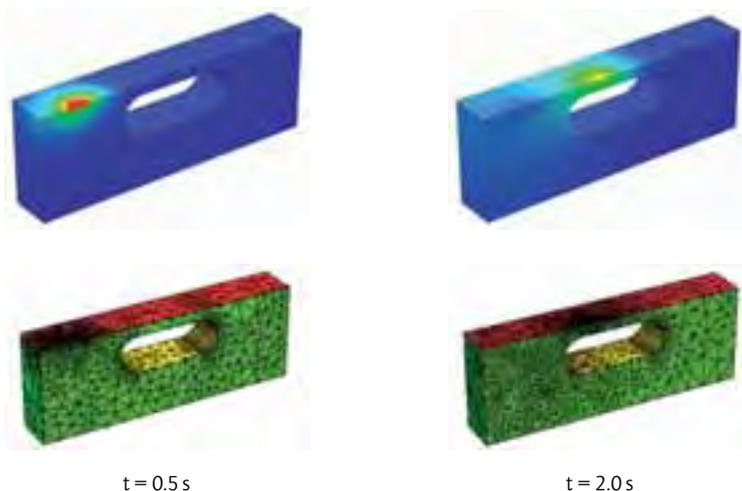
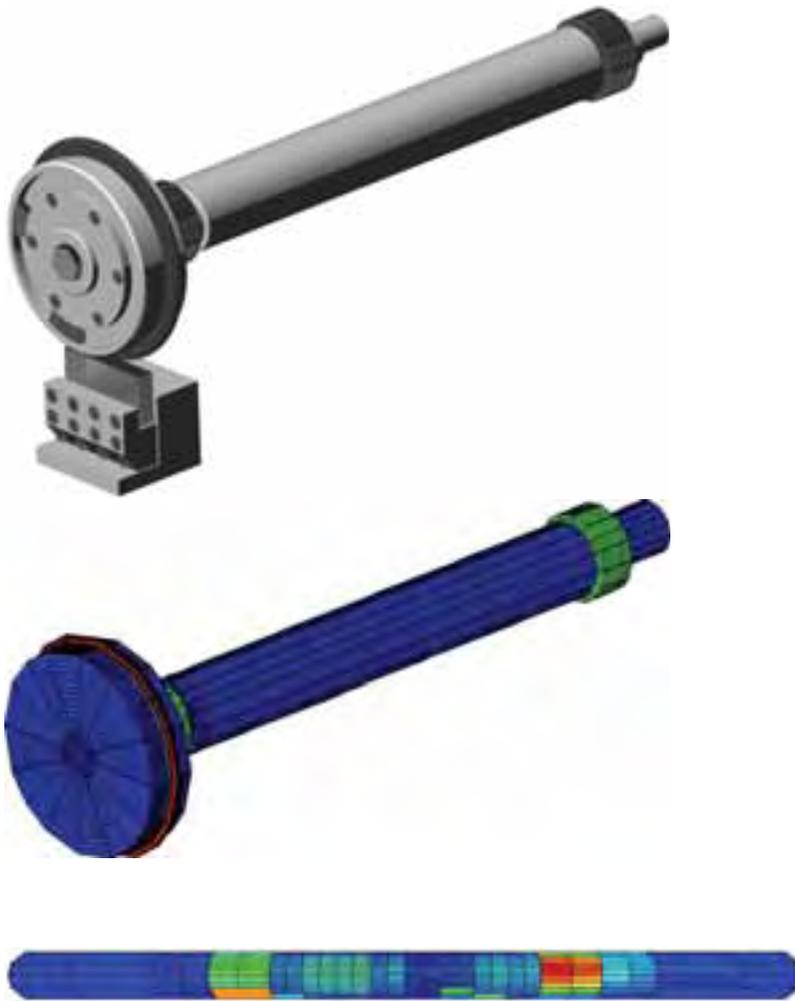


Abb. 7: Adaptive Simulation des NC-Formschleifens [25]. System Spindel-Schleifscheibe (oben), FE-Vernetzung (Mitte) und adaptive Vernetzung des Kontaktbereichs (rot: großer, blau: kleiner Fehler)



3.3 OPTIMALSTEUERUNG

Das ultimative Ziel jeder Prozesssimulation ist es, den Prozessablauf zu optimieren. Typische Ziele dabei sind die Verbesserung der Produktqualität, die Reduktion der Taktzeit oder die Senkung der Energiekosten.² Mathematisch bedeutet das die Lösung eines Optimalsteuerungsproblems, d. h. es wird eine Zielfunktion oder ein Gütefunktional definiert, mit deren Hilfe man das Prozessresultat bewerten kann. Dann versucht man, dieses Funktional zu minimieren unter der Nebenbedingung, dass der optimale Zustand eine Lösung der entsprechenden Feldgleichungen ist. Die mathematische Theorie der Optimalsteuerung partieller Differentialgleichungen ist ein aktives Forschungsgebiet in Deutschland. Durch eine Reihe von Neuberufungen in den vergangenen Jahren ist dieses für die Anwendungen äußerst wichtige Gebiet mittlerweile an fast allen technischen Universitäten vertreten. Ein guter Überblick über die mathematische Theorie findet sich in dem kürzlich erschienenen Lehrbuch [19].

Abb. 8 zeigt die Formulierung eines Optimalsteuerungsproblems für die Wärmebehandlung von Stahl. Ziel einer Wärmebehandlung ist es, eine bestimmte Gefügestruktur im Werkstück zu erreichen, die durch den Vektor der gewünschten Phasenanteile z_d beschrieben wird. Zur Endzeit t_E wird versucht, den Abstand der tatsächlich berechneten Gefügeanteile $z(x, t_E)$ von z_d zu minimieren, unter der Nebenbedingung, dass das Temperaturfeld T und der Vektor der Gefügeanteile z die entsprechenden Modellgleichungen erfüllen.

Abb. 8: Formulierung eines Optimalsteuerungsproblems für die Wärmebehandlung von Stahl

- Minimiere

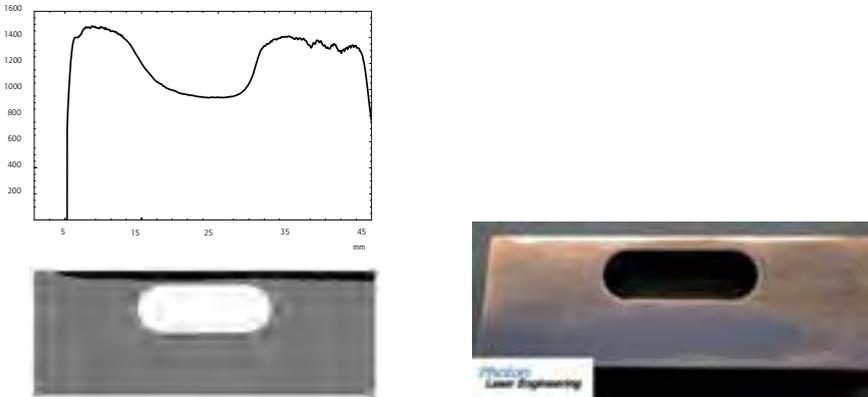
$$J(T, z; p) = \int |z(x, t_E) - z_d|^2 dV$$

- unter den Nebenbedingungen

- o Temperatur T erfüllt die Wärmeleitungsgleichung;
- o Phasenanteile z genügen einer Ratenbilanz für die Gefügeanteile in Stahl.

² Eine kürzlich vom BMBF in Auftrag gegebene Studie zeigt den Handlungs- und Forschungsbedarf im Bereich Energieeffizienz in der Produktion auf [14]. Hier öffnet sich ein weites Feld für Anwendungen der Optimalsteuerung.

Abb. 9: Optimale berechnete Temperatur im Laserfokus und resultierende simulierte Härtetiefe (links), experimentell mit numerisch berechneter optimaler Fokustemperatur als Sollgröße für die maschinenseitige Regelung erzielte Härtetiefe (rechts)



Die Wahl der Steuerungsvariablen p hängt ab vom jeweiligen Wärmebehandlungsverfahren. Bei der Induktionshärtung kann das die Generatorfrequenz oder der Einkopplungsabstand zwischen Spule und Werkstück sein, bei der Laserhärtung die Vorschubgeschwindigkeit und/oder die Leistung des Lasers. Häufig ist es so, dass das zugrunde liegende Modell auf vereinfachenden Annahmen beruht oder die Materialparameter nicht genau bekannt sind. Bei der Laserhärtung ist beispielsweise der Absorptionskoeffizient, der beschreibt, welcher Anteil der Laserleistung ins Werkstück einkoppelt, eine schwer bestimmbare Größe, da er nicht nur von der Rauigkeit der Oberfläche sondern auch von anderen schwer vorhersehbaren Faktoren wie Farbe und Verschmutzung abhängt.

Die besten und auch für die Praxis relevanten Resultate bringt die Optimalsteuerung daher in Kopplung mit einer maschinenseitigen Regelung. In [1, 12] wurde gezeigt, dass eine alleinige maschinenseitige Regelung mit konstanter Oberflächentemperatur nicht die gewünschte konstante Einhärtetiefe liefert. Andererseits lässt sich durch numerische Simulation bzw. durch die Lösung eines Optimalsteuerungsproblems die Laserleistung berechnen, die die gewünschte Einhärtetiefe liefert. Aufgrund der erwähnten Ungenauigkeit bei der Identifikation des Absorptionskoeffizienten würde die direkte Verwendung der berechneten optimalen Laserleistung zu einem schlechten Ergebnis führen. Allerdings lässt sich aus der optimalen Laserleistung auch eine resultierende optimale Fokustemperatur berechnen. Diese Temperatur muss unabhängig vom jeweiligen Absorptionskoeffizienten erreicht werden. Das beste Resultat erhält man also, wenn die berechnete optimale Fokustemperatur als Sollgröße für die maschinenseitige Regelung verwendet wird (vgl. Abb. 9).

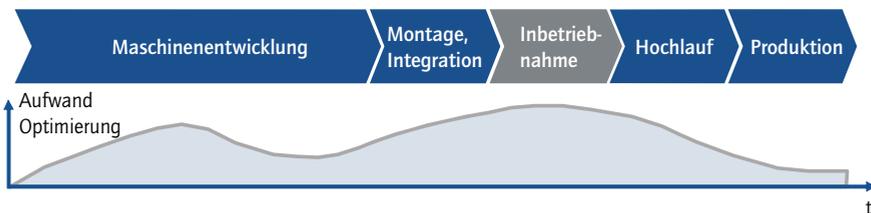
4 HERAUSFORDERUNGEN

4.1 WECHSELWIRKUNG ZWISCHEN STRUKTUR UND PROZESS

Werkzeugmaschinenhersteller entwickeln sich in den letzten Jahren mehr und mehr von Maschinenlieferanten zu Systemlieferanten, die dem Endanwender neben der eigentlichen Werkzeugmaschine auch Werkzeuge, Messmittel und Dienstleistungen anbieten. Dabei müssen die vom Auftraggeber festgelegten Kriterien wie Bauteilqualität, Bearbeitungszeit und Liefertermin erfüllt werden. Zu den Aufgaben des Werkzeugmaschinenherstellers gehört es also, nach Festlegung eines geeigneten Maschinenkonzepts schon in der Planungsphase für die von der Maschine durchzuführenden Bearbeitungsaufgaben geeignete Werkzeuge, Prozessparameter und Werkzeugbahnen zu bestimmen. Je nach Komplexität des zu fertigenden Bauteils werden die NC³-Programme zur Steuerung der Maschine manuell erstellt, oder es werden CAM⁴-Softwaresysteme zur Erzeugung der Werkzeugbahnen eingesetzt. CAM-Systeme bieten die Möglichkeit, bei gegebenem Werkstückgeometriemodell und nach Festlegung der benötigten Bearbeitungswerkzeuge die Werkzeugwege zur Herstellung des fertigen Werkstücks zu erzeugen und durch die Verwendung eines Postprozessors in einen auf der Maschine anwendbaren NC-Code zu übersetzen.

Nach Abschluss der Planungsphase erfolgen die Inbetriebnahme und anschließend die Abnahme der Maschine bei dem Endanwender. Der Aufwand in den einzelnen Phasen von der Maschinenentwicklung bis zu ihrem Einsatz ist schematisch in Abb. 10 dargestellt, wobei die Inbetriebnahmephase den größten Aufwand verursacht [6].

Abb. 10: Aufwand bei der Inbetriebnahme von Werkzeugmaschinen [6]



³ Numerical Control

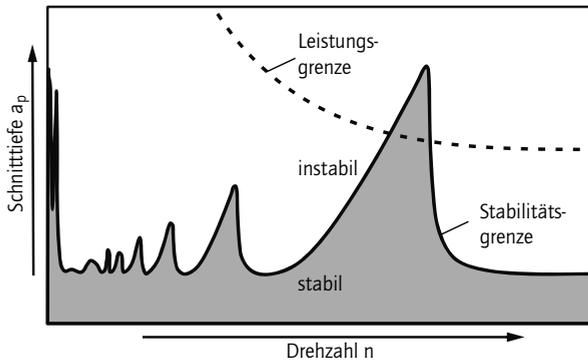
⁴ Computer Aided Manufacturing

Bei der Inbetriebnahme folgt auf die Montage der Fertigungseinrichtung das so genannte Einfahren der NC-Programme. In dieser Phase wird Rohmaterial in der Maschine aufgespannt und das durch die Planung erstellte NC-Programm ausgeführt, um Probewerkstücke herzustellen. Die erzeugten Werkstücke werden nach der Bearbeitung z. B. in einem 3D-Koordinatenmessgerät vermessen, und ein Protokoll über die vorhandenen Maßabweichungen an den zu prüfenden Funktionsflächen wird erstellt. Da bei Maschinenübergabe nachgewiesen werden muss, dass alle relevanten Maße des Abnahmebauteils innerhalb der Toleranzen liegen, sind die Ursachen, die zu den Maßabweichungen führen, zu ermitteln und abzustellen. Eine übliche Vorgehensweise zur Abstellung von Maßfehlern ist es, dass der für das Einfahren der Maschine verantwortliche Programmierer die Verfahrenswege der Werkzeuge solange anpasst, bis die Maße an den Funktionsflächen in der Mitte des Toleranzbandes liegen. Durch die Anpassung von NC-Programmen und die damit verbundenen zusätzlichen Testreihen kann es zu einer mehrtägigen Verzögerung in der Anlaufphase kommen [6]. Bei der Planung und Erstellung von NC-Programmen wird meist von idealen Bedingungen ausgegangen. Es wird angenommen, dass sowohl die Maschine mit ihren Komponenten als auch Werkzeug und Werkstück in ihren Maßen den Konstruktionsvorgaben entsprechen und dass während der Bearbeitung keine Verformungen auftreten. Chargenschwankungen führen jedoch zu variierenden Werkstoffeigenschaften und Rohteilmaßen, die neben nicht wiederholgenauen Einspannbedingungen oder aber durch Prozesskräfte hervorgerufene Werkzeugabdrängungen in der Realität zu Abweichungen an den gefertigten Werkstücken führen.

Eine weitere Problematik stellen schlecht auf die vorhandenen Systemeigenschaften abgestimmte Prozessparameter, wie Spindeldrehzahl und Schnitttiefe, dar, die zu instabilen Prozessen und damit zu Qualitätseinbußen und einer geringen Prozesssicherheit führen. Treten instabile Zustände auf, durch welche charakteristische Marken auf der Werkstückoberfläche entstehen können, sind die Prozessparameter so anzupassen, dass eine stabile Bearbeitung möglich ist. Hierfür wird für den betreffenden Bearbeitungsschritt z. B. die Schnitttiefe soweit verringert oder die Spindeldrehzahl angepasst, bis der Prozess stabil läuft. Eine Reduktion der Schnitttiefe ist jedoch direkt mit einer Verringerung des Zeitspannvolumens verbunden und führt zu einer nicht optimalen Ausnutzung der Maschinenleistung. Zudem dauert der Prozess länger als ursprünglich vorgesehen, wodurch die vom Auftraggeber geforderten Durchlaufzeiten nicht mehr zuverlässig eingehalten werden können.

Um schon in der Planungsphase Werkzeugbahnen und Prozessparameter festlegen zu können, bei denen in der Inbetriebnahmephase nur noch sehr geringe Anpassungen nötig sind, ist es zwingend erforderlich, die Wirkmechanismen, die zu Maßabweichungen oder Instabilitäten führen, zu kennen und zu berücksichtigen. Die in diesem Zusammenhang wichtigen Wirkmechanismen zwischen dem Fertigungsprozess und den Eigenschaften der Maschinenstruktur lassen sich unter dem Begriff Wechselwirkungen

Abb. 11: Stabilitätskarte eines Zerspanprozesses [7]



zwischen Struktur und Prozess zusammenfassen und werden am Beispiel der Fräsbearbeitung im Folgenden näher erläutert.⁵

4.2 WECHSELWIRKUNGEN BEIM FRÄSEN

Durch Zerspanprozesse angeregte Schwingungen führen neben einer Verkürzung der Werkzeugstandzeit und der Maschinenlebensdauer zu schlechten Oberflächengüten und einer hohen Geräuschentwicklung. Besonders kritisch sind in diesem Zusammenhang Schwingungen, die zu instabilen Zuständen mit plötzlich stark ansteigenden Amplituden führen („Rattern“). Das Auftreten von Maschinenschwingungen hängt von mehreren Parametern ab. Zum Beispiel erhöhen lang auskragende Werkzeuge, große Schnitttiefen, oder dünnwandige Bauteile die Wahrscheinlichkeit, dass Schwingungen auftreten. Hauptziel der Untersuchungen zu den Wechselwirkungen bei Zerspanprozessen ist es, diese Schwingungen vorhersagen zu können, um Fräsprozesse so auszulegen, dass eine möglichst unkritische Schwingungsanregung stattfindet.

Die Stabilitätseigenschaften einer Werkzeugmaschine werden häufig mit Hilfe von Stabilitätskarten bestimmt. Auf diesen auch „Ratterkarten“ genannten Diagrammen

⁵ Um die Zusammenhänge zwischen dem Fertigungsprozess, dem Maschinenverhalten und den resultierenden Werkstückeigenschaften besser verstehen und prognostizieren zu können, ist im Jahr 2005 durch die Deutsche Forschungsgemeinschaft das Schwerpunktprogramm 1180 „Prognose und Beeinflussung der Wechselwirkungen zwischen Struktur und Prozess“ eingerichtet worden (www.prowesp.de). In diesem auf insgesamt sechs Jahre ausgelegten Forschungsprogramm arbeiten deutschlandweit ca. 50 Forscher aus den Bereichen Fertigungstechnik, Mechanik, Mathematik und Werkstoffwissenschaften interdisziplinär zusammen. Über Fächergrenzen hinweg werden innovative Ansätze und Methoden für die Vorhersage und gezielte Beeinflussung der Wechselwirkungen erforscht, vgl. [7].

werden in Abhängigkeit der aktuellen Spindeldrehzahl die Schnitttiefen dargestellt, bis zu der eine stabile Bearbeitung möglich ist. Auf der in Abb. 11 dargestellten Stabilitätskarte würden Betriebspunkte oberhalb der Stabilitätsgrenze zu instabilen Prozessen führen, und umgekehrt. Neben der Stabilitätsgrenze muss bei der Auslegung von Fräsprozessen auch die Leistungsgrenze der Maschine berücksichtigt werden.

Um eine qualitativ hochwertige und aussagekräftige Stabilitätskarte zu erhalten, ist heutzutage ein sehr hoher experimenteller Aufwand erforderlich. Dabei werden Zerspanversuche mit systematisch variierten Schnitttiefen und Drehzahlen durchgeführt. Führt die Parameterkombination zu einem stabilen oder instabilen Prozess, wird dies in der Stabilitätskarte markiert. Die Güte des verwendeten Stabilitätskriteriums beeinflusst dabei auch die Güte der Stabilitätskarte. Zur Einschränkung der zu untersuchenden Parameterbereiche werden vor den Stabilitätsuntersuchungen meist weitergehende Analysen durchgeführt. Die experimentelle Modalanalyse bietet z. B. die Möglichkeit, die Eigenfrequenzen einer Maschine zu bestimmen. Da die Eigenfrequenzen zumeist auch den Ratterfrequenzen entsprechen, müssen dann nur Spindeldrehzahlen untersucht werden, die in der Nähe einer Maschineneigenfrequenz liegen.

Um den experimentellen Umfang bei der Erzeugung von Stabilitätskarten zu verringern, werden vermehrt auch Simulationen eingesetzt. Dabei zeigt sich, dass durch Simulationen prinzipiell der Verlauf der real gemessenen Stabilitätsgrenzen nachgebildet werden kann. Da das für die Simulation zugrunde liegende Modell mit Unsicherheiten behaftet ist, kann es jedoch zu Abweichungen zwischen experimentell ermittelter und simulierter Stabilitätskarte führen, die dessen Aussagekräftigkeit und Einsatz stark reduzieren. Um die Simulationsgüte dieser Instabilitätsuntersuchungen zu erhöhen, werden im Schwerpunktprogramm 1180 unterschiedliche Ansätze zur simulativen Erzeugung von Stabilitätskarten erforscht. Ein Zugang beruht auf einem gekoppelten mathematischen Modell (vgl. Abbildung 5), welches die Dynamik der Werkzeugmaschine als Mehrkörpersystem abbildet und das Werkstück als thermoelastisches Kontinuum beschreibt. Abb. 12 zeigt die reduzierte Darstellung der Werkzeugmaschine und das Werkstück.

Auch wenn hinsichtlich der Parameteridentifikation für das Maschinenmodell und die Modellierung der Fügestellen im Mehrkörpermodell noch Fragen offen sind, konnte doch schon im bisherigen Projektverlauf ein vertieftes Verständnis der Wechselwirkung von Maschinen- und Bauteildynamik gewonnen werden [16].

4.3 REKONFIGURATION VON PRODUKTIONSANLAGEN

Abb. 13 zeigt die rapide anwachsende Modellvielfalt in der Automobilindustrie. Mit jeder Modellgeneration verstärkt sich der Druck zu kürzeren Entwicklungszeiten (time to market). Dies hat nicht nur große Auswirkungen auf die Bauteileentwicklung, auch die Zeit für Entwicklung und Realisierung von Produktionssystemen wird permanent gekürzt. Die immer größer werdende Varianz der Modelle, die kürzeren Lebenszyklen und die Forderung nach mehr Flexibilität bilden enorme Zielkonflikte.

Abb. 12: Fräsmaschine als Mehrkörpersystem in Wechselwirkung mit einem thermoelastischen Werkstück

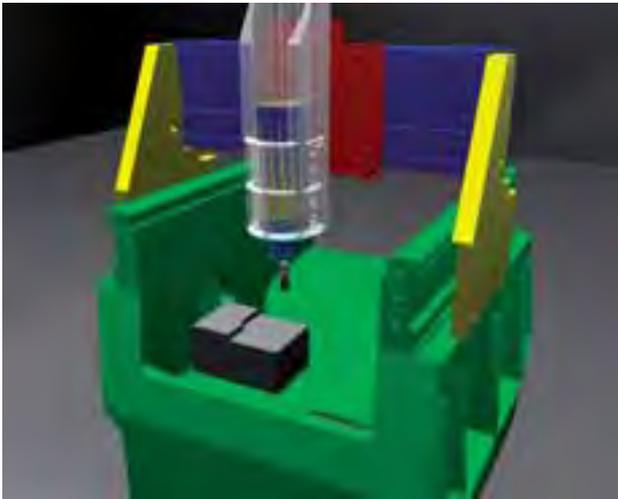
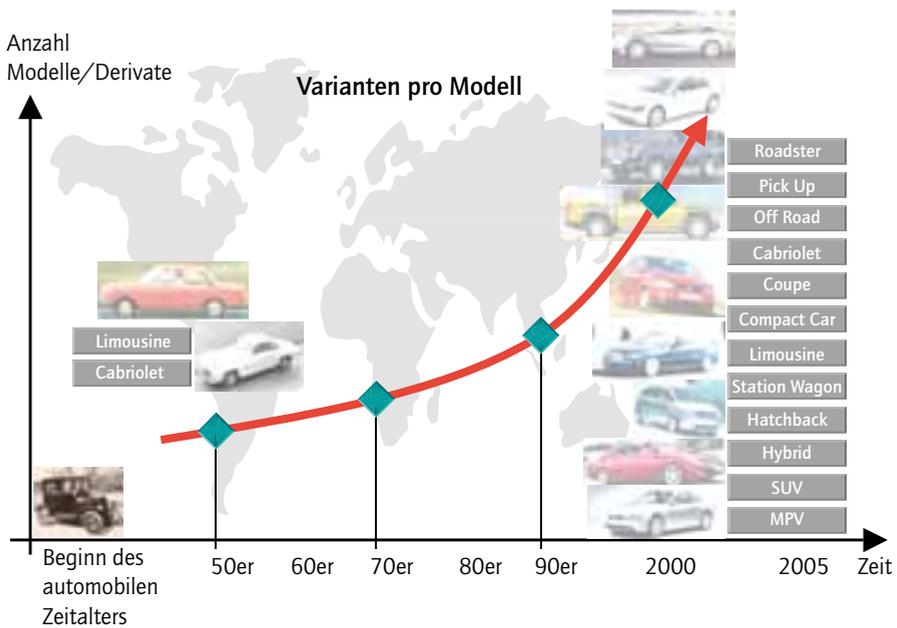


Abb. 13: Variantenentwicklung in der Automobilproduktion (mit freundlicher Genehmigung von Rücker EKS, Weingarten)



Angesichts dieser Situation ist die virtuelle Simulation komplexer Produktionsanlagen von wachsender Bedeutung. Abb. 14 zeigt ein typisches Produktionsszenario. Man erkennt Gruppen von Robotern, die an verschiedenen geostationären Vorrichtungen arbeiten, dazu kommen Förderbänder für Materialtransport und weitere Vorrichtungen für die Zu- und Abfuhr sowie Zwischenlagerung von Bauteilen. Für die virtuelle Produktion sind in den vergangenen Jahren einige kommerzielle Tools entwickelt worden, die in der Lage sind, Produktionsprozesse von hoher Komplexität abzubilden. Damit lassen sich ganze Produktionsanlagen am Computer aufbauen und simulieren. Allerdings sind sie noch weit davon entfernt, eine zumindest halbautomatische Planung neuer bzw. Rekonfiguration bestehender Anlagen zu ermöglichen.

Für die Rekonfiguration einer solchen Anlage lassen sich mindestens drei Ebenen unterscheiden: Im einfachsten Fall muss lediglich das Programm eines Roboters geändert werden, weil er z. B. einen weiteren Schweißpunkt anbringen muss. Bei einer Rekonfiguration mittlerer Komplexität könnte sich z. B. die Geometrie eines Bauteils ändern, so dass die Zuordnung der Arbeitspunkte und die Bahnplanung aller Roboter einer Gruppe modifiziert werden muss. Eine komplexe Rekonfiguration würde einen zusätzlichen Hardwareumbau erfordern.

Hier bietet sich ein weites Feld für Kooperationen zwischen Produktionstechnik und Mathematik. Die Änderung eines Roboterprogramms aufgrund mathematischer Bahnplanung gehört seit den neunziger Jahren zum Stand der Technik [11, 17]. Aber man muss wohl konstatieren, dass die Entwicklung stehen geblieben ist. Schon bei der Bahnplanung für kooperierende Roboter gibt es noch viele offene Fragen [4, 13].

Eine spannende Herausforderung im Schnitt von diskreter und gemischt-ganzzahliger Optimierung, Optimalsteuerung und Produktionstechnik wäre beispielsweise die Entwicklung eines Planungstools für eine Roboterschweißzelle. Basierend auf den CAD-Daten eines Bauteils und den vorgegebenen Koordinaten der aufzubringenden Schweißpunkte, müsste dieses Tool eine automatische Zuordnung der Schweißpunkte zu den Robotern vornehmen, eine optimale Reihenfolge festlegen und schließlich automatisch optimale und kollisionsfreie Roboterbahnen generieren.

5 PERSPEKTIVEN

Aus industrieller Sicht besteht großer Bedarf an Planungs- und Optimierungstools, die eine beschleunigte Neuplanung und Rekonfiguration komplexer Produktionsanlagen erlauben.

Die Modellierung und Simulation von grundlegenden Produktionsprozessen wie Fräsen, Schleifen und Umformen und von Fügeprozessen wie Löten, Schweißen und Kleben, ist seit den achtziger Jahren ein aktives Forschungsgebiet der Produktionstechnik und der Angewandten Mathematik. Das Gleiche gilt für die Mehrkörpersimulation von Werkzeugmaschinen und Robotern. Demgegenüber ist die Untersuchung von Wechselwirkungen zwischen Prozess, Maschine und Werkstück erst in den letzten Jah-

Abb. 14: Virtuelle Simulation einer komplexen Produktionsanlage (mit freundlicher Genehmigung von Rücker EKS, Weingarten)



ren in Angriff genommen worden, getrieben vor allem durch die Einrichtung eines DFG-Schwerpunktprogramms zu diesem Thema.

Außerdem hat es in den vergangenen Jahren große Fortschritte bei der Beschreibung des globalen Produktions- und Materialflusses gegeben. Die Untersuchung von Supply Chains sowie die Discrete-Event Simulation haben sich als aktive Forschungsgebiete der Angewandten Mathematik und Informatik [5, 9] etabliert.

Der nächste Schritt auf dem Weg zu automatischen Planungstools für die Neuplanung und Rekonfiguration komplexer Produktionsprozesse ist die Entwicklung eines mehrskaligen mathematischen Standardmodells. Dazu bedarf es einer gemeinsamen Anstrengung von Produktionstechnik, Mathematik und Informatik:

- Auf dem niedrigsten Level geht es dabei um die Modellierung elementarer Produktionsprozesse wie Schweißen, Bohren, Fräsen unter Einbeziehung der ausfüh-

- renden Werkzeugmaschine bzw. des entsprechenden Manipulators. Mathematisch geht es hier um partielle Differentialgleichungs- und Mehrkörper-Modelle.
- Auf der mittleren Ebene werden diese Prozesse zusammengefasst zu Arbeitszellen, bestehend aus einer Gruppe von Robotern, Werkzeugmaschinen, Halteapparaturen und Förderbändern. Mathematische Aufgaben sind hier vor allem gemischt-ganzzahlige Optimierung sowie Bahnplanung.
 - Schließlich werden diese Arbeitszellen verbunden zu komplexen Produktionssystemen. Hier geht es vor allem um die Beschreibung des Workflow und die Entwicklung einer entsprechenden Sprache, d. h. die Aufgaben liegen im Bereich der Angewandten Informatik.

Die Arbeiten könnten z. B. im Rahmen eines interdisziplinären DFG-Schwerpunktprogramms in Angriff genommen werden.

Der Vorteil in der Entwicklung eines solchen Konzeptes liegt darin, dass er problemlos die Berücksichtigung verbesserter Simulations- und Optimierungstechniken auf den verschiedenen Skalen erlaubt. In Verbindung mit geeigneter Sensorik ließe sich das Fabrikmodell für die Regelung und Überwachung des gesamten Produktionsprozesses verwenden. Es würde eine intelligente Fehlerkontrolle liefern und könnte für das kontrollierte Hoch- und Herunterfahren der Anlage verwendet werden. Durch Kopplung mit entsprechender Lagerhaltungssoftware ließe sich schließlich eine Business-to-Manufacturing Integration des Produktionsablaufes erreichen. Mit anderen Worten: Das vielbeschworene digitale Unternehmen kann Wirklichkeit werden.

6 LITERATUR

- [1] Alder, H.; Hömberg, D.; Weiss, W.: *Simulationsbasierte Regelung der Laserhärtung von Stahl*, HTM Z. Werkst. Wärmebeh. Fertigung, **61** (2006), pp. 103–108.
- [2] Ardelt, T.: *Einfluss der Relativbewegung auf den Prozess und das Arbeitsergebnis beim Planschleifen mit Planetenkinematik*, Dissertation TU Berlin, 2001.
- [3] Brenner, S.; Carstensen, C.: *Finite element methods*, Kapitel 3 in: *Encyclopedia of Computational Mechanics* (Eds. E. Stein, R. de Borst, T.J.R. Hughes), 2004.
- [4] Caccavale, F.; Villani, L.: *Impedance control of cooperative manipulators*, *Machine Intelligence and Robotic Control*, **2** (2000), pp. 51–57.
- [5] Degond, P.; Göttlich, S.; Herty, M.; Klar, A.: *A network model for supply chains with multiple policies*, *SIAM Multiscale Modeling & Simulation (MMS)*, **6** (2007), No. 3, pp. 820–837.

- [6] Denkena, B.; Brecher, C.: *Ramp-Up/2: Anlaufoptimierung durch Einsatz virtueller Fertigungssysteme*, VDMA Verlag, Frankfurt, 2007.
- [7] Denkena, B.; Deichmüller, M.: *Wechselwirkungen zwischen Prozess und Maschine*, in: *Spanende Fertigung: Prozesse Innovationen Werkstoffe* (Hrg. D. Biermann), Vulkan Verlag, Essen, 2008.
- [8] Denkena, B.; Günther, G.; Mehrmann, V.; Möhring, H.-C.; Steinbrecher, A.: *Kalibrierverfahren für hybride Parallelkinematiken*, in: *Fertigungsmaschinen mit Parallelkinematiken – Forschung in Deutschland* (Hrsg. U. Heisel, H. Weule), erscheint im Shaker Verlag.
- [9] Fowler, J.W.; Rose, O.: *Grand challenges in modeling and simulation of complex manufacturing systems*, SIMULATION, **80** (2004), No. 9, pp. 469–476.
- [10] Günther, G.; Steinbrecher, A.: *Strukturausnutzung hybrider Parallelkinematiken hinsichtlich ihrer Kalibrierung*, in: *Tagungsband, 3. Dresdner WZM-Fachseminar*, Technische Universität Dresden, 2001.
- [11] Heim, A.; von Stryk, O.: *Trajectory optimization of industrial robots with application to computer-aided robotics and robot controllers*, Optimization, **47** (2000), pp. 407–420.
- [12] Hömberg, D.; Weiss, W.: *PID control of laser surface hardening of steel*, IEEE Trans. Control Syst. Technol., **14** (2006), pp. 896–904.
- [13] Lippiello, V.; Siciliano, B.; Villani, L.: *An open architecture for sensory feedback control of a dual-arm industrial robotic cell*, Industrial Robot, **34** (2007), pp. 46–53.
- [14] Neugebauer, R.: *Energieeffizienz in der Produktion*, Studie der Fraunhofer Gesellschaft im Auftrag des BMBF, 2007.
- [15] Rannacher, R.; Suttmeier, F.-T.: *Error estimation and adaptive mesh design for FE models in elasto-plasticity theory*, in: *Error-Controlled Adaptive Finite Elements in Solid Mechanics* (Ed. E. Stein), J. Wiley & Sons, 2002.
- [16] Rott, O.; Rasper, P.; Hömberg, D.; Uhlmann, E.: *A milling model with thermal effects including the dynamics of machine and work piece*, WIAS Preprint No. 1338 (2008).
- [17] Steinbach, M.C.; Bock, H.G.; Kostin, G.V.; Longman, R.W.: *Mathematical optimiza-*

- tion in robotics: towards automated highspeed motion planning*, *Surveys Math. Indust.*, **7** (1998), No. 4, pp. 303–340.
- [18] Tönshoff, H.K.; Günther, G.; Grendel, H.: *Influence of manufacturing and assembly errors on the pose accuracy of hybrid kinematics*, in: *Proceedings of the 2000 Parallel Kinematic Machines – International Conference*, September 13th-15th 2000, Ann Arbor, Michigan, USA, pp. 255–262.
- [19] Tröltzsch, F.: *Optimale Steuerung partieller Differentialgleichungen*, Vieweg, Wiesbaden, 2005.
- [20] Uhlmann, E.; Mattes, A.; Zettier, R.; Graf von der Schulenburg, M.: *Investigations on the adjustment of the modeling action in 2D simulation of milling processes*, in: *Proceedings of the 10th CIRP Int. Workshop on Modeling in Machining Operations*, 27.–28. August 2007, Kalabrien, Italien, 2007, pp. 157–164.
- [21] Uhlmann, E.; Paesler, C.; Ardelt, T.: *Qualitäts- und Leistungssteigerung beim Planschleifen mit Planetenkinematik*, Vortrag, 10. Internationales Braunschweiger Feinbearbeitungskolloquium, Braunschweig, 7.–9. Oktober 2002.
- [22] Uhlmann, E.; Zettier, R.: *3D FE simulation of turning processes*, CIRP Workshop on “FE Simulation of Cutting and Forging Processes”, 29. 1. 2003, Paris, France.
- [23] Uhlmann, E.; Zettier, R.: *Experimentelle und numerische Untersuchungen zur Spannbildung beim Hochgeschwindigkeitsspanen einer Nickelbasislegierung*, in: *Hochgeschwindigkeitsspanen metallischer Werkstoffe* (Hrsg. H.K. Tönshoff, F. Hollmann), Wiley-VCH, Weinheim 2005, pp. 404–425.
- [24] Uhlmann, E.; Zettier, R.; Sievert, R.; Clos, R.: *FE simulation of high-speed turning of Inconel 718*, (Eds. A. Moisan, G. Poulachon), *Proceedings of the 7th CIRP International Workshop on Modeling of Machining Operations*, 4.–5. 5. 2004, Cluny, France, pp. 67–74.
- [25] Weinert, K.; Blum, H.; Jansen, T.; Rademacher, A.: *Simulation based optimization of the NC-shape grinding process with toroid grinding wheels*, *Production Engineering – Research and Development*, **1** (2007), No. 3, Springer, Berlin, Heidelberg, pp. 245–252.

> HERSTELLUNG UND EINSATZ MODERNER MATERIALIEN

WOLFGANG DREYER

1 EXECUTIVE SUMMARY

Der Einsatz neuer Materialien in Anwendungen der Hochtechnologie steigt seit längerer Zeit rasant an. Stähle mit neuartigen Eigenschaften werden entwickelt. Die enorm steigende Nachfrage nach Sonnenkollektoren bedingt neue Verfahren zur Herstellung der hierzu benötigten Halbleiter. Zur Erzielung höherer Taktraten im Computer, aber auch zur Steigerung der Leistung von Sonnenkollektoren wird die reine Siliziumtechnologie verlassen; es werden neuartige Halbleiter konzipiert. Im Zuge der Miniaturisierung von Chips kommt dem Halbisolator große Bedeutung zu.

Hier die Marktführerschaft zu gewinnen, setzt ein tiefes und insbesondere interdisziplinäres Verständnis der physikalischen Vorgänge in der Technologiekette *Charakterisierung, Herstellung und Lebensdauer neuer Materialien* voraus. Dies zu erlangen, gelingt aber in hohem Maße nur bei Einsatz der angewandten Mathematik mit ihren Methoden *Modellierung, Analysis und Simulation*. Die geschickte Anwendung mathematischer Methoden kann nämlich

- die Entwicklungszeit neuer Materialien drastisch verkürzen,
- sogar zu Patenten führen, und schließlich
- mittels subtiler Methoden den Ingenieur vor numerischen Fehlern und anderen Artefakten bewahren.

Illustriert wird das hier anhand ausgewählter Aspekte der Themen: Herstellung eines Kristalls, Verbesserung der Eigenschaften des Halbisolators Galliumarsenid, Bruchtest eines Wafers, Alterung und Schädigung von Lotverbindungen auf Chips. Der wesentliche mathematische Kontext sind Anfangs- und Randwertaufgaben für gekoppelte Systeme nichtlinearer partieller Differentialgleichungen.

2 ERFOLGSGESCHICHTEN

Kristallzucht im wandernden Magnetfeld. Die essentiellen Bauteile elektronischer Geräte basieren auf gezüchteten Kristallen. Deren Herstellungsprozess beginnt bei hoher Temperatur mit einer Schmelze, woraus durch gezielte Abkühlung der Kristall entsteht. Bei dem sogenannten Czochralski-Verfahren wird der Kristall aus einer rotierenden Schmelze gezogen.

Abb. 1 zeigt einen Ausschnitt einer typischen Czochralski-Anlage inklusive einer berechneten Temperaturverteilung [10]. Zur Erzielung eines homogenen Kristalls sind drei Dinge in der Schmelze sicherzustellen, nämlich eine räumlich und zeitlich möglichst homogene Temperatur, eine laminare Strömung im Schmelztiigel, und eine konvexe Phasengrenze zwischen Schmelze und Kristall. Dies zu erreichen erfordert langjährige Erfahrung und große Experimentierkunst. Sind nun zu diesem Zweck alle Parameter zur Züchtung eines Kristalls aus einer Schmelze mit gegebenem Durchmesser und gegebener Schmelzhöhe optimal eingestellt, und entscheidet sich dann der Kristallzüchter, größere Kristalle herzustellen, was eine Vergrößerung von Tiegeldurchmesser und Schmelzhöhe bedingt, so wird dies im allgemeinen zu einem Kristall von minderer Qualität führen. Wieder sind langwierige Änderungen der Kristallzuchtanlage und neue Versuche zur Auffindung der neuen optimalen Parameter notwendig. Parallel zur industriellen Produktion von Kristallen kann dies nicht geleistet werden kann.

In diesem Zusammenhang wurde in Berlin ein Konsortium, bestehend aus zwei Industriefirmen, Physikern, Ingenieuren und Mathematikern, zur Durchführung des interdisziplinären Projektes *KristMAG* gegründet.¹ Es sollte herausgefunden werden, ob mit einem sogenannten wandernden Magnetfeld, welches durch die selben Spulen induziert wird, die auch die (Induktions-) Heizung der Anlage bewirken, Temperatur, Strömung und Phasengrenze im obigen Sinn günstig zu beeinflussen sind. Das seit kurzem abgeschlossene Projekt hat dies eindrucksvoll aufgezeigt, es sind mehrere Modellanlagen gebaut worden, und es gibt fünf erteilte Patente sowie einen Patentantrag, welcher insbesondere durch die Mathematiker im Projekt initialisiert wurde.

Das kombinierte Heizungs-Magnet-Modul ist schematisch in der Czochralski-Anlage der Abb. 1 eingezeichnet. Es besteht aus drei zylindrischen Spulen, die von phasenverschobenen Strömen durchflossen werden und das wandernde Magnetfeld erzeugen. Diese Spulen sind seitlich zum Schmelztiigel angebracht. Die Abb. 2 zeigt auf der linken Seite die Momentaufnahme einer berechneten typischen Temperatur- und Geschwindigkeitsverteilung in der Schmelze ohne Magnetfeld. Auf der rechten Seite gibt es ein nach unten wanderndes Magnetfeld, welches augenscheinlich die Temperaturverteilung ändert und außerdem die Richtung und Form des rechten Wirbels umkehrt.

¹ Gefördert durch den Zukunftsfonds der Stadt Berlin und durch die Technologiestiftung Berlin und kofinanziert durch die EU (EFRE).

Abb. 1: Ausschnitt einer Kristallzuchtanlage mit Magnetspulen und einigen berechneten Temperaturen (aus [10])

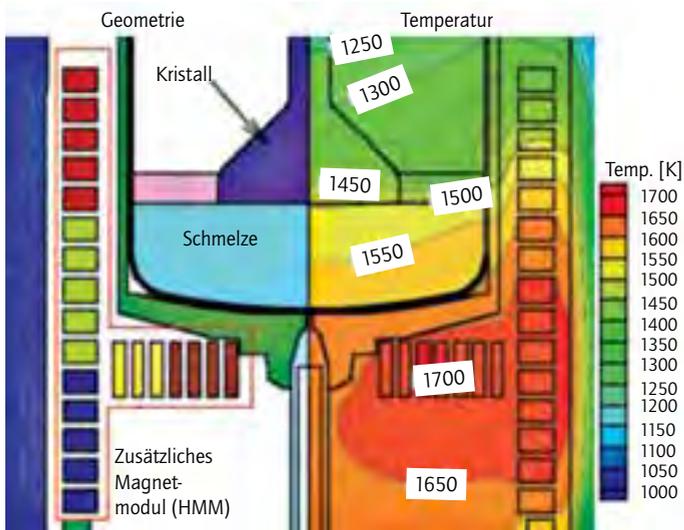


Abb. 2: Schnappschuss der Temperatur- und Strömungsverteilung in der Schmelze, aus [10]. Links: Ohne Magnetfeld, Rechts: Mit Magnetfeld. Man beachte die unterschiedlichen Skalen von Geschwindigkeit und Temperatur.

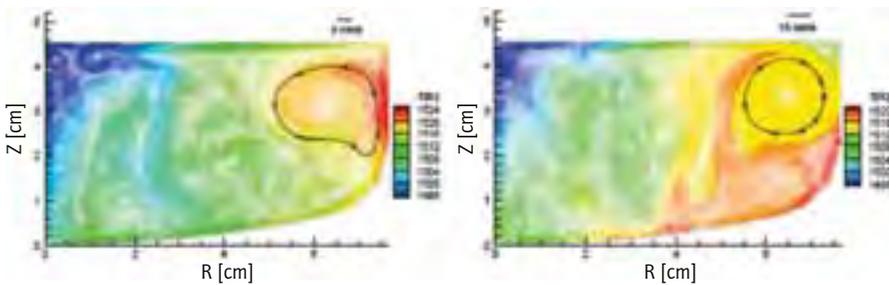


Abb. 3: Mathematischer Kontext zur Kristallzucht im wandernden Magnetfeld

Das Problem der Beeinflussung von Temperatur, Strömung und Phasengrenze führt mathematisch gesehen auf ein gekoppeltes System von nichtlinearen partiellen Differentialgleichungen für ein nichtglattes Gebiet mit freien Rändern, siehe [6]. Beispielsweise sind die Variablen in der Flüssigkeit die Temperatur T , die Strömungsgeschwindigkeit $v = (v_i)_{i=1,2,3}$ und die magnetische Induktion $B = (B_i)_{i=1,2,3}$. Die Differentialgleichungen enthalten die Navier-Stokes-Gleichungen in der Boussinesq-Aproximation,

$$\begin{aligned} \rho \left(\frac{\partial v}{\partial t} + v \cdot \nabla v \right) + \nabla p \\ + \operatorname{div} \left(\eta(T)(\nabla v + (\nabla v)^T) \right) \\ = \rho(1 + \alpha(T - T_R))g + \frac{1}{\mu}(\operatorname{rot} B \times B), \end{aligned}$$

$$\operatorname{div} v = 0.$$

Auf der rechten Seite steht hier die Auftriebskraft, welche in hohem Maße die unerwünschten Wir-

bel induziert, sowie die Lorentzkraft, mittels der ein vorhandenes Magnetfeld die Strömung beeinflussen kann. Das Magnetfeld wird aus reduzierten Maxwell-Gleichungen berechnet, nämlich

$$\frac{\partial B}{\partial t} + \operatorname{rot} \left(\frac{1}{\mu \sigma(T)} \operatorname{rot} B \right) = 0$$

und schließlich gibt es die Wärmeleitungsgleichung mit Konvektion,

$$\begin{aligned} \rho c_v \left(\frac{\partial T}{\partial t} + v \cdot \nabla T \right) \\ = \operatorname{div} (\kappa(T) \nabla T) + \eta(T) \nabla v \cdot \nabla v \\ + \frac{\operatorname{rot} B \cdot \operatorname{rot} B}{\mu^2 \sigma(T)}. \end{aligned}$$

In anderen Teilen der Anlage tritt zusätzlich noch Strahlung auf. Hierauf, sowie auf die Randbedingungen, wird hier aber nicht eingegangen.

Darüber hinaus gibt es zwei zusätzliche Spulen unterhalb des Schmelzriegels, die ein nach außen laufendes Magnetfeld erzeugen, und hierauf bezieht sich der neue Patentantrag [7]. Nahegelegt wurde diese Anordnung durch verschiedene Simulationen, die gezeigt haben, dass hiermit bei geeignet eingestellten Parametern die zeitlichen Temperaturosillationen gedämpft werden und darüber hinaus die Form der Phasengrenze zwischen Schmelze und Kristall variiert werden kann. Diese wichtige Erkenntnis hätte man natürlich auch experimentell heraus bekommen. Nun ist es aber so, dass bereits kleinste Umbauten einer Kristallzuchtanlage, beispielsweise an dem System der Magnetspulen, mehrere Wochen, teilweise bis zu Monaten in Anspruch nehmen. In dieser Zeit kann weder experimentiert noch kann ein Kristall gezüchtet werden. Die virtuelle Kristallzuchtanlage, wie sie durch das mathematische Modell repräsentiert wird, gibt deshalb dem Kristallzüchter einen großen Wettbewerbsvorteil hinsichtlich Zeit als auch materiellem Aufwand.

Unerwünschte Präzipitatbildung in Galliumarsenid-Kristallen. Aus dem gezüchteten Kristall werden so genannte Wafer hergestellt, dünne einkristalline Scheiben, welche dann zur Produktion von Chips, Lasern oder Photovoltaikanlagen weiterverarbeitet werden. Zur Homogenisierung der elektrischen, optischen und mechanischen Eigenschaften

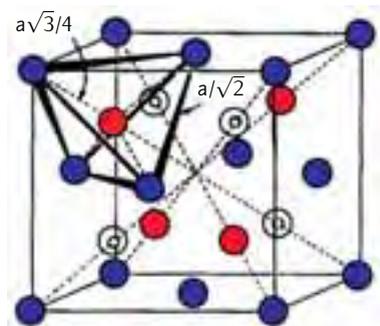
muss der Wafer einer Wärmebehandlung oberhalb einer gewissen Temperatur unterzogen werden. Hierbei kommt es zu unerwünschten Ausscheidungen. In halbleitenden Kristallen aus Galliumarsenid (GaAs) sind dies beispielsweise flüssige Tropfen mit einem Arsenanteil von über 90 %. In diesem Zustand kann ein Einsatz in mikro- und optoelektronischen Anwendungen nicht erfolgen, so dass ein Verfahren zur Auflösung der Tropfen gefunden werden muss, und auch hierzu haben die Methoden der angewandten Mathematik einen wichtigen Beitrag geliefert.

Der Erläuterung des Problems sei ein kurzer Exkurs über den Kristallaufbau voran gestellt [4]. Abb. 5 zeigt die Grundstruktur des Kristallgitters von Galliumarsenid, bestehend aus drei kubischen Untergittern mit gleichen Kantenlängen, die periodisch fortgesetzt werden. Der ideale Kristall besteht aus gleich vielen Ga- und As-Atomen, die wie folgt auf die Gitterplätze verteilt sind. Das erste Untergitter ist vollständig mit Ga-Atomen (blaue Kugeln), das zweite mit ausschließlich As-Atomen (rote Kugeln) besetzt, und das dritte Untergitter ist leer (weiße Kugeln). Man sagt, es ist mit Vakanzan besetzt. Dieser so aufgebaute Kristall ist allerdings kein Halbleitator, wie er für die Anwendungen benötigt wird. Dieser entsteht aus einer Schmelze mit leicht erhöhtem Arsenanteil; zum Beispiel 0.500082 bei Freiberger Compound Materials (FCM), einem der weltweit größten Hersteller von halbleitendem GaAs. Aus dieser Konzentration ergibt sich auch eine andere Verteilung der Atome auf die Untergitter. Die Plätze des ersten Untergitters werden von allen vorhandenen Galliumatomen, von einigen Arsenatomen und von Vakanzan eingenommen, während das zweite und dritte Untergitter nun sowohl von Arsenatomen als auch von Vakanzan besetzt sind. Diese Verteilung führt aber bei

Abb. 4: Kristalle und Wafer aus GaAs



Abb. 5: Untergitterstruktur des GaAs-Kristalls



der Wärmebehandlung zur Diffusion der Arsenatome, und dies ist letztlich die Ursache der unerwünschten Tropfenbildung.

Zur Beschreibung der Diffusionsvorgänge mit Tropfenbildung gibt es bis jetzt bereits sieben Variable, nämlich die Teilchenzahldichten der Atome auf den jeweiligen Untergittern. Hinzu kommen die N Radien der N Tropfen, mit $N \approx 10^6$, und die Verschiebungen des umgebenden Kristallgitters in den drei Raumrichtungen. Diese Verschiebungen entstehen dadurch, dass die Oberflächen der Tropfen mit Energie versehen sind, und darüber hinaus benötigt ein Tropfen, aufgrund des Dichteunterschiedes zwischen Kristall und Flüssigkeit, mehr Platz als der Kristall. Die reale Situation ist leider noch komplizierter, denn die soeben beschriebene Situation mit sieben Teilchenzahldichten stellt eine Vereinfachung der realen Verhältnisse dar. Die von FCM hergestellten Kristalle enthalten elf weitere Fremdstoffe in kaum messbarer Menge, die aber trotzdem berücksichtigt werden müssen, beispielsweise tritt Silizium mit einer Konzentration 10^{-9} auf.

Es ist also klar, dass eine Reduktion des vollständigen mathematischen Modells vorgenommen werden muss. Aber trotz vielfältiger Experimente, inklusive mehrerer experimenteller Dissertationen, das Regime, in welchem die Evolution der Tropfen stattfindet, konnte bisher nicht ermittelt werden. Hier setzt die mathematische Modellierung ein, welche in diesem Fall eine Hierarchie von testbaren Modellen aufstellen kann.

Das einfachste Modell betrachtet das Diffusionsproblem in der Umgebung eines einzelnen Tropfens in einem Kristall. Es sind zwei Klassen von möglichen Randbedingungen an der Phasengrenze zwischen Flüssigkeit und Kristall untersucht worden, und ein Vergleich der Resultate mit den wenigen vorliegenden experimentellen Daten hat bereits eine Klasse von Randbedingungen ausgeschlossen, während die andere Klasse noch im Rennen ist.

Das einfachste Vieltropfenmodell ist das Becker-Döring-Modell. Es wurde ursprünglich konzipiert, um Nukleation von Nebel in Luft zu simulieren. Dreyer & Duderstadt haben dieses Modell auf das Entstehen flüssiger Tropfen in einem Kristall erweitert und auf das Galliumarsenid-Problem angewandt [3]. Das Modell besteht aus einem großen System gewöhnlicher Differentialgleichungen und berechnet zu jeder Zeit die Größenverteilung der im System vorhandenen Tropfen. Aus Sicht der Mathematischen Physik ist das Becker-Döring-Modell hochinteressant. Jedoch es gibt einige Indizien, dass hier das industrielle Problem nicht angemessen repräsentiert wird.

Ein anderes Vieltropfenmodell reduziert das vollständige, natürlich nicht lösbare, Diffusionsproblem mit vielen Tropfen auf ein homogenisiertes Modell, welches nur noch die Evolution des mittleren Tropfenradius berechnet. Zur Zeit gibt es Hinweise, dass die Realität hier gut abgebildet wird.

Aber worin liegt der praktische Wert dieser Studien? Was nützt dem industriellen Anwender die Erkenntnis, dass die Reduktion des vollen Modells durch Homogenisierung dem beobachteten Phänomen sehr nahe kommt? Nun, jedes der hier vorgestellten Modelle identifiziert andere Größen als die entscheidenden Parameter, durch die der unerwünschte Prozess der Tropfenbildung beeinflusst werden kann. Wird also beispielsweise

vorausgesagt, dass eine Änderung des Druckes keinen Einfluss auf die unerwünschten Tropfen hat, sehr wohl aber eine Variation des Temperaturprofils bei der Wärmebehandlung, so reduziert sich der experimentelle Aufwand zur Ermittlung der Ursachen des Phänomens beträchtlich.

Der Biegetest für GaAs Wafer. Ein Wafer ist bei der Weiterverarbeitung verschiedenen Prozessen ausgesetzt, die aufgrund mannigfacher mechanischer Belastungen sogar zum Bruch führen können. Der Waferhersteller muss deshalb dem Kunden eine bestimmte Bruchfestigkeit garantieren, und dessen Ermittlung erfordert neben einem Bruchtest subtile mathematische Methoden zur Berechnung der nicht direkt messbaren Spannungen, die im Wafer während des Testes auftreten.

Abb. 7 zeigt das Prinzip des Testes [5]. Der runde Wafer mit Durchmesser 15 cm und Dicke 0,5 mm liegt zentrisch auf einem Stützring. Über eine Stahlkugel wird eine

Abb. 6: Mathematischer Kontext zum Becker-Döring-Modell

Das Becker-Döring-Modell beschreibt die Evolution der Zahl von Tropfen $Z(t, \alpha)$ mit α Atomen durch das System

$$\frac{\partial Z(t, \alpha)}{\partial t} = J_{\alpha-1} - J_{\alpha}$$

mit $J_{\alpha} = \Gamma_{\alpha}^C Z(t, \alpha) - \Gamma_{\alpha+1}^E Z(t, \alpha + 1)$,
 $\alpha \in \{2, 3, \dots, \infty\}$,

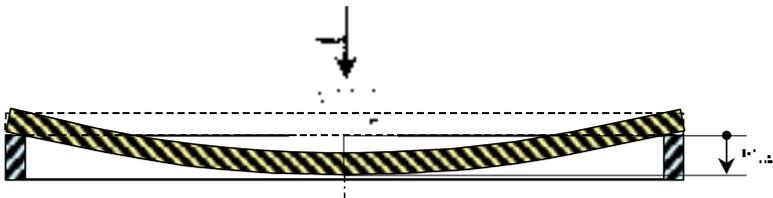
wobei $Z(t, 1)$ durch die Erhaltung der Gesamtteil-

chenzahl

$$N = \sum_{\alpha=1}^{\infty} \alpha Z(t, \alpha)$$

bestimmt wird. Der Fluss J_{α} gibt die Rate der produzierten Tropfen mit $\alpha + 1$ Atomen an, und Γ_{α}^C , Γ_{α}^E bezeichnen die Kondensationsrate bzw. die Verdampfungsrates. Diese beiden Größen hängen, je nach Anwendungsfall, in komplizierter Weise von α und von den Tropfenzahlen $Z(t, \alpha)$ ab, so dass ein nichtlineares und manchmal auch nichtlokales System von gewöhnlichen Differentialgleichungen entsteht.

Abb. 7: Querschnitt eines Ring-Platte-Last-Systems (aus [5])



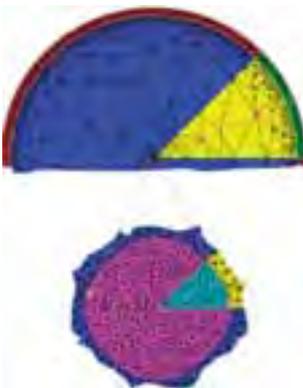
Kraft aufgebracht, und es kommt zu einer Biegung des Wafers. Gemessen wird die zu einer gegebenen Kraft gehörige Durchbiegung des Wafers im Zentrum.

Mit diesem Test kann natürlich auch die Kraft ermittelt werden, bei der es zum Bruch der Platte kommt. Der resultierende Wert ist aber nur bedingt eine Kennzahl für die Bruchfestigkeit, da er nicht ausschließlich die Materialeigenschaft bewertet, sondern auch abhängt von den Abmessungen des Wafers und den Eigenschaften des Testes. Eine nur materialabhängige Kennzahl gewinnt man aus den im Wafer auftretenden Spannungen, und diese lassen sich nicht messen, sondern nur berechnen.

Die Berechnung der Spannungen basiert auf der Beobachtung, dass bei Wafers aus Galliumarsenid ausschließlich Sprödbrech auftritt, was bedeutet, dass dem plötzlich auftretenden Bruch nur eine *elastische* Deformation vorausgeht. Somit können die Spannungen mittels der Gleichungen der Elastizitätstheorie berechnet werden. Aufgrund des großen Unterschiedes zwischen Durchmesser und Dicke des Wafers kommt aber eine Finite-Element-Approximation der dreidimensionalen Elastizitätsgleichungen nicht in Frage. Zunächst muss eine Reduktion dieser Gleichungen auf ein zweidimensionales Plattenmodell durchgeführt werden.

Eine sorgfältige Analyse der vielen möglichen Modellreduktionen motiviert für den vorliegenden Fall das von Kármán-Modell als das adäquate Plattenmodell, und hierfür wurde eine Finite-Element-Approximation durchgeführt. Eine typische Vernetzung ist in Abb. 8 eingezeichnet ([5]). Man beachte die Verdichtung des Netzes auf der sehr kleinen Kontaktfläche. Sehr schön läßt sich hier auch die kubische Symmetrie des Kristalls erkennen. Der Abdruck der Lastkugel auf dem Wafer ist kein Kreis, sondern zeigt die Kristallsymmetrie bei Drehungen um 90° .

Abb. 8: Finites-Element-Netz des Plattenproblems ([5]). Oben: Hälfte des vollen Netzes. Unten: Netz der Lastfläche mit 350-fachem Zoom



Ein heikler Punkt sind übrigens die Randbedingungen auf der Kontaktlinie Waver-Stützring. Es ist nämlich dem Experiment nicht zu entnehmen, ob der Wafer hier während der Lastaufbringung nach innen rutscht oder liegen bleibt. Deshalb werden mittels des mathematischen Modells einfach beide Möglichkeiten über geeignete Randbedingungen simuliert. Die linke Abb. 9 zeigt den Unterschied auf, und der nebenstehende Vergleich mit den experimentellen Daten zeigt, dass der mathematisch schwierigere Fall des Rutschens im Experiment realisiert ist. Interessant ist auch das völlige Versagen des linearen Plattenmodells.

Abschließend soll hier noch von einem subtilen Phänomen berichtet werden, dessen Ignorierung zu Berechnungsfehlern in der Größenordnung von 40 % führen kann [1]. Es ist das Babuška-Paradox, welches unter Mathematikern bekannt ist, aber überraschenderweise von Ingenieuren aus Unkenntnis ignoriert wird. Das Paradox hängt mit der Diskretisierung des Kreises zusammen, welcher durch die Kontaktlinie Waver-Stützring generiert wird. Es ist natürlich klar, dass ein Polygon einen Kreis um so besser approximiert, je mehr Polygonzüge es gibt. Trotzdem approximiert das Randwertproblem für das Polygon das Randwertproblem für den Kreis um so schlechter, je feiner die Vernetzung ist, jedenfalls dann, wenn geradlinige Elemente verwendet werden. Dies ist das Babuška-Paradox. Man kann aber Methoden entwickeln, dass es sogar bei geradlinigen Elementen nicht auftritt, aber natürlich nur dann, wenn das Problem als solches erkannt wird.

Alterung – Schädigung – Bruch. Am Ende der Produktionskette von Halbleitern steht ihr Einsatz in elektronischen Geräten. Hier treten nun Phänomene auf, die typisch für ein gänzlich anderes Thema sind. Es geht um Alterung, Schädigung und Bruch, was schließlich zu einem Totalausfall des betreffenden Bauteiles führt. Weil dem Bruch aber immer

Abb. 9: Maximale Durchbiegung vs. Kraft. Links: Vergleich verschiedener Modelle. Rechts: Vergleich Theorie und Experiment.

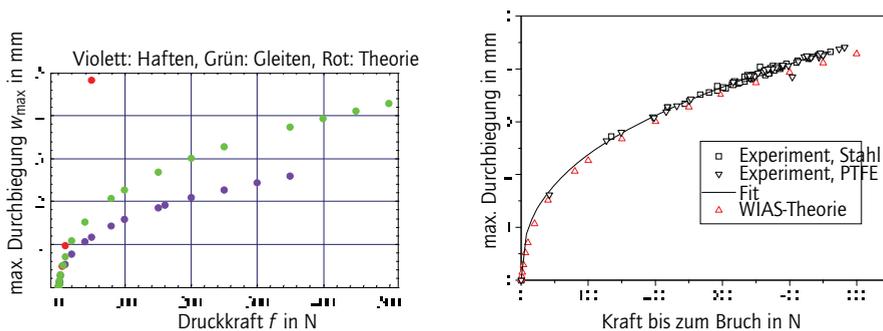


Abb. 10: Mathematischer Kontext zum Biegetest für GaAs Wafer ([5])

Das System der Plattengleichungen nach von Kármán besteht aus drei gekoppelten nichtlinearen partiellen Differentialgleichungen vom elliptische Typ für die beiden Verschiebungen $U = (U_1(X_1, X_2), U_2(X_1, X_2))$ der Wafermittelfläche und der vertikalen Verschiebung $W(X_1, X_2)$, wobei (X_1, X_2) die Lagrange-Koordinaten der Mittelfläche sind. Für eine gegebene äußere Last $p(X_1, X_2)$ stellen die Gleichungen

$$\frac{\partial N_{\alpha\beta}}{\partial X_\beta} = 0, \quad \frac{\partial^2 M_{\alpha\beta}}{\partial X_\alpha \partial X_\beta} + \frac{\partial^2 W}{\partial X_\alpha \partial X_\beta} N_{\alpha\beta} = -p, \quad \alpha, \beta \in \{1, 2\},$$

das mechanische Gleichgewicht sicher. Diese werden komplettiert durch die Spannungs-Dehnungs-Beziehungen für kubisch anisotropes Galliumarsenid mit den drei Lamé-Konstanten λ, μ und μ' :

$$N_{\alpha\beta} = h \left(\lambda \frac{2\mu + \mu'}{2\mu + \mu' + \lambda} G_{\gamma\gamma} \delta_{\alpha\beta} + 2\mu G_{\alpha\beta} + \mu' \delta_{\alpha\beta} G_{\beta\beta} \right) \quad \text{mit} \quad G_{\alpha\beta} = \frac{1}{2} \left(\frac{\partial U_\alpha}{\partial X_\beta} + \frac{\partial U_\beta}{\partial X_\alpha} \right),$$

$$M_{\alpha\beta} = -\frac{h^3}{12} \left(\lambda \frac{2\mu + \mu'}{2\mu + \mu' + \lambda} \frac{\partial^2 W}{\partial X_\gamma \partial X_\gamma} \delta_{\alpha\beta} + 2\mu \frac{\partial^2 W}{\partial X_\alpha \partial X_\beta} + \mu' \delta_{\alpha\beta} \frac{\partial^2 W}{\partial X_\beta \partial X_\beta} \right).$$

Zu diesem System treten noch Randbedingungen, die hier aber nicht angegeben werden.

Abb. 11: Lotkugeln auf einem Chip



größere Veränderungen der Eigenschaften der beteiligten Materialien vorausgehen, sie *altern*, ist er kein unvorhersehbares Ereignis, sondern lässt sich bei genauer Kenntnis der zugrunde liegenden Mechanismen vorhersagen. Allerdings fehlt dem Technologen oft die Zeit, den Alterungsvorgang im Detail zu verstehen und stattdessen zieht dieser pragmatisch-phänomenologische Lösungen heran, obwohl ein tieferes Durchdringen des Problems langfristig zu besseren Ergebnissen führen würde.

Zur Illustration der ganzen Problematik sei hier die Herstellung und der Einsatz von Löt, die den Chip mit der Außenwelt verbinden, betrachtet. Das bisher populärste Lot besteht aus einer Legierung aus Zinn und Blei, weil in diesem Fall zwei notwendige, aber sich eigentlich widersprechende Prämissen einigermaßen gut erfüllt werden. Zur Herstellung des elektronischen Bauteils muss der Schmelzpunkt des Lotes niedrig sein. Ist aber

das Bauteil im Einsatz, wird aufgrund der dann auftretenden hohen Betriebstemperatur ein hoher Lotschmelzpunkt notwendig.

Aufgrund der Giftigkeit des Bleis ist seit 2006 das Zinn-Blei-Lot aus der großtechnischen Produktion gebannt, und es muss ein Ersatz geschaffen werden. Die letzten zwei Jahre haben gezeigt, dass dies kein einfaches Unterfangen ist. Übrigens, das Versagen eines Chips, einer Photozelle oder eines Airbags geht zu 90 % auf das Versagen der eingesetzten Lote zurück. Der volkswirtschaftliche Schaden ist immens. Nach einer Schätzung der deutschen Umweltorganisation BUND lassen sich allein in Deutschland mit dem jährlich anfallenden Elektronikschrott vier Cheopspyramiden (Höhe 140 m) füllen.

Wo liegen nun aber die Ursachen des Versagens der Lote? Grob gesagt trifft man auf zwei unterschiedliche Probleme. Aufgrund der hohen Betriebstemperatur, die 70 % der Schmelztemperatur betragen kann, gibt es eine Entmischung der beteiligten Stoffe. Wohingegen das andere Problem bereits bei der Herstellung des Kontaktes, beispielsweise zwischen dem Lot und einem Kupfersubstrat, auftritt.

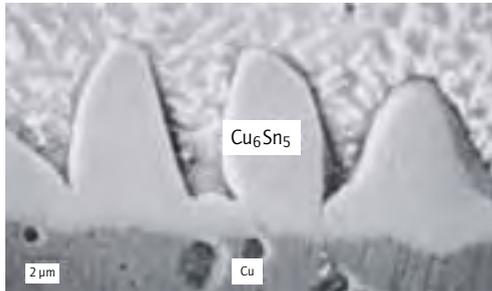
Abb. 12: Wohin mit dem Elektronikschrott?



Abb. 13: Entmischung bei hoher Temperatur (aus [2]): Links: nach Erstarrung. Mitte: nach 5 Stunden. Rechts: nach 40 Stunden



Abb. 14: Unerwünschte Zwischenschicht zwischen Substrat und Lot (aus [9])



Sofort nach der Erstarrung kommt es im Innern des Lotes zu einer Entmischung der chemischen Bestandteile [2]. Im Fall des Beispiels aus Abb. 13 bilden sich silberreiche Inseln in einer kupferreichen Umgebung. Es ist wichtig, die zeitliche Entwicklung zu kennen, denn die Erfahrung zeigt, dass mit der Vergrößerung der Lotstruktur die verstärkte Bildung kleiner Risse längs der Phasengrenzen einhergeht. Deshalb ist es ein großer Erfolg, dass diese Evolution mittlerweile in realistischer Weise mittels eines mathematischen Modells simuliert werden kann. Zu diesem Zweck wird ein sogenanntes Phasenfeldmodell verwendet, welches die in Abb. 13 auftretenden scharfen Grenzen zwischen den Inseln und der Umgebung leicht diffus modelliert. Dieser Trick trägt erheblich zu der Fähigkeit bei, derart komplexe Strukturen überhaupt berechnen zu können.

Das zur Zeit vielversprechendste bleifreie Lot besteht aus den drei chemischen Komponenten Zinn, Silber und Kupfer. Wird dieses Lot im flüssigen Zustand in Kontakt mit dem festen Kupfersubstrat gebracht, bildet sich allerdings unerwünscht eine äußerst spröde Schicht mit dem chemischen Namen Cu_6Sn_5 . Insbesondere die geometrische Form der Grenze zwischen der Zwischenschicht und dem Innern des Lotes wird als bruchauslösend vermutet. Eines der Ziele ist die Identifizierung der entscheidenden Parameter, so dass diese Grenze einigermaßen eben wird. Auch hierzu gibt es bereits das mathematische Modell, aber nicht in der Version mit diffusen sondern mit scharfen Grenzflächen.

3 STATUS QUO

Die dargestellten Erfolgsgeschichten betreffen den Einsatz von angewandter Mathematik bei *modernen Materialien*. Die hierdurch bedingte Arbeitsmethodik wird nun an zwei Beispielen illustriert.

Abb. 15: Mathematischer Kontext zum Thema Alterung-Schädigung-Bruch (aus [2])

Die Evolution der Mikrostruktur des Lotinnern in Zeit und Raum wird für ein Lot mit zwei chemischen Komponenten beschrieben durch eine Konzentration $c(t, X)$ und die Verschiebung des Kristallgitters $u = (u_1(t, X), u_2(t, X), u_3(t, X))$. Das Phasenfeldmodell besteht aus einem parabolisch-elliptischen System von Differentialgleichungen der Form

$$\rho_0 \frac{dc}{dt} + \frac{\partial J^i}{\partial X^i} = 0 \quad \frac{\partial \sigma^{ij}}{\partial X^i} = 0,$$

welche die Erhaltung der Masse und mechanisches Gleichgewicht beschreiben. Hierzu treten zwei neuartige Gleichungen für den Diffusionsfluss und den Spannungstensor, die speziell die Verhältnisse in einem Lot beschreiben:

$$j^i = -\rho_0 \mathcal{M}^{ij}(T) \frac{\partial}{\partial X^j} \left(\frac{\partial \psi}{\partial c} - 2A^{kl} \frac{\partial^2 c}{\partial X^k \partial X^l} - \frac{\partial A^{kl}}{\partial c} \frac{\partial c}{\partial X^k} \frac{\partial c}{\partial X^l} - \frac{\partial A^{kl}}{\partial E^{mn}} \frac{\partial c}{\partial X^k} \frac{\partial E^{mn}}{\partial X^l} - \frac{\partial a^{kl}}{\partial E^{mn}} \frac{\partial^2 E^{mn}}{\partial X^k \partial X^l} - \frac{\partial^2 a^{kl}}{\partial E^{op} \partial E^{mn}} \frac{\partial E^{op}}{\partial X^k} \frac{\partial E^{mn}}{\partial X^l} \right),$$

$$\sigma^{ij} = K^{ijkl} (E^{kl} - E_*^{kl}) - \frac{\partial a^{kl}}{\partial E^{ij}} \frac{\partial^2 c}{\partial X^k \partial X^l} - \frac{\partial b^{kl}}{\partial E^{ij}} \frac{\partial c}{\partial X^k} \frac{\partial c}{\partial X^l}.$$

Zu diesem System gehören noch Randbedingungen, aber nur auf dem Außenrand des Lotes. Würde hier ein Modell mit scharfen Grenzflächen verwendet, wären die Differentialgleichungen von viel einfacherer Struktur, aber stattdessen werden Übergangsbedingungen für jeden Grenzflächenpunkt benötigt.

Bei einigen Kristallzuchtverfahren befindet sich das Schmelze-Kristall-System in einem Behälter, dessen Wand mit einer dünnen Polymerschicht versehen ist. Wo nun Schmelze und Kristall an der selben Stelle auf die Polymerschicht treffen, wird eine sogenannte freie Triplekurve mit den drei beteiligten Materialien induziert. Diese Linie ist nicht gegeben, sondern Bestandteil der Lösung des mathematischen Modells. Das Problem ist nun, dass die Lage und Form der Tripellinie von entscheidendem Einfluss auf die Kristallgüte ist, nur sind leider die relevanten Gleichungen auf dieser Linie nicht vollständig bekannt. Insbesondere liegt dies an komplizierten mechanischen Deformationen, die durch die Elastizität der beteiligten Materialien ausgelöst werden, und die sowohl mit dem Stofftransport als auch mit der Wärmeleitung wechselwirken. Bevor also die angewandte Mathematik ihren eigentlichen Beitrag leisten kann, setzt zunächst ein subtiler und häufig langwieriger Prozess ein, der einhergeht mit Iterationen zwischen physikalischer Modellierung und numerischen Tests.

Ein weiteres Beispiel betrifft die Verfügbarkeit von Materialdaten. Die Simulation der Evolution der unerwünschten Tropfen in halb-isolierendem Galliumarsenid benötigt etwa 50 Kenngrößen, und weil es sich hier um ein *modernes Material* handelt, sind viele dieser Größen nicht verfügbar. Dies resultiert zum Teil aus dem Phänomen, dass viele

Materialkenngrößen nicht direkt messbar sind, sondern zu ihrer Bestimmung bereits ein theoretisches Modell benötigen. Hier aber ist die Analysis von großer Bedeutung, da durch geschickte Anwendung eine Klassifizierung in wichtige und unwichtige Materialkenngrößen möglich ist.

Dies sind typische Problemfelder bei der mathematischen Modellierung moderner Materialien; sie sollen natürlich suggerieren, dass eine starke interdisziplinäre Verzahnung unabdingbar für eine erfolgreiche Behandlung dieses Gegenstandes ist. Zur Zeit gibt es aber diese Verzahnung nur in ersten Ansätzen. Die Industrie stellt für ihre Probleme im allgemeinen keine interdisziplinären Konsortien zusammen. Stattdessen werden deren Forschungsgelder meistens einzelnen disjunkten Fachgebieten übertragen. Das oben angeführte Beispiel zur Kristallzucht im Magnetfeld stellt leider immer noch einen Ausnahmefall dar. Aber auch die Deutsche Forschungsgemeinschaft hat zur Zeit zu wenig Instrumente zur Förderung interdisziplinärer Projekte. Beispielsweise sind die Sonderforschungsbereiche mit interdisziplinärer Fragestellung zu groß und schwerfällig für eine industrielle Fragestellung. Ihre Behandlung ist in Kleingruppen mit zwei oder drei unterschiedlichen Partnern viel besser aufgehoben, nur scheint deren Förderung verwaltungstechnisch zu aufwendig. In diesem Zusammenhang sollte auch die Dauer eines interdisziplinären Projektes diskutiert werden. Sie ist zur Zeit häufig dem industriellen Nutzer der Ergebnisse zu lang! Es könnten hier viele Beispiele zu folgendem Sachverhalt angeführt werden. Die mathematische Modellierung löst ein gestelltes Problem auf hervorragende Weise; nur ist niemand mehr an der Lösung interessiert, da die Industrie das ursprüngliche Problem mittlerweile durch andere Sichtweise pragmatisch wegschaffen konnte.

4 STÄRKEN/SCHWÄCHEN-ANALYSE, HERAUSFORDERUNGEN

Die soeben geschilderte Problematik trifft die Behandlung neuer Materialien in viel stärkerer Weise als andere Themen, wo die angewandte Mathematik fast immer großartige Erfolge aufweisen kann. Auch dies sei noch einmal an einem Beispiel erläutert.

Die *aktive Strömungsbeeinflussung* von Fahrzeugen, siehe das entsprechende Kapitel, verwendet als zentrales mathematisches Modell die Navier-Stokes-Gleichungen für Wasser oder für Luft, und darüber hinaus stehen auch die möglichen Randbedingungen im Prinzip seit langer Zeit fest. Dies gilt selbst dann, wenn auftretende Turbulenz mit reduzierten Modellen beschrieben wird. Mit anderen Worten: Das Problem der Übersetzung des betrachteten technischen Prozesses in ein mathematisches Modell geschieht nach etablierten und einfachen Regeln. Zudem werden für Wasser bzw. Luft, um die es hier geht, nur wenige und zudem bekannte Materialkenngrößen benötigt. Es sind dies höchstens drei Zahlen, nämlich die Viskosität, der Ausdehnungskoeffizient und eventuell ein Gleitreibungskoeffizient, der in die Randbedingungen eingeht.

Wenn es um moderne Materialien geht, ist dies jedoch anders. Hier geht es nicht darum vorhandene Modellgleichungen hoher Komplexität möglichst effizient und in

möglichst kurzer Zeit zu lösen. Die momentane Hauptaufgabe der Mathematik ist vielmehr ein Einsatz bei der Modellbeschaffung im Sinne der obigen Beispiele.

Zur Zeit werden die hierzu notwendigen Techniken nach Meinung des Autors an deutschen Universitäten kaum oder nicht sachgemäß vermittelt. Die Modellierung in der Mechanik ignoriert die meisten thermodynamischen Aspekte, die Thermodynamik behandelt im allgemeinen die Mechanik in einer zu stark vereinfachten Form, und zur Erlangung möglichst vieler expliziter Lösungen des Modells vereinfacht die Physik häufig in unrealistischer Weise die reale Geometrie des betrachteten Körpers. Aber auch die Modellierung der Mathematiker lässt viele Wünsche offen. Hier ist es nämlich zu oft das Ziel, mit einem Modell zu enden, für das man bereits die mathematischen Methoden beherrscht. Dies aber sollte kein Kriterium für realistische Modellierung sein. Fazit: Die Interdisziplinarität muss verstärkt ausgebaut werden.

5 VISIONEN UND HANDLUNGSEMPFEHLUNGEN

Problemlösungen in dem Gebiet *Herstellung und Einsatz moderner Materialien* bestehen aus den folgenden Punkten:

- Mathematische Modellierung inklusive der Beschaffung der Materialdaten;
- Analyse der mathematischen Eigenschaften der Modellgleichungen;
- Entwicklung numerischer Algorithmen inklusive zuverlässiger Fehlerabschätzungen;
- Numerische Simulation, Visualisierung und Interpretation der Lösungen.

Üblicherweise wird diese Aufgabenkette mehrmals durchlaufen und das Fazit ist deshalb: Nur eine interdisziplinär zusammengesetzte Arbeitsgruppe wird auf diesem Gebiet erfolgreich sein.

6 LITERATUR

- [1] I. Babuška and J. Pitkäranta. The plate paradox for hard and soft simple support. *SIAM J. Math. Anal.*, 29(5):1261–1293, 1992.
- [2] Th. Böhme. *Investigations of Microstructural Changes in Lead-Free Solder Alloys by Means of Phase Field Theories*. PhD thesis, Technical University of Berlin, 2008.
- [3] W. Dreyer and F. Duderstadt. On the Becker–Döring theory of nucleation of liquid droplets in solids. *J. Stat. Phys.*, 123:55–87, 2006.

- [4] W. Dreyer and F. Duderstadt. On the modelling of semi-insulating GaAs including surface tension and bulk stresses. *Proc. Roy. Soc. London Ser. A*, 2008. 10.1098/rspa.2007.0205.
- [5] W. Dreyer, F. Duderstadt, St. Eichler, and M. Jurisch. Stress analysis and bending tests for GaAs wafer. *Microelectron. Reliability*, 46:822–835, 2006.
- [6] P.E. Druet. On weak solutions to the stationary mhd-equations coupled to heat transfer with nonlocal radiation boundary conditions. *WIAS-Preprint*, 29(1290):1–44, 2008.
- [7] Ch. Frank-Rotsch, P. Rudolph, P. Lange, O. Klein, and B. Nacke. Vorrichtung und Verfahren zur Herstellung von Kristallen aus elektrisch leitenden Schmelzen (Anordnung und Verfahren zur optimierten Schmelzzüchtung von Kristallen). Patent application DE 10 2007 028 548.7. submitted June 18th 2007.
- [8] M. Herrmann, M. Naldzhieva, and B. Niethammer. On a thermodynamically consistent modification of the Becker–Döring equation. *Phys. D*, 222:116–130, 2006.
- [9] H.K. Kim and K.N. Tu. Kinetic analysis of the soldering reaction between eutectic snpb alloy and cu accompanied by ripening. *Physical Review B*, 53:16027–16034, 1996.
- [10] O. Klein, Ch. Lechner, P.-É. Druet, P. Philip, J. Sprekels, Ch. Frank-Rotsch, F.-M. Kießling, W. Miller, U. Rehse, and P. Rudolph. Numerical simulations of the influence of a travelling magnetic field, generated by an internal heater-magnetic module, on Czochralski crystal growth. In *Proceedings of the International Scientific Colloquium Modelling for Electromagnetic Processing (MEP 2008), Hannover, October 27–29, 2008*.
- [11] P. Rudolph. Travelling magnetic fields applied to bulk crystal growth from the melt: The step from basic research to industrial scale. *J. Crystal Growth*, 310(7–9):1298–1306, 2008.

> TOPOLOGIE UND DYNAMISCHE NETZWERKE: ANWENDUNGEN DER OPTIMIERUNG MIT ZUKUNFT

GÜNTER LEUGERING, ALEXANDER MARTIN UND MICHAEL STINGL

1 EXECUTIVE SUMMARY

Die optimale Auslegung von Infrastrukturen z. B. bei der Verkehrsplanung und bei der Planung von Versorgungssystemen, die optimale Strukturierung bzw. Formgebung von Materialien und Werkstücken z. B. im Leichtbau sind aktuelle Themen angewandter Forschung. In beiden Bereichen wurde bis in die jüngste Zeit vornehmlich eine simulationsbasierte Optimierung auf der Grundlage einer Parameterjustierung vorgenommen, die oft wenig systematisch und zeit- und kostenintensiv ist. Stattdessen erweisen sich modellbasierte mathematische Optimierungsalgorithmen zusammen mit moderner numerischer Simulations- und Visualisierungstechnologie zunehmend als Katalysator neuer Technologien. Eine so verstandene Mathematische Optimierung kann bereits auf beeindruckende Erfolgsgeschichten verweisen und so den Anspruch als eine Zukunftsdisziplin behaupten. Diesem Anspruch trägt die Einrichtung des DFG-Schwerpunktprogramms 1253 ‚Optimierung mit partiellen Differentialgleichungen‘ im Jahre 2006 Rechnung, in dem über 25 Projekte im Bundesgebiet sowohl auf die theoretische Fundierung, als auch und insbesondere auf die Verzahnung zwischen Methodenentwicklung und numerischer Realisierung fokussieren. Forschung im Bereich der mathematischen Optimierung und Steuerung von Prozessen bzw. Eigenschaften, die mit Hilfe partieller Differentialgleichungen, so genannten ‚verteilten Systemen‘, beschrieben werden, erfolgt im Kontext konkreter und exemplarischer Anwendungssituationen, die neue mathematische Herausforderungen markieren: Sicherheitsvorgaben etwa bei der Belastung von Gas- und Frischwasserleitungen oder solche für die Belastbarkeit von Verkehrsflugzeugen führen auf Druckbeschränkungen bzw. Materialschränken, die in Form von Ungleichungen bei der Modellierung, der Optimierung und der Simulation zu berücksichtigen sind und eine neuartige Herausforderung an die Mathematik darstellen. Entscheidungsgetriebene Prozesse, etwa im Gas-, Wasser- oder Verkehrsnetzbetrieb, tragen Unstetigkeiten bzw. diskrete Komponenten in die mathematische Beschreibung hinein und führen damit auf völlig neuartige Modelle und Verfahren, die diskrete und kontinuierliche Methoden verzahnen.

In diesem Beitrag werden erfolgreiche Anwendungsfälle im Bereich der Struktur- und Topologieoptimierung sowie der Optimierung von vernetzten Transportsystemen vorgestellt.

Die Optimierung von Infrastrukturen und die ressourcenschonende Materialoptimierung sind Kerngebiete künftiger Forschung, die nach Einschätzung der Autoren durch BMBF- und DFG-Förderungsstrukturen unterstützt werden müssen, um den Standortvorteil im internationalen Vergleich zu wahren und auszubauen.

2 VON DEN EIGENSCHAFTEN ZU OPTIMALEN STRUKTUREN

‚Strukturbildung‘ und deren aktive zielorientierte Beeinflussung zur Herstellung funktionaler Werkstoffe ist zentraler Gegenstand der aktuellen Forschung im Bereich der Materialwissenschaften, im Chemie- und Bioingenieurwesen bis hin zur Theoretischen Chemie und Physik. *Engineering of Advanced Materials: Hierarchical Structure Formation of Functional Devices* ist der Titel eines seit November 2007 an der Universität Erlangen durch die DFG geförderten Exzellenzclusters. Die damit assoziierten Begriffe ‚Neue Materialien‘, ‚Metamaterialien‘, ‚Materialschäume‘, ‚druckbare Elektronik‘ etc. haben mittlerweile Einzug in den Sprachgebrauch auch populärwissenschaftlicher Veröffentlichungen gefunden und gehören zum festen Wortschatz der Zukunftstechnologien.

Die Versorgung mit Wasser, Strom und Gas ist ein Problem der zukünftigen Gesellschaft von schicksalhafter Bedeutung. Erhalt und Ausbau von Verkehrs- und Kommunikationsnetzwerken werden die Mobilität künftiger Generationen bestimmen. Kurz: die Funktionalität von Infrastrukturen rückt mehr und mehr in den Mittelpunkt.

Um die kontextübergreifende Katalysatorfunktion der Mathematik und insbesondere der mathematischen Optimierung für diese Zukunftstechnologien zu verstehen, muss man zunächst einen Schritt von der unmittelbaren technischen Realität zurück treten und die ökonomischen Bedingungen reflektieren. So wird schnell deutlich, dass die Modellierung, Simulation, Optimierung und Visualisierung realer Prozesse, sozusagen deren Antizipation auf dem Computer, durch geeignete Modelle und deren quantitative numerische Umsetzung in Form von Algorithmen eine klassische Prototypen-orientierte Produktentwicklung erweitert bzw. bereits abgelöst haben. Basierend auf den dramatisch zunehmenden Möglichkeiten moderner Computertechnologie, gepaart mit einer ebenso rasanten Entwicklung moderner Optimierungs- und Simulationsalgorithmen, ist es nunmehr möglich, den umgekehrten Schritt, den von der Funktion zur Struktur eines Materials oder eines Prozesses zu gehen. Salopp gesagt: Gegeben die gewünschten Eigenschaften eines Lackes (Brillanz, Kratzfestigkeit, Pflegeleichtigkeit etc.), gesucht die Struktur der Nano-Partikel, die den Lack wesentlich bestimmen. Gegeben die Notwendigkeit der Gewichtsreduktion bei Großraumflugzeugen, Schiffen etc., gesucht die Topologie der Struktur bzw. die Charakteristika der Werkstoffe. Gegeben die wechselnden Bedingungen bei den Gasversorgern und den Verbrauchern, gesucht die Stellung von Ventilen und die kostengünstigste Kompressorleistung. Dieser neue Weg des ‚reverse enginee-

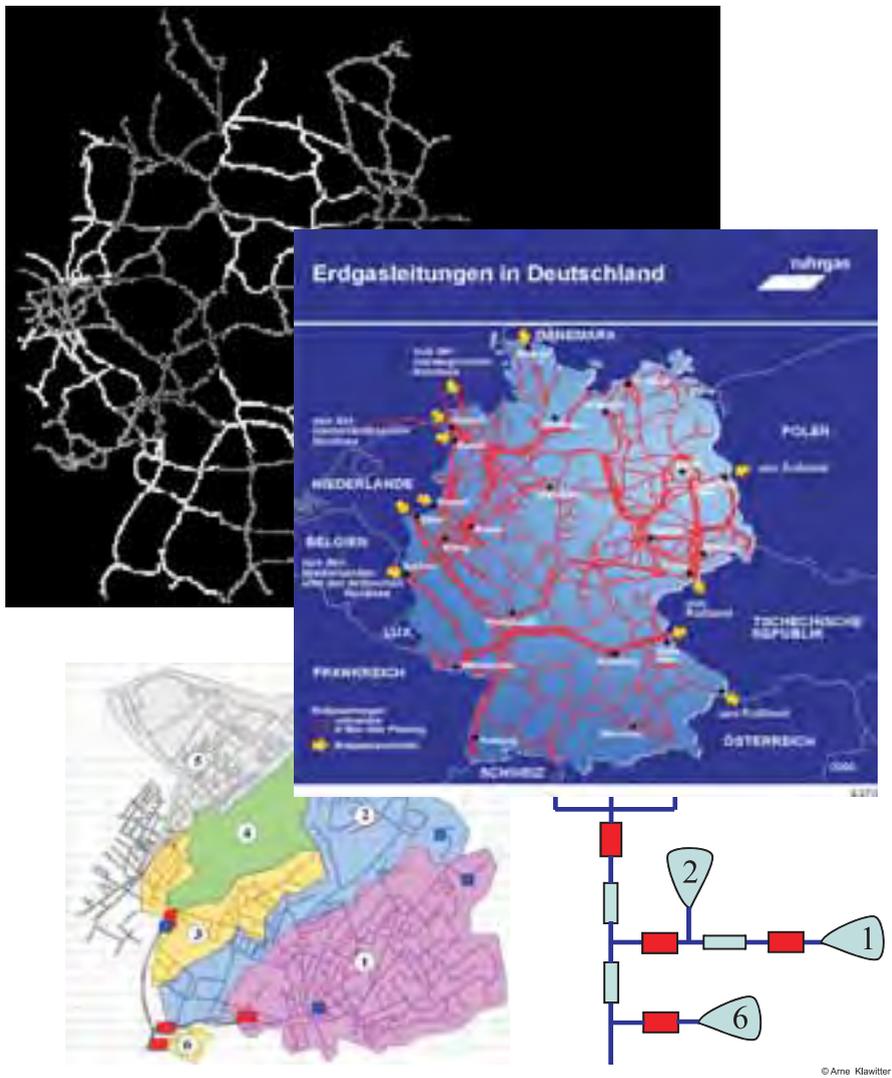
ring', der im Sprachgebrauch der Ingenieure oft auch einfach ‚inverse Optimierung' genannt wird, wird in der ‚Mathematischen Optimierung' mit Erfolg beschriftet. Um die Instrumente der mathematischen Optimierung nutzen zu können, müssen Zielfunktionen und Nebenbedingungen sowie das zu optimierende Objekt selbst in die Begrifflichkeit der Mathematik, letztlich in differential-algebraische Gleichungen oder Ungleichungen, übersetzt werden.

Die Mathematik bietet dabei als Wissenschaft von den Strukturen Denk- und Visualisierungsmodelle an, die in der Phase der Modellierung von fundamentaler Bedeutung sind. Konkret: Wie kann man vielfach verzweigte Partikelstrukturen beschreiben, die sich in einem chemischen Reaktor bilden und die für die Funktionalität von daraus zu bildenden Produkten marktentscheidend sind?

Die mathematische Antwort ist das Konzept des Graphen, dargestellt durch Linien (Kanten), die gegebene Knoten (Ecken) miteinander verbinden. So können ‚Nanoröhrchen' und ‚Nanodrähte', aber eben auch ‚zusammengebackene Partikel' durch Graphen, manchmal durch so genannte ‚fraktale Graphen' dargestellt werden (siehe exemplarisch Abb. 1). Die für die Anwendung entscheidende Vernetzungsstruktur der real beobachteten Objekte spiegelt sich in dem ‚Kantengrad', der Anzahl und der Geometrie der in einen Knoten hineinlaufenden Kanten wider. Das Wechselspiel von lokalen Eigenschaften in den Knoten und Kanten einerseits und dem gesamten Graphen andererseits, ist Gegenstand mathematischer Forschung seit Jahrhunderten, angefangen mit dem Königsberger Brückenproblem von Euler (1736) und der Existenz einer Rundreise, die jede Kante (Brücke) genau einmal enthält (globale Eigenschaft), genau dann, wenn jeder Knoten geraden Grad hat (lokale Eigenschaft). Fragen nach einer ‚optimalen Vernetzungsstruktur', wie sie in der Materialwissenschaft im Zusammenhang mit hierarchischen Materialien (Metall- oder Keramikschäumen) oder in der ‚Photonik' bei so genannten ‚photonischen Kristallen' gestellt werden, finden ihre Umschrift in der Sprache mathematischer Objekte (z. B. Differentialgleichungen) auf Graphen. Vor diesem Strukturhintergrund lässt sich fragen: Warum verzweigen die Kanäle auf Blättern immer auf die gleiche Weise [7]? Wie muss eine mechanische Gitterstruktur aussehen, damit sie möglichst gut gegebenen Lasten standhält [10, 14]? Wie sollte man bei der Planung und Nutzung von Gas-, Wasser- oder Transportnetzwerken vorgehen? Welche Topologie ist optimal?

Topologie lässt sich etwa in der Abwasser-/Frischwasser-Wirtschaft und im Gasnetzbetrieb auch aktiv durch Öffnen oder Schließen von Ventilen steuern. In Verkehrsnetzen wird diese Rolle von Lichtsignalanlagen und online-Schildern (Blockierung/Freigabe von Fahrspuren etc.) übernommen. Das optimale Zusammenspiel zwischen derartigen aktiven Eingriffen und dem dynamischen Betrieb von Kompressoren, Rückstaubecken und anderen Steuerungsmechanismen (etwa Geschwindigkeitsvorgaben im Verkehr) ist eine herausfordernde Aufgabe in der optimalen Auslegung und Betriebsführung von Infrastrukturen (vgl. Abb. 2). Fragen der Robustheit von Transportnetzen mit Blick auf Störungen (Unfälle, Ausfall einer Pipeline oder eines anderen Instruments, Versagen einer

Abb. 2: Infrastrukturen: Transportnetzwerke (Verkehr, Gas, Wasser)



surd belächelt werden mag, hat sich bereits als so genannte ‚Freie Materialoptimierung‘ (FMO) etabliert und ist Gegenstand der kooperativen Forschung im Rahmen eines von der EU geförderten Projekts, an dem fünf global operierende Industrien (EADS, Airbus, Altair, Eurocopter, RISC) und vier Forschergruppen an Universitäten aus vier Ländern, darunter Israel, Dänemark, England und Deutschland beteiligt sind.

Zusammenfassend kann schon hier festgehalten werden: Die Mathematik gewinnt abstrahierend Struktureinsichten: der ‚Schritt zurück‘ aus der Unmittelbarkeit ermöglicht den produktübergreifenden struktur-orientierten Blick, der das Fenster in Richtung anderer Technologien öffnet. Technologietransfer ist in diesem Sinne eine genuin mathematische Angelegenheit. Die Mathematik ist jedoch darüber hinaus auch in der quantitativen Simulation – also in einem Schritt auf den konkreten Anwendungskontext hin – Katalysator für neue Technologien. In dieser polyvalenten Funktion wirkt sie schon ‚hinter den Kulissen‘, schreibt sie am Drehbuch für die Technologie der Zukunft mit.

3 ERFOLGSGESCHICHTE: STRUKTUR-, TOPOLOGIE- UND MATERIALOPTIMIERUNG

Eines der Hauptprobleme in der Strukturmechanik lässt sich folgendermaßen formulieren: *Gegeben seien Auslenkungen oder Kräfte am Rand und innerhalb des betrachteten Gebietes. Gesucht ist die leichteste Struktur, die den gegebenen Belastungen standhält.*

Die Optimierung von Strukturen wird typischerweise über die Variation der Geometrie der beteiligten Strukturelemente (Stäbe, Balken, Membranen, Platten, Schalen etc.) via sogenanntes ‚sizing‘ also einer ‚Querschnitts- bzw. Parameteroptimierung‘ oder über die Veränderung der Strukturgeometrie via ‚shape‘ z. B. mit mathematischen Freiformflächen als ‚Formoptimierung‘ vollzogen. Dieses Verfahren, das in vielen Bereichen immer noch sehr erfolgreich ist und bleiben wird, kann im Bereich ‚neuer Materialien‘, also solcher Materialien, die das Ergebnis einer gezielten ‚Materialkonfiguration‘ sind, durch ein radikaleres Konzept ersetzt bzw. ergänzt werden, das die Optimierung des Materials selbst in den Blick nimmt. Die daraus entstehende Theorie und deren praktische Umsetzung firmiert unter den Begriffen ‚Material-Design‘ und ‚Freie Material- und Topologieoptimierung‘ [4, 5, 16–18].

Die ersten Arbeiten, die diesem neuen Forschungsgebiet zugerechnet werden können, wurden in den achtziger Jahren des letzten Jahrhunderts geschrieben. Sie betrafen eher theoretische Hintergrundprobleme der Relaxierung und Homogenisierung [3], [1]. Im Lichte des wachsenden Interesse seitens der Ingenieure wurden numerische Verfahren auf der Basis des Homogenisierungsansatzes entwickelt und in die Praxis umgesetzt [1, 14].

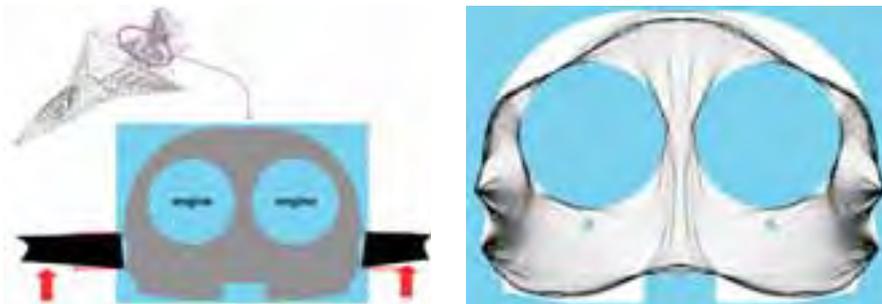
Nach einem vielversprechenden Start gab es zunächst enttäuschende Erkenntnisse, da die Homogenisierungsmethode zu einem stark binnenstrukturierten, lokal unterschiedlichen Schichtenmaterial führte, das wiederum schwer zu realisieren war. Der mathematische Hintergrund dieser artifiziell erscheinenden Mikrostrukturen liegt in dem

0-1-Verhalten des Designzugangs: entweder wird Material in einem Punkt angesetzt oder es wird kein Material zugewiesen. Derartige Funktionen, die nur die Werte 0 oder 1 annehmen, nennt man ‚charakteristische Funktion‘ oder ‚Indikatorfunktionen‘. Da man einen iterativen Optimierungs-Prozess betrachtet, in dem jeweils diese für Anwesenheit oder Abwesenheit von Material ‚zuständigen‘ Funktion angepasst werden, steht man vor dem mathematischen Problem, dass die Grenzwerte solcher Funktionen nicht mehr den 0-1-Charakter haben bzw. im Optimierungsprozess das optimale Material sozusagen ‚verschmiert‘, eben als ‚homogenisiert‘, erscheint. Das Problem, das in mathematischer Begrifflichkeit als ‚G-closure-problem‘ bezeichnet wird, ist bis heute im Rahmen der dreidimensionalen Elastizität ungelöst. Die Homogenisierungsmethode markiert gleichwohl einen Durchbruch im ‚material design‘. Sie brachte das Problem auf den Punkt: ‚Jedes Material ist eine Struktur, wenn man es nur unter einem hinreichend starken Mikroskop betrachtet‘ [5]. Für die Optimierung von Materialien und Werkstoffen ist wieder der ‚umgekehrte Vorgang‘, die ‚inverse Homogenisierung‘, zentral: während in der mathematischen asymptotischen Analysis, die der Homogenisierung zugrunde liegt, aus einer gegebenen Mikrostruktur in einem ‚Referenz-Zellbereich‘ auf das homogenisierte Material geschlossen wird, geht es bei der inversen Homogenisierung darum, die Mikrostruktur in dem Zellbereich aus den gewünschten Materialeigenschaften vermöge mathematischer Optimierung zu berechnen. Entscheidend aber ist, dass immer von realen isotropen Materialien ausgegangen wird, aus denen die Mikrostrukturen konfiguriert werden. Die technische Aufgabe besteht dann darin, die optimalen Mikrostrukturen mechanisch zu realisieren. Für Leichtbaukonstruktionen ist die Methode der Homogenisierung und deren Inversion aufgrund der vorausgesetzten lokalen Isotropie und der als periodisch gedachten zellularen Strukturen nicht immer gut geeignet.

Der entscheidende Schritt, der schließlich zum Erfolg führte, bestand darin, die einschränkenden Modellannahmen fallen zu lassen und neue potentielle Materialien zuzulassen. Dieser Ansatz war der Beginn der ‚Freien Materialoptimierung‘ (FMO). Es wurden neben den notwendigen technischen Einschränkungen (etwa an das Volumen) mit Blick auf das Material nur die wesentlichen physikalischen Nebenbedingungen (Symmetrie und ‚Positivität‘ der Materialmatrizen) berücksichtigt. Auf diese Weise wird ein ‚Utopia-Material‘, eine Art Metamaterial, berechnet, das optimale Eigenschaften besitzt.

Das Ergebnis bedurfte (und bedarf) allerdings der Interpretation durch den Anwender. Dazu ist es unerlässlich, eine sinnvolle Visualisierung der Resultate anzubieten, die es gestattet, aus den Daten ein technisch realisierbares Material sozusagen zu antizipieren. Hier bot sich die Visualisierung durch sogenannte Spannungslinien an, die es gestatten, Zug- und Druckbelastungen im Material abzulesen. Die Ingenieure bei EADS vermochten sodann, diese Linien durch eine Tape-laying-Strategie zu verfolgen. Ein alternativer Zugang bestand in der Betrachtung der Spannungslinien in Kombination mit der so genannten ‚Materialdichte‘, die dann durch das ‚geschulte Expertenauge‘ als topologisches Resultat interpretiert werden konnte. Auf diese Weise konnten die ‚utopischen Daten‘ in ein reales Material übersetzt werden.

Abb. 3: Querspantentwurf. Links Heck, rechts 40 000 Finite Elemente (Quelle: EADS)



Eine ähnliche Technik wurde verwendet, um gewisse Rippenstrukturen im Flügel des Airbus 380 so zu optimieren, dass massiv Gewicht eingespart werden konnte. Das Resultat war sensationell: es konnte über 30 % Gewicht eingespart werden [10].

Die ‚Interpretation‘ durch FMO optimierter Materialien blieb und bleibt ein kritisches Problem. Während das ‚manuelle‘ Verfahren im Falle der Konstruktion der Rippe des Tragflügels erfolgreich war, bleibt das Bedürfnis nach einer mehr automatisierten Interpretation. Insbesondere lässt sich fragen, ob es möglich ist, sozusagen im Schritt eines ‚postprocessing‘, die FMO-Resultate gewissen mechanisch realisierbaren Materialklassen im Zuge von Optimierungsverfahren zuzuordnen. Das daraus entstehende mathematische Problem ist ein nichtlineares Mixed-Integer-Problem (MINLP), in dem sowohl diskrete als auch kontinuierliche Variablen in die Optimierung eingehen. Es kann an eine Verschränkung des FMO-Problems mit dem MINLP-Problem gedacht werden, das eine Alternanz zwischen beiden Zugängen realisiert. Eine Parallelisierung der Optimierungsprobleme ist unter Umständen notwendig [11].

Der Erfolg wurde zum Programm: schnell fanden sich mehrere Forschergruppen aus Industrie und Academia im europäischen Raum in einem von der EU geförderten Forschungsprogramm (PLATO-N: www.plato-n.org) zusammen, welches das Potenzial der Material- und Topologieoptimierung von der Grundlagenforschung zur Industriereife ausschöpft.

3.1 FAZIT/EMPFEHLUNGEN

Form-, Topologie- und Materialoptimierung stellen eine große Herausforderung an die industriennahe mathematische Optimierung dar. Zu einer erfolgreichen Forschung müssen diskrete und kontinuierliche Optimierungsmethoden verknüpft und neue Paradigmen ihrer Verzahnung entwickelt werden. Der Grundstein ist gelegt. Es ist möglich, dass

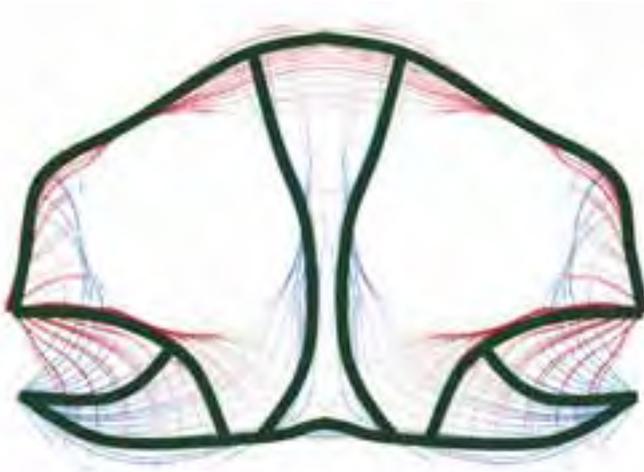
Abb. 4: Tape-laying Technik (Quelle: EADS)



Spannungslinien



Tape

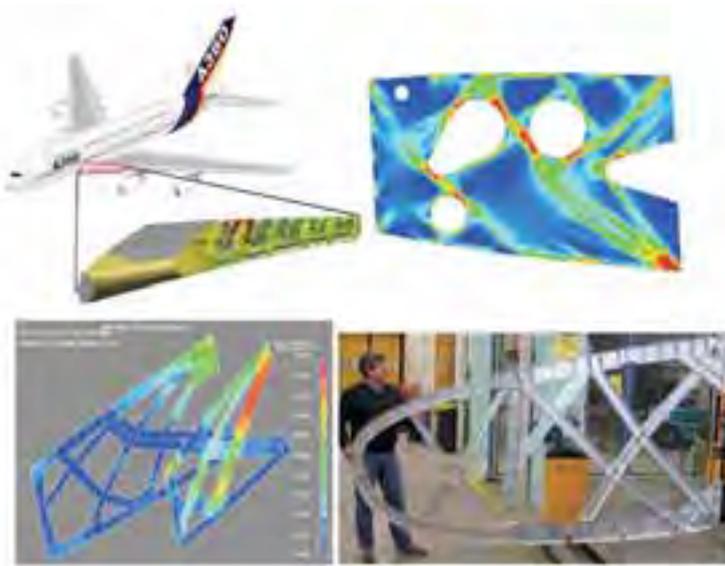


vor dem Hintergrund der hervorragenden Nachwuchssituation in Deutschland durch geeignete Förderungsmaßnahmen eine international herausragende Rolle deutscher Wissenschaftler erzielt werden kann.

4 OPTIMIERUNG VON TRANSPORTNETZEN

Netzwerke aus ‚pipelines‘ dienen dem Transport von Gas von den Anbietern zum Verbraucher. Im Gegensatz zu einer vergleichsweise einfachen Konstellation, in der nur wenige Anbieter Gas bestimmter Qualität in das Netz einspeisten, hat sich die Situation in den letzten zehn Jahren dramatisch dahingehend geändert, dass nunmehr viele auch internationale Anbieter Gas unterschiedlichster Qualität in das Gasnetz einleiten und die Verbraucher einen zunehmend kritischen Blick auf Preise und Qualität richten. Hinzu kommen durch die Liberalisierung der Märkte Fragen nach den variablen Kosten eines

Abb. 5: Praktische Umsetzung: Die Rippe (Quelle: EADS)



Gasnetzes oder allein nach den Durchleitungskosten von Gas durch ein bestehendes Netz. Den Gasfluss in einem Netz optimal zu steuern, so dass den Kundenprofilen störungssicher Rechnung getragen wird und gleichzeitig die Betriebskosten minimal gehalten werden, ist ein Problem von großer Tragweite. Die Betriebskosten sind geprägt durch viele Faktoren. Ein entscheidender Faktor ist der Gasverbrauch der Steuerungsinstrumente, insbesondere der Kompressoren. Kompressoren sind zum Erhalt des vorgeschriebenen Gasdrucks im Leitungssystem notwendig. Aufgrund mannigfacher Reibungsphänomene verliert das Gas nämlich während des Transportes massiv an Druck. Während das Netz tausende von Gasrohren signifikanter Größe beinhaltet, stehen nur sehr wenige dieser Kompressoren zur Verfügung. Einen Eindruck von der Größe und der Komplexität dieser Kompressoren vermittelt Abb. 8.

Das Gasnetzwerk mit der Gesamtheit der Rohre und Bauwerke sowie seiner Entscheidungsstrukturen entzieht sich angesichts seiner Komplexität einer exakten Beschreibung. Dies führte traditionell dazu, das System durch sehr einfache algebraische Gleichungen zu modellieren, die lediglich die Beziehungen der Druck- und Flussverhältnisse an den jeweiligen Enden der Rohre zum Ausdruck bringen. Auf diese Weise entsteht ein diskretes Flussmodell auf einem Graphen, dessen Knoten die Endpunkte der Rohre und dessen Kanten die Rohre selbst durch die Flussrate repräsentieren. Die Ventile werden

Abb. 6: Praktische Umsetzung: Gewichtsreduktion (Quelle: EADS)

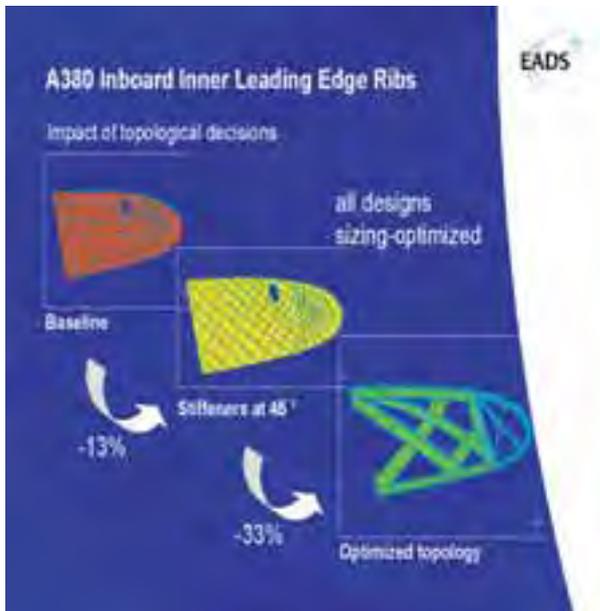


Abb. 7: Praktische Umsetzung Airbus 380: Links Prototyp, rechts Produktvorstellung (Quelle: EADS)



Abb. 8: Kompressoranlage (Quelle: E.ON Ruhrgas)



als ‚offen‘ (1) oder ‚geschlossen‘ (0) dargestellt und die Kompressoren werden als ‚in Betrieb‘ (1) oder ‚abgeschaltet‘ (0) abgebildet. Ist ein Kompressor in Betrieb (1), so verbraucht er Gas, das dem Netz entnommen wird. Dieser Verbrauch, der eine hochgradig nichtlineare Funktion der Druckverhältnisse vor und nach dem Kompressor und anderer Parameter ist, ist vermöge zulässiger Operationsbedingungen so zu minimieren, dass die Kundennachfrage befriedigt wird. Das gesamte Problem lässt sich als ein so genanntes ‚Gemischt ganzzahliges Optimierungsproblem (englisch ‚Mixed Integer Program‘ oder kurz MIP)‘ mathematisch approximieren. Im Falle einer zu groben Approximation muss auf feinere physikalische Modelle, z. B. partielle Differentialgleichungen, im Zuge einer Simulation zurückgegriffen werden. Testrechnungen haben gezeigt, dass mit dieser Vorgehensweise globale Optimallösungen innerhalb einer vorgegebenen Genauigkeit bestimmt werden können. Vergleiche auf bereits sehr einfach strukturierten Netzwerken dokumentieren, dass damit signifikante Verbesserungen gegenüber einer ausschließlichen Verwendung von nichtlinearen oder simulationsbasierten Optimierungsverfahren

erzielt werden können. Zukunftsweisend sind hier diskret-kontinuierliche hybride Modelle, die dem Entscheidungscharakter durch Schalten zwischen verschiedenen Modellen Rechnung tragen [9].

Die Komplexität der geschilderten Optimierungsaufgaben war für die damalige Ruhrgas AG (heute E.ON Ruhrgas AG) Anlass, Kontakt mit Optimierern am Konrad-Zuse-Zentrum Berlin aufzunehmen, woraus sich ein BMBF-Verbundprojekt auch mit anderen Forschergruppen entwickelte, das die numerische Lösung der Probleme zum Ziel hatte [12, 13].

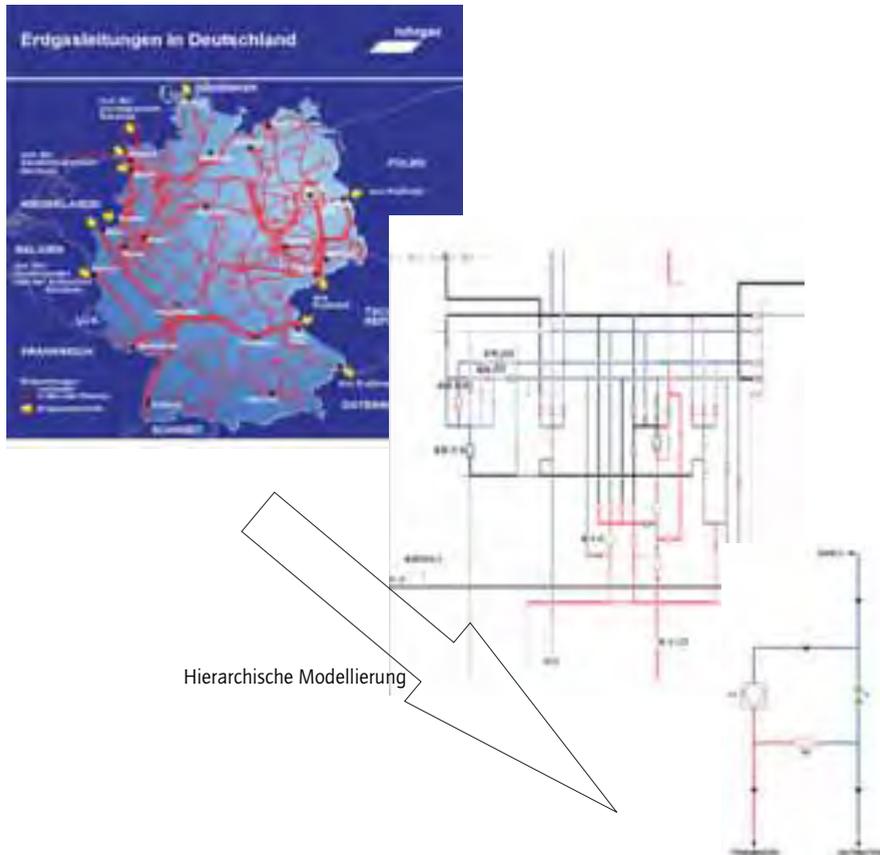
Sehr ähnliche Fragestellungen ergeben sich bei Frischwasser- und Abwassernetzwerken. In diesem Kontext wird eine Kooperation verschiedener Arbeitsgruppen an der TU Darmstadt und der Friedrich-Alexander-Universität Erlangen-Nürnberg mit SIEMENS, Hessenwasser, Steinhardt und den Dresdener Stadtwerken seit 2007 durch das BMBF-Programm gefördert (www.odysseus.tu-darmstadt.de) [8] (siehe auch die Referenzen darin).

4.1 FAZIT/EMPFEHLUNGEN

Die Optimierung, Instandhaltung und Planung von Infrastrukturen im Bereich dynamischer Transportnetzwerke sind Problemstellungen von höchster gesellschaftlicher Relevanz. Milliarden von Euro werden in naher Zukunft in diesen Bereich zu investieren sein. Die Verteilung von Ressourcen (Wasser, Gas, Öl etc.) ist nachgerade *das* Zukunftsproblem der Menschheit. Viele der der Verteilung zugrundeliegenden Prozesse sind entscheidungs- und regelbasiert und oft einer Automatisierung nur schwer zugänglich. Ein großer Teil dieser Prozesse lässt sich aber durch mathematische Objekte wie algebraische Gleichungen, gewöhnliche oder partielle Differentialgleichungen und Hierarchien solcher Modelle beschreiben. Entscheidungsvorgänge lassen sich durch 0-1-Strukturen mathematisch abbilden. Im Rahmen einer Multiskalenmodellierung lassen sich entscheidungs-orientierte diskrete-kontinuierliche Optimierungsmethoden mit technik-orientierten kontinuierlichen Optimierungsmethoden verzahnen. Dynamischen diskret-kontinuierlichen Optimierungsverfahren für Systeme mit Millionen Variablen kommt eine Schlüsselrolle in der künftigen Entwicklung zu. Anfänge sind gemacht: einige BMBF-Projekte konzentrieren sich auf diese und verwandte Fragestellungen. Das Potential an jungen Nachwuchswissenschaftlern steht bereit. Es bedarf kooperierter Forschungsinitiativen seitens des Bundes, dieses Potential auszuschöpfen.

Wir haben gezeigt: Reale Prozesse im Kontext von Schlüsseltechnologien sind nur mit Methoden einer Multiskalenmodellierung abbildbar. Fragen der Optimierung entstehen auf natürliche Weise in der Umkehrung der phänomenologischen ‚Vorwärtsmodellierung‘, nämlich von den Eigenschaften zur Struktur. Variationen von Strukturen oder Strukturelementen lassen sich in den Bereich der Form-, Struktur- oder Topologieoptimierung sowie in den Rahmen einer diskret-kontinuierlichen Optimierung mathematisch abbilden. Aufgrund der komplexen Anwendungssituation, die allein den Einsatz mathe-

Abb. 9: Netzwerkhierarchie



matischer Optimierung rechtfertigt, hat man es mit Millionen von Variablen und Parametern zu tun. Diese Aspekte zusammenzutragen und sie in ein Konzept zu gießen, ist eine herausragende Aufgabe, die der Förderung durch koordinierte Förderprogramme dringend bedarf. Ein solches sehr erfolgreiches Förderprogramm ist der DFG-Schwerpunkt 1253 ‚Optimization with PDE constraints‘ (Koordinator: Prof. G. Leugering). In diesem Programm werden über 25 Projekte gefördert, die dem beschriebenen Forschungsumfeld zugeordnet werden können. Durch eine gezielte Berufungspolitik ist es gelungen, in den vergangenen Jahren Deutschland zur führenden Nation in der Optimierung mit partiellen Differentialgleichungen zu machen. Jetzt besteht die einmalige Chance, diesen

Vorsprung nachhaltig zu machen und die weltweite Führung auf das Gebiet einer ‚Hybriden diskret-kontinuierlichen Multiskalen-Optimierung großer Systeme‘ auszuweiten.

5 DANKSAGUNG

Die Bilder von der Fertigung der Strukturelemente wurden uns freundlicherweise von der Firma EADS Deutschland GmbH, und die Bilder der Gasnetze und Kompressoren von der Firma Ruhrgas AG (heute E.ON Ruhrgas) im Rahmen eines BMBF-Projekts (2000–2003) zur Verfügung gestellt.

6 LITERATUR

- [1] G. Allaire, Shape optimization by the homogenization method Applied Mathematical Sciences 146, Springer Verlag 2002
- [2] G. Allaire F. Gournay, F. Jouve und A.-M. Toader, Structural optimization using topological and shape sensitivities via a level set method, Ecole Polytechnique, R.I. Nr. 555, 2004.
- [3] G. Allaire und R. Kohn. Optimal design for minimum weight and compliance in plane stress using extremal microstructures. European J. on Mechanics (A/Solids), 12:839–878, 1993.
- [4] A. Ben-Tal, M. Kocvara, A. Nemirovski, und J. Zowe. Free material design via semi-definite programming. The multi-load case with contact conditions. SIAM J. Optimization, 9(4): 813–832, 1999.
- [5] M. P. Bendsøe und O. Sigmund. Toplogy Optimization. Theory, Methods and Applications, Springer-Verlag, Heidelberg, 2004.
- [6] M. Bernot, V. Caselles und J.-M. Morel, Branched transportation networks, Springer-Verlag 2007/2008
- [7] Durand, M., Architecture of optimal transport networks, Physical Review E (2006) 73, 016116.
- [8] M. Gugat und G. Leugering, Global boundary controllability of the de St. Venant system for sloped channels with friction. Annales de l'Institut Henri Poincaré Non Linear Analysis, Vol. 25, 2008, in press, published online.

- [9] F. Hante, G. Leugering und T.I. Seidman, Modeling and Analysis of Modal Switching in Networked Transport Systems, Appl. Math. and Optim. (in press 2008)
- [10] M. Kocvara, M. Stingl und R. Werner. MOPED User's Guide. Version 1.02. Research Report 262, Institute of Applied Mathematics, University of Erlangen, 2000.
- [11] J. E. Lagnese und G. Leugering, Domain decomposition methods in optimal control of partial differential equations., ISNM. International Series of Numerical Mathematics 148. Basel: Birkhäuser. xiii, 443 p.,2004.
- [12] A. Martin, M. Möller und S. Moritz, Mixed Integer Models for the Stationary Case of Gas Network Optimization, Mathematical Programming B, 2006, 105, 563-582.
- [13] D. Mahlke, A. Martin und S. Moritz, A simulated annealing algorithm for transient optimization in gas networks, Mathematical Methods of Operations Research, 2007, 66, 99-116.
- [14] G. I. N. Rozvany, Topology optimization of multi-purpose structures, Math. Methods Oper. Res., (1998), 47/2, 265-287.
- [15] J. Sokolowski und A. Zochowski, Topological derivatives for elliptic problems, Inverse problems (1999), 15, 123-134.
- [16] M. Stingl, M. Kocvara und G. Leugering, Free Material Optimization with Control of the Fundamental Eigenfrequency, FAU Institute of Applied Mathematics, Preprint 319 (in review SIAM J Optim.)
- [17] M. Stingl, M. Kocvara und G. Leugering, A Sequential Convex Semidefinite Programming Algorithm for Multiple-Load Free Material Optimization, FAU Institute of Applied Mathematics, Preprint 317 (in review SIAM J Optim.)
- [18] J. Zowe, M. Kocvara und M. Bendsøe. Free material optimization via mathematical programming. Mathematical Programming, Series B, 79: 445-466, 1997.

ENERGIE UND BAUEN

> RESSOURCENEINSATZPLANUNG IN DER ENERGIE- UND VERFAHRENSTECHNIK

SEBASTIAN ENGELL, EDMUND HANDSCHIN, CHRISTIAN REHTANZ
UND RÜDIGER SCHULTZ

1 EXECUTIVE SUMMARY

Der ökologisch nachhaltige, wirtschaftlich effiziente und politisch faire Umgang mit den Ressourcen der Erde ist eine der größten Herausforderungen an die Menschheit im 21. Jahrhundert und impliziert Paradigmenwechsel in allen Bereichen des gesellschaftlichen Lebens. Energie- und Verfahrenstechnik sind traditionell Bereiche, in denen der optimale Umgang mit kostbaren und beschränkt verfügbaren Ressourcen von erstrangiger Bedeutung ist. Zur Lösung der damit verbundenen komplexen Entscheidungsprobleme sind mathematische Methoden schon seit längerem im Einsatz. Hierbei können aber noch längst nicht alle bislang aufgetretenen Probleme fundiert angegangen werden. Der wachsende wirtschaftliche und ökologische Druck auf die Effizienz der Produktions- und Versorgungssysteme sowie stärkere Schwankungen von Angebot und Nachfrage führen zu neuen Herausforderungen, für die neue mathematische Methoden und Algorithmen entwickelt werden müssen.

Dieser Artikel berichtet über den Stand und die Bedeutung des Einsatzes von Mathematik bei der Ressourceneinsatzplanung in der Energie- und Verfahrenstechnik, geht auf neue Herausforderungen ein und unterbreitet Vorschläge für innovative Forschungsaktivitäten zu deren Bewältigung.

Zu den Beispielen für den erfolgreichen Einsatz von Mathematik in der Energie- und Verfahrenstechnik zählen die Kraftwerksoptimierung zum effizienten Ressourceneinsatz bei der elektrischen Energieversorgung, die Regelung elektrischer Netze und die kontinentweite Gewährleistung von deren Stabilität, die optimierte Durchlaufplanung von Materialchargen durch Produktionsprozesse wie u.a. die Herstellung von Kupfer oder die optimale Produktionsplanung bei der Herstellung pharmazeutischer Wirkstoffe.

Allen diesen Beispielen ist gemein, dass die mathematische Modellierung und Optimierung bisher unter der Annahme erfolgt, dass die relevanten Problemdaten vollständig bekannt sind. Tatsächlich handelt es sich aber um Entscheidungsprobleme in dynamischen Kontexten, in denen die künftige Entwicklung mit Unsicherheiten behaftet ist, aber in die Entscheidung einbezogen werden muss. Die mathematische Modellierung

und Optimierung müssen unvollständige Information und Informationszuwachs im Laufe der Optimierung adäquat berücksichtigen.

Durch den Wechsel hin zu liberalisierten Energiemärkten wird beispielsweise das elektrische Netz von einem rein technischen Gebilde zu einer Handelsplattform mit bidirektionalem Informationsaustausch zwischen Produzenten, Händlern und Kunden. Treten noch exogene Einflüsse wie z. B. zufallsbehaftete Windenergieeinspeisung hinzu, so ergeben sich für die Ressourcenplanung hochkomplexe Entscheidungsprobleme unter unvollständiger Information.

Ganz analoge Effekte zeigen sich bei der Optimierung verfahrenstechnischer Produktionsprozesse. Auch hier nimmt der sachgerechte Umgang mit unvollständiger Information eine Schlüsselstellung bei der Ressourceneinsatzplanung ein. Solche Unsicherheiten können beispielsweise bestehen in Bezug auf Prozesszeiten, Rohstoffpreise, Kundennachfrage oder Anlagenverfügbarkeit. Angesichts der Komplexität der Entscheidungsprobleme kommt neben der Einbeziehung von Unsicherheit in die Entscheidungsfindung der für Anwender nachvollziehbaren Kommunikation von Lösungen mittels innovativer Instrumente der Mensch-Maschine-Kommunikation große Bedeutung zu.

In enger Wechselwirkung mit anderen Teilgebieten der mathematischen Optimierung zielt die stochastische Optimierung auf die Entwicklung von Modellen und Algorithmen zur Entscheidungsfindung bei Ungewissheit. Erste Anwendungen der stochastischen Optimierung befinden sich im Einsatz. Die Bewältigung der oben skizzierten Herausforderungen erfordert allerdings noch intensive grundlagenbezogene und anwendungsorientierte interdisziplinäre Forschung von Mathematikern, Ingenieuren und Informatikern. Die Autoren schlagen dazu vier exemplarische Themenbereiche vor:

- Optimierungsstrategien für den Betrieb, den Ausbau und die Erneuerung der elektrischen Versorgungsnetze des 21. Jahrhunderts,
- Entscheidungshilfen zur Optimierung des bidirektionalen Informations- und Energiemanagements in innovativen Strukturen der Energieversorgung,
- Problemangepasste Modellierung und effiziente Handhabung von Ungewissheit bei Ressourcenzuteilungsproblemen in der verfahrenstechnischen Produktion,
- Mensch-Maschine-Interaktion bei der Optimierung komplexer Produktionsprozesse.

Aufbauend auf bereits bestehenden und noch zu schaffenden interdisziplinären Vernetzungen können durch das BMBF, das BMWi oder die DFG zukunftsweisende Forschungsprojekte in diesen Bereichen gefördert werden.

2 ERFOLGSGESCHICHTEN

2.1 ELEKTRISCHE ENERGIEVERSORGUNG

Voraussetzungen für einen sicheren, wirtschaftlichen und umweltverträglichen Betrieb der elektrischen Energieversorgung sind die ständige Einhaltung aller technischer Grenzwerte sowie des (n-1)-Prinzips¹ für die Führung und Optimierung der elektrischen Transport- und Verteilnetze. Dazu sind umfangreiche Netzsimulationen erforderlich, um den aktuellen Zustand der Netze zuverlässig und sicher bestimmen zu können. Diese Simulationen beruhen auf der Lösung sehr großer, nichtlinearer Gleichungssysteme, so dass innerhalb weniger Sekunden auch für Netze mit sehr vielen Knoten und Leitungen die gewünschten Lösungen zusammen mit einer graphischen Anzeige zur Verfügung stehen.

NETZREGELUNG UND RESERVELEISTUNG

Das Hoch- und Höchstspannungsnetz (380/220 kV) ist in Regelzonen eingeteilt, in der zu jedem Zeitpunkt das Leistungsgleichgewicht zwischen Erzeugung und Verbrauch exakt eingehalten werden muss. Dazu muss der Übertragungsnetzbetreiber (ÜNB) ausreichend Regelenergie einkaufen und bereitstellen, um so die Frequenz von 50 Hz stets einhalten zu können. Dabei wird zwischen Primär-, Sekundär- und Minutenreserve unterschieden. Die erste wird dezentral im Kraftwerk bei Bedarf automatisch aktiviert. Sie muss sofort bereitgestellt werden und bis zu 15 Minuten zur Verfügung stehen. Die Sekundärreserve wird zentral in der Netzleitstelle aktiviert und kann von Kraftwerken innerhalb oder außerhalb der eigenen Regelzone geliefert werden. Sie muss spätestens 30 Sekunden nach Auftreten eines Leistungsungleichgewichts aktivierbar sein und für mehr als 15 Minuten zur Verfügung stehen. Durch den telefonischen Abruf der Minutenreserve kann Primär- und Sekundärreserve wieder auf den ursprünglichen Wert zurückgefahren werden. Im Hinblick auf die Größe des europäischen Verbundnetzes mit einer installierten Leistung von 600 GW handelt es sich hier mit Sicherheit um die komplexeste regelungstechnische Aufgabe eines von Menschen geschaffenen technischen Systems. Abb. 1 zeigt die drei genannten Regelreserven in ihrer zeitlichen Anforderung.

Im Bereich der Mittel- und Niederspannungsnetze ist der Verteilnetzbetreiber (VNB) ebenso wie der ÜNB im Bereich der Hoch- und Höchstspannungsnetze für die Bereitstellung ausreichender Transportkapazität für die von den Kunden angeforderte elektrische Energie verantwortlich. Durch die Liberalisierung der Energiemärkte und die starke Förderung der erneuerbaren Energien, die insbesondere im Bereich der Windenergie durch stark schwankende Leistungseinspeisungen gekennzeichnet sind, sind die Anforderungen an den Netzbetrieb erheblich gestiegen. Durch die Entflechtung von Erzeugung,

¹ D. h. die Funktionsfähigkeit des Netzes auch nach eventuellem Ausfall einer beliebigen Komponente.

Handel und Anwendung einerseits und den Netzbetrieb als natürliches Monopol andererseits, sind neue Aufgaben entstanden, die nur durch die enge Zusammenarbeit zwischen Bilanzkreisverantwortlichen und ÜNB resp. VNB zu bewältigen sind.

Da der physikalische und der vertragliche Stromhandel völlig voneinander entkoppelt sind, ist der Netzbetreiber gezwungen, die geforderte Transportleistung ständig auf ihre Realisierbarkeit hin zu überprüfen. Dazu stehen ihm umfangreiche Methoden der mathematischen Optimierung zur Verfügung:

- Zustanderkennung, die auf der Ausgleichsrechnung beruht,
- Optimale Netzberechnung, die die numerische Lösung eines sehr großen, nicht-linearen Gleichungssystems erfordert, um die Netzverluste zu minimieren,
- Netzsicherheitsberechnungen zur Überprüfung des (n-1)-Prinzips und der Kurzschlußfestigkeit aller Netzkomponenten,
- Feststellung und Lösung von Netzengpassproblemen.

Durch den zunehmenden europäischen Handel mit elektrischer Energie hat die lokale Netzengpassbildung im europäischen Verbundnetz stark zugenommen. In der Praxis hat sich bisher das Auktionsverfahren durchgesetzt, das aber insbesondere in stark vermaschten Netzen mit erheblichen Nachteilen verbunden ist. Erste Erfolge sind mit dem Einsatz der hierarchischen Optimierung erzielt worden, bei dem berücksichtigt wird, dass die Netzdaten zwischen den einzelnen ÜNB nicht vollständig ausgetauscht werden müssen [3].

Abb. 1: Zeitliche Bereitstellung der Regelreserve in Transportnetzen

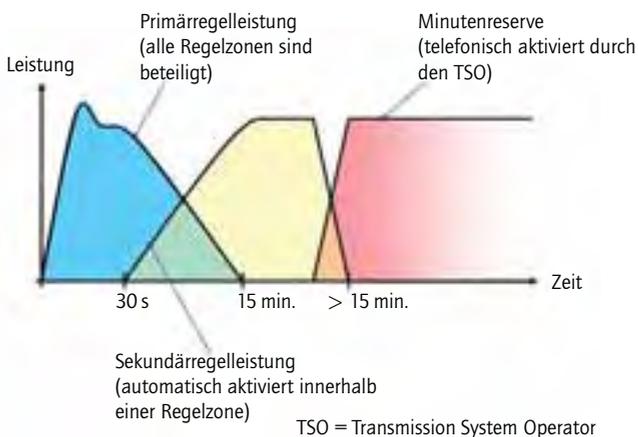


Abb. 2: Netzleitstelle eines großen Übertragungsnetzbetreibers (Quelle: RWE)



Abb. 2 zeigt die Ansicht einer Netzleitstelle für ein großes Übertragungsnetz. Von hier aus wird das gesamte Transportnetz überwacht, geführt und optimiert.

KONTINENTWEITE STABILITÄT IN ECHTZEIT

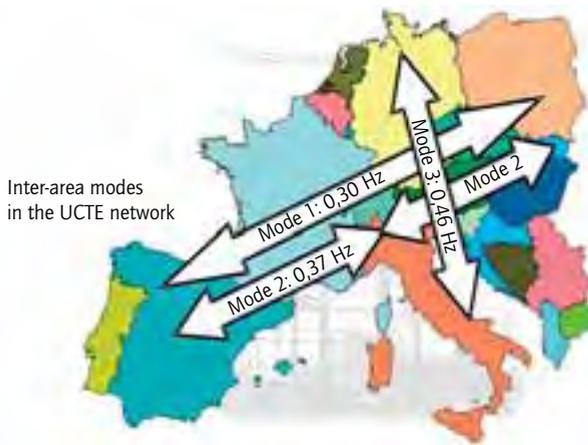
Durch die Einspeisung stark fluktuierender Windenergieanlagen muss die Stabilität des gesamten UCTE-Übertragungsnetzes fortlaufend geprüft werden. Dabei zeigt sich, dass sowohl in West-Ost-Richtung wie auch in der Nord-Süd-Richtung Eigenfrequenzen auftreten, die zu unerwünschten Leistungspendelungen führen können. Abb. 3 zeigt die Ergebnisse der Eigenwertanalyse für das UCTE-Netz [33].

Derartige Eigenfrequenzen müssen sorgfältig überwacht werden. Dazu werden heute transkontinentale Weitbereichsmessungen und zugehörige Auswertemethoden eingesetzt, die durch eine ausführliche Signalanalyse der gemessenen Größen zuverlässige Aussagen über Art und Ort der Einsatzorte für dämpfende Betriebsmittel aus dem Bereich der leistungselektronischen Netzregler (sogenannte FACTS-Geräte) erlauben [30].

KRAFTWERKSOPTIMIERUNG ZUM EFFIZIENTEN RESSOURCENEINSATZ

Im Bereich der Kraftwerksoptimierung sind sowohl das Einsatzproblem wie auch die wirtschaftliche Lastaufteilung zu lösen. Bei der Einsatzplanung ist kurz-, mittel- und langfristig festzulegen, welche Erzeugungseinheiten wann zur Lastdeckung eingesetzt werden sollen. Neben der dynamischen Programmierung spielt die gemischt-ganzzahlige lineare Optimierung eine wichtige Rolle. Bei der mathematischen Modellierung ist zu beachten, dass neben Leistungen auch Energieebenenbedingungen zu erfüllen sind, die als

Abb. 3: Eigenfrequenzen im europäischen UCTE-Netz



integrale Größen der Leistung in die Optimierung einzubauen sind. Weiterhin sind die Mindestbetriebs- und Mindestausschaltzeiten zu beachten. Hier stehen Optimierungsverfahren zur Verfügung, mit denen auch sehr große Probleme erfolgreich gelöst werden können [7, 11].

Bei der wirtschaftlichen Lastaufteilung ist zu bestimmen, mit welcher Leistung welche Kraftwerksblöcke eingesetzt werden. Dabei sind die Last- und Netznebenbedingungen zu jedem Zeitpunkt exakt einzuhalten. Die Unsicherheiten der Preisentwicklung wie auch der Einspeisung regenerativer Erzeuger werden in absehbarer Zeit anstelle der etablierten deterministischen Optimierungsverfahren den Einsatz der stochastischen Optimierung erforderlich machen. Dadurch kann die Unsicherheit, die mit Voraussagen verknüpft ist, innerhalb realistischer Grenzen berücksichtigt werden [39].

2.2 RESSOURCENEINSATZPLANUNG IN DER VERFAHRENSTECHNISCHEN PRODUKTION

Verfahrenstechnische Produktionsanlagen unterscheiden sich von Maschinen der Fertigungstechnik durch längere Dauern der einzelnen Schritte, komplexere Übergangsvorgänge und -bedingungen zwischen unterschiedlichen Arbeitspunkten oder Chargen und komplexe stoffliche Verkopplungen. Für den wirtschaftlichen und energieeffizienten Betrieb der Produktionsanlagen und die Reduktion der von der Lagerung gefährlicher und giftiger Stoffe ausgehenden Risiken ist die günstige zeitliche Anordnung und Abstim-

mung der einzelnen Schritte von wesentlicher Bedeutung. Die effiziente Nutzung der Produktionskapazitäten ist in verfahrenstechnischen Produktionsprozessen in den letzten Jahren immer wichtiger geworden. Dies dient zum einen der Einsparung von Energie und Rohstoffen und der Vermeidung von Abfällen und nicht spezifikationsgerechter Produktion, die teuer entsorgt werden müssen, zum anderen der Steigerung der Wirtschaftlichkeit.

Produktionsplanungsprobleme treten vor allem dort auf, wo die Produktion in einzelnen Chargen erfolgt und wo eine Vielzahl unterschiedlicher Produkte parallel und nacheinander hergestellt werden muss um spezifische Bedarfe des Marktes zu decken. Nachfolgend wird an zwei Beispielen auf den Einsatz von Optimierungsverfahren in diesem Bereich eingegangen.

SIMULTANE OPTIMIERUNG UND REIHENFOLGEPLANUNG IN DER KUPFERPRODUKTION

Kupfer ist das am längsten von der Menschheit genutzte Metall und auch heute noch von großer Bedeutung in einer Vielzahl von Produkten, beispielsweise in der Elektronik und in der Bautechnik. Ein großer Vorteil von Kupfer ist die Möglichkeit eines problemlosen Recyclings. Kupfer kann leicht von Verunreinigungen befreit werden und hat dann dieselben Eigenschaften wie aus Erz gewonnenes „frisches“ Kupfer. Ein erheblicher Teil des Bedarfs kann daher durch Aufschmelzen und Reinigen von Schrott und Abfällen aus der Produktion gedeckt werden.

Bei der Herstellung von Kupfer (Abb. 4) durchläuft das Material mehrere Öfen und Konverter, in denen alle störenden Bestandteile entfernt werden, bis das Kupfer schließlich mit einer Reinheit von 99,6 % vorliegt und in Stranggussanlagen vergossen wird. In einer typischen Kupferraffinerie sind mehrere parallele Öfen und Konverter vorhanden, zwischen denen das Material mit großen Kränen bewegt wird. Diese Kräne teilen sich einen Arbeitsbereich, die Bewegungen eines Kranes werden deshalb durch die des anderen eingeschränkt. Die Produktion in den Konvertern und Anodenöfen muss so arrangiert werden, dass die gemeinsam genutzte Ressource „Kräne“ zeitlich koordiniert beansprucht wird, und so dass die nachfolgende Stranggussanlage kontinuierlich mit Material versorgt wird. Gleichzeitig sind Wartezeiten möglichst zu vermeiden, da hierdurch Energieverluste eintreten und die Anlagen schlechter ausgelastet werden. Da Schlacken mit einem erheblichen Kupfergehalt im Prozess zu den davorliegenden Stufen zurückgeführt werden, muss eine genaue Materialbilanz aufgestellt und berücksichtigt werden.

Das Problem der Planung der Durchläufe der einzelnen Chargen von Material durch den Produktionsprozess wird durch zwei Aspekte stark erschwert: die Prozesszeiten und erreichten Reinheiten in den einzelnen Apparaten hängen stark von der Zusammensetzung des Eingangsmaterials ab und es müssen unregelmäßige geplante Wartungen von Anlagen berücksichtigt werden. Neben geplanten Wartungsarbeiten kommen im

Betrieb der Anlagen natürlich unvorhergesehene Ausfälle vor, auf die in der Ressourcenbelegungsplanung reagiert werden muss.

In einer von ABB entwickelten Lösung für die simultane Rezeptoptimierung und Produktionssteuerung für die Kupferproduktion wird das Gesamtproblem als gemischt-ganzzahliges lineares Programm (MILP) formuliert [14]. Das Optimierungsproblem enthält knapp 1000 kontinuierliche Variable, beispielsweise die Startzeiten von Operationen auf den einzelnen Einheiten und knapp 100 binäre Variable für die Belegung der Ressourcen. Durch eine problemangepasste Modellierung gelang es, das integrierte Optimierungsproblem für einen Zeitraum von 1–1½ Tagen mit dem kommerziell verfügbaren MILP-Löser CPLEX [19] in wenigen Sekunden zu lösen. Dadurch ist eine fast verzögerungsfreie Reaktion auf Störungen oder Änderungen der Materialzusammensetzung möglich. Die Kranbewegungen, die zur Ausführung der geplanten Produktionsschritte notwendig sind, werden nach Festlegung der Produktionsreihenfolgen unterlagert simuliert, um sicher zu stellen, dass der berechnete Produktionsplan tatsächlich ausführbar ist. Wenn sich herausstellt, dass dies nicht der Fall ist, müssen die im Planungsmodell abgeschätzten Zeiten für die Krantransporte angepasst und die Planung wiederholt werden. Die Lösung hat sich im industriellen Einsatz bewährt. In dem realisierten System wird der aktuelle Planungsstand in allen Anlagenteilen angezeigt um allen Verantwortlichen nicht nur die aktuellen Vorgaben für die nächsten Schritte anzuzeigen sondern auch einen Überblick über den Stand der Planung und die Einbindung ihrer Einheiten in den Gesamtprozess zu geben [13].

Abb. 4: Kupferproduktion



PRODUKTIONSPLANUNG BEI DER HERSTELLUNG PHARMAZEUTISCHER WIRKSTOFFE

Produktionsplanungs- und Produktionssteuerungsprobleme in der chemischen und pharmazeutischen Industrie unterscheiden sich von den klassischen „Flow-Shop“ und „Job-Shop“ Problemen der Stückgutproduktion durch eine Reihe von zusätzlichen Beschränkungen und durch sehr lange Planungshorizonte. Ein typisches Planungsproblem betrifft 10 000 verschiedene Substanzen (Rohstoffe, Zwischen- und Endprodukte), die nach spezifischen Rezepten in bis zu 30 oder mehr Schritten hergestellt werden, einige Hundert Apparate oder Verpackungsmaschinen, zahlreiche Standorte und einen Planungshorizont von mehreren Jahren. Typisch für verfahrenstechnische Planungsprobleme ist ein „divergierender Baum“ von Produkten – aus wenigen Rohstoffen wird eine große Zahl verschiedener Zwischen- und Endprodukte hergestellt. Dabei fallen vielfach Nebenprodukte oder unverbrauchte Einsatzstoffe an, die zurückgeführt werden, ggf. nach einer Reinigung, oder an anderer Stelle im Produktionsprozess verwendet werden. Typischerweise verursacht der Wechsel von einem (Zwischen-)Produkt zum anderen auf einem Apparat beträchtlichen Aufwand für die Reinigung und ggf. Umrüstung, so dass man eine sog. Kampagnenfahrweise bevorzugt, in der dasselbe Produkt möglichst lange auf einer Anlage hergestellt wird. Dies erzeugt andererseits aber Lagermengen und damit Kosten, und sowohl die Menge als auch die Dauer der Zwischenlagerung können beschränkt sein. Zusätzlich zur Einsatzplanung der Anlagen ist die Verfügbarkeit von Personal ein erheblicher beschränkender Faktor, insbesondere bei Umrüstoperationen.

Derartige Planungsprobleme sind heute und auf absehbare Zeit nicht „im Ganzen“ mit mathematischen Optimierungsalgorithmen lösbar. Es existieren jedoch kommerzielle Werkzeuge, die näherungsweise optimale Lösungen liefern. Diese basieren auf zwei Grundmethoden der Vereinfachung der Lösung: Hierarchische Zerlegung und näherungsweise Lösung mit Heuristiken und lokaler Optimierung. Bei der hierarchischen Zerlegung wird typischerweise über den gesamten Planungshorizont ein vereinfachtes Problem gelöst, in dem nur die wichtigeren Beschränkungen und die Hauptprodukte und Haupteinsatzstoffe berücksichtigt werden. Ebenso werden Ressourcen, die bekanntermaßen nie Engpässe darstellen, nur in vereinfachter Form und vergleichsweise kurze Nebenzeiten pauschal berücksichtigt. Bei der Optimierung werden feste Zeitabschnitte von beispielsweise je einem Monat verplant, die genaue zeitliche Festlegung innerhalb dieser Slots wird einem unterlagerten Scheduling-Algorithmus überlassen. Die Optimierungsprobleme können dann als gemischt ganzzahlig-lineare Programme formuliert und gelöst werden. Planungsaufgaben der oben genannten Größenordnung führen auf Optimierungsprobleme mit einigen 100 000 kontinuierlichen und ca. 1000 binären Variablen, die innerhalb einer Stunde gelöst werden können [8]. Da die Lösungen des vergrößerten Problems nicht direkt implementierbar sind, wird die Produktion über den unmittelbar in der Zukunft liegenden Zeithorizont, beispielsweise für drei Monate detailliert geplant, wobei die Ergebnisse der langfristigen Optimierung als Aufträge, die zu erfüllen sind, angenommen werden. Die beiden Ebenen benutzen unterschied-

lich genaue Modelle, weshalb ihre Abstimmung sorgfältig durchdacht werden muss [8].

Während bei der hierarchischen Zerlegung unterschiedlich genaue Abbildungen des Problems auf unterschiedlich langen Horizonten verwendet werden, um eine Lösung mit mathematischen Optimierungsalgorithmen zu ermöglichen, wird bei der heuristischen Planung ein einheitliches Modell benutzt. Hier werden alle Abhängigkeiten und Beschränkungen über den gesamten Planungshorizont im Detail berücksichtigt. Dadurch ist es ausgeschlossen, dass sich eine Lösung, die unter vereinfachten Annahmen bestimmt wurde, nachher als nicht implementierbar erweist, andererseits entsteht ein hoher Aufwand bei der detaillierten Planung in der fernen Zukunft, in der die Annahmen der Planung längst überholt sein können. Der Kern erfolgreicher Werkzeuge für die Lösung solcher hochkomplexer Planungsprobleme auf langen Horizonten ist die iterative Verbesserung einmal gefundener zulässiger Lösungen [29]. Zunächst wird mit heuristischen Regeln (beispielsweise Rückwärtsterminierung von den Lieferterminen aus mit angenommenen Durchlaufzeiten für die einzelnen Einheiten und Lösung von Konflikten mit Prioritätsregeln wie „dringlichster Auftrag“ ein zulässiger Plan erzeugt. Dieser wird dann durch Anwendung lokaler Optimierungsverfahren und Heuristiken verbessert. Dieser Suchprozess kann vom Anwender durch manuelle Umplanungen, Fixieren von bestimmten Produktionen und Auswahl der Suchmethoden und der erlaubten Rechenzeit stark beeinflusst werden.

Beide Ansätze sind in kommerziellen Planungswerkzeugen implementiert [8, 29] und im industriellen Einsatz.

3 STATUS QUO

3.1 ENERGIEVERSORGUNG

PARADIGMENWECHSEL VOM TECHNISCHEN NETZ ZUR HANDELSPLATTFORM

Um die energiepolitischen Vorgaben der EU bis zum Jahr 2020,

1. 20 % der elektrischen Energie durch regenerative Anlagen zu erzeugen,
2. 20 % der CO₂ Emissionen zu reduzieren,
3. 20 % Effizienzsteigerung in der Erzeugung elektrischer Energie,

zu erreichen, werden neue Anforderungen an die Netze der Zukunft gestellt. Im Rahmen des Paradigmenwechsels hin zu liberalisierten Energiemärkten stellt die EU-Technologie-Plattform „Smart Grids“ den Schlüssel für die Erreichung der oben genannten Ziele dar. In Zukunft wird der bisherige unidirektionale Energiefluss von Kraftwerken zu den Kunden durch die Entwicklung dezentraler Erzeugungsanlagen im kW-Bereich in unmittelbarer Nähe der Kunden zu neuen Netzbelastungen führen. Diese Entwicklung wird durch die technologische Entwicklung neuer kleiner Energieerzeugungsanlagen für Strom und

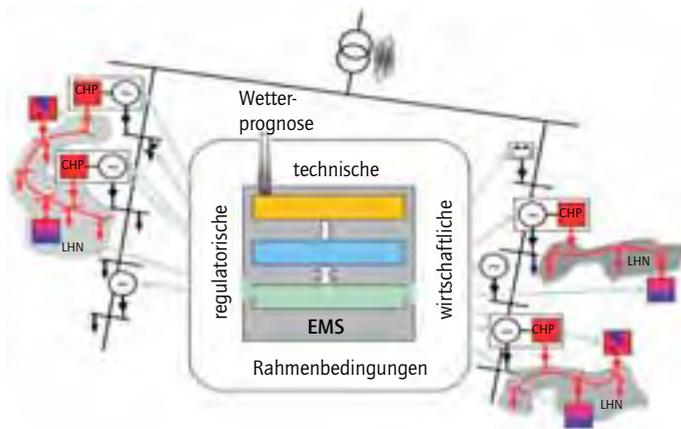
Wärme stark gefördert [15]. Man bezeichnet diese Entwicklung als Internet der Energie, da in Zukunft Millionen von Anlagen gleichzeitig in Betrieb stehen werden und über das Verteilungsnetz Energie austauschen können. Allerdings können diese Anlagen nur dann wirtschaftlich sinnvoll in das Netz integriert werden, wenn zwischen Handel, Netzbetreiber und Kundenanlagen bidirektionale Informationswege bestehen.

Durch den Einbau digitaler Messgeräte zur Erfassung des Energieverbrauchs werden die Voraussetzungen geschaffen, um über derartige Gateways beim Kunden Informationen in beiden Richtungen auszutauschen. Dabei ist zu beachten, dass erst der zusammenfassende, koordinierte Betrieb vieler dezentraler Anlagen zu wirtschaftlichen Lösungen führt. Deshalb muss eine elektronische Plattform aufgebaut werden, zu der alle Akteure einen einfachen, diskriminierungsfreien Zugang haben. Das Internet wird beim Aufbau einer derartigen gemeinsamen Handelsplattform eine zentrale Rolle spielen. Allerdings müssen dazu gemeinsame Datenprotokolle entwickelt werden, um eine nahtlose Kommunikation zwischen allen Beteiligten ermöglichen zu können. Gleichzeitig ist dem Problem der Datenintegrität große Aufmerksamkeit zu schenken, indem durch entsprechende Verschlüsselungstechnik sichergestellt wird, dass nur die jeweils Berechtigten den entsprechenden Datenzugang erhalten. Mit Hilfe eines Agentensystems sind dabei sowohl die Belange des Handels wie auch des Netzbetreibers zu jedem Zeitpunkt zu berücksichtigen. Man spricht in diesem Zusammenhang von einem virtuellen Kraftwerk, das von einer Stelle aus nach wirtschaftlichen Gesichtspunkten mindestens im 15-Minuten-Raster geführt und optimiert werden muss [25]. Dabei spielt die Berücksichtigung von Unsicherheiten bezüglich zukünftiger Lastanforderungen und Preisentwicklungen eine ganz zentrale Rolle, da nur so ein Weg zu Echtzeittarifen gefunden werden kann. Die mehrstufigen stochastischen Optimierungsverfahren weisen ein großes Potential auf, um den extrem komplexen Anforderungen der Praxis gerecht werden zu können. Ferner muss dabei die Möglichkeit bestehen, dass einzelne Geräte auf der Kundenseite über das oben genannte Gateway ein- und ausgeschaltet werden können, um so den Kunden die Möglichkeit zu geben, Zeitbereiche mit niedrigem Strompreis (z. B. wegen starker Windeinspeisung) optimal nutzen zu können.

OPTIMIERTES ENERGIEMANAGEMENT FÜR VIRTUELLE KRAFTWERKE

Virtuelle Kraftwerke (siehe Abb. 5, [25]) benötigen für deren Betrieb innovative Energiemanagementsysteme, die auf stochastischer Optimierung aufbauen. Das virtuelle Kraftwerk besteht aus einer Reihe kleiner dezentraler Energieumwandlungsanlagen, die alle gemeinsam in unmittelbarer Kundennähe eingesetzt und optimiert werden. Da das virtuelle Kraftwerk nicht nur mit ungenauen Prognosen für den Strom- und Wärmebedarf sondern auch mit unbekanntem Strompreisen des Spot Handels konfrontiert ist, muss ein stochastisches Verfahren eingesetzt werden, das diese Unsicherheiten berücksichtigt. In enger Zusammenarbeit der Autoren dieses Beitrages ist ein für die Praxis sehr gut geeignetes Optimierungsverfahren entwickelt worden, das im Viertelstundenraster den

Abb. 5: Energiemanagement für ein virtuelles Kraftwerk



optimalen Einsatzplan für alle im virtuellen Kraftwerk zusammengefassten Anlagen berechnet und überwacht. Die Berücksichtigung aller Details führt zu einer sehr komplexen Zielfunktion. Damit ist es erstmals gelungen, ein komplexes Optimierungsproblem unter realistischen Randbedingungen zu lösen. Da die stochastische Optimierung aus zwei Stufen besteht, können für die Unsicherheit der zweiten Stufe verschiedene Szenarien mit unterschiedlichen Wahrscheinlichkeiten angenommen werden. Natürlich nimmt die Rechenzeit mit der Anzahl der Szenarien sehr stark zu, so dass hier eine sorgfältige Auswahl zwingend erforderlich ist. Die entwickelte Lösung wurde auf ein reales virtuelles Kraftwerk angewendet. Es wurde dabei gezeigt, dass aus betrieblicher Sicht erhebliche Vorteile sowohl in technischer wie auch wirtschaftlicher Hinsicht erzielt werden können.

KOMPLEXITÄT DER NETZE UND MATHEMATISCHE HERAUSFORDERUNGEN

Gleichzeitig mit der Entwicklung dezentraler Einspeiser führt der Bau großer Windparks sowohl on-shore wie auch off-shore zwangsläufig zum Bau neuer Leitungen innerhalb des bestehenden Transportnetzes, um diese Energie in die entsprechenden Lastzentren transportieren zu können. Durch das stark fluktuierende Einspeiseverhalten von Windenergieanlagen müssen stabilisierende Maßnahmen gefunden werden, um unerwünschte Leistungspendelungen unterdrücken zu können. Dazu muss mit umfangreichen Netzanalysen festgelegt werden, wo entsprechende FACTS-Geräte eingebaut werden müssen, um einen größtmöglichen dämpfenden Effekt erzielen zu können. Die resultierenden dynamischen Simulationsmodelle sind außerordentlich umfangreich und komplex, denn

sie müssen einen großen Zeitbereich abdecken. Die dabei zu lösenden steifen Differentialgleichungssysteme stellen große Anforderungen an die zu entwickelnden numerischen Lösungsmethoden, da sonst die Kapazität der heute verfügbaren Rechner schnell überschritten wird. Mit Hilfe geeigneter Parallelisierungsmethoden lassen sich die Probleme in handhabbare Unterprobleme unterteilen.

Neben den netzspezifischen Problemen der Windenergieanlagen ist auch die Windenergieveredelung zu sehen. Da im Strommarkt nur konstante Leistungsbänder von jeweils 15 Minuten Dauer gehandelt werden, muss die stochastische Windeinspeisung in ein Band umgewandelt werden (Windstromveredelung) [36]. Die bisherige Praxis sieht vor, dass die Windeinspeisung innerhalb der vier deutschen Regelzonen in jeweils einem eigenen Bilanzkreis gehandelt und abgerechnet wird.

Neben einer zuverlässigen Windprognose sind optimierungsbasierte Verfahren zu entwickeln, um die erforderliche Regel- und Ausgleichsenergie rechtzeitig und preisgünstig zu beschaffen. Nur so wird es möglich sein, die mit der Windenergieveredelung verbundenen Kosten in einem akzeptablen Rahmen zu halten. Da bereits heute die technischen Voraussetzungen bei den Windenergieanlagen erfüllt sind, ist in Zukunft auch zu überlegen, in welchem Maße diese zur Erbringung von Netzdienstleistungen herangezogen werden können. Dies gilt insbesondere auch für die Bereitstellung von Regelreserve, um so diese sehr teure Energie trotz des weiteren Ausbaus der Windenergie begrenzen zu können.

NEUE HERAUSFORDERUNGEN DES 21. JAHRHUNDERTS

Zusammenfassend ist festzuhalten, dass die Netze des 21. Jahrhunderts völlig neue und anspruchsvolle Aufgaben zu erfüllen haben. Tab. 1 zeigt in vereinfachter Form die wesentlichen Unterschiede zu den bisherigen Netzen.

Natürlich ist der Übergang zu den Netzen des 21. Jahrhunderts ein langfristiger und kostenintensiver Prozess, da die Betriebsmittel der elektrischen Energietechnik eine sehr lange Lebensdauer von 40 Jahren und mehr haben. Deshalb sind die zukünftigen Netzinvestitionen in engem Zusammenhang mit einem modernen Asset Management aller Betriebsmittel zu sehen [20]. Neben der sorgfältigen Wahl geeigneter Instandhaltungs- und Wartungsstrategien unter angemessener Berücksichtigung des Risikos müssen auch Erneuerungsstrategien festgelegt werden, mit denen dieser Übergang technisch und wirtschaftlich sinnvoll realisiert werden kann. Da die Planung und der Bau von neuen Kraftwerken und dezentralen Einspeisungen unabhängig vom Netzbetreiber durchgeführt werden, müssen alle Vorkehrungen getroffen werden, um Fehlinvestitionen im Netz zu vermeiden.

Tab. 1: Netze des 21. Jahrhunderts

Netze des 20. Jahrhunderts	Netze des 21. Jahrhunderts
Unidirektionale Kommunikation	Bidirektionale Kommunikation
Wenige Sensoren und Datenerfassungsgeräte	Sensoren und Datenerfassung im gesamten System
Beschränkte Leistungsfluss-Steuerung	Umfangreiche Steuermöglichkeiten in Stationen und Netzen
Suboptimale Nutzung der Betriebsmittel	Modernes Asset Management aller Betriebsmittel mit Zustandserfassung und dynamischen Grenzwerten
Allein operierende Informationssysteme innerhalb eines Energieversorgungsunternehmens	Integrierte Informationssysteme auf der Basis des Internet mit diskriminierungsfreiem Zugang für alle Marktteilnehmer
Wenige, einzelne dezentrale Einspeisungen	Millionenfache dezentrale Einspeisung und aktive Beteiligung der Kunden durch aktives Lastmanagement
Begrenzte Preisinformation für Kunden, feste Tarife	Volle Preistransparenz, Echtzeittarife, aktive Kundenbeteiligung

3.2 VERFAHRENSTECHNISCHE PRODUKTIONSPROZESSE

SUPPLY-CHAIN OPTIMIERUNG

In allen großen Unternehmen der Metallerzeugung, der Spezialchemie und der Nahrungsmittelproduktion wurden und werden Projekte zur Verbesserung der Produktionslogistik durchgeführt, meist unter dem Schlagwort der „Supply-Chain-Optimierung“. Dies bedeutet eine Gesamtsicht des Produktions- und Verteilungsprozesses vom Einkauf und der Lagerung der Rohmaterialien über die Produktion bis zur Distribution über dezentrale Lager oder Großhändler [40]. Tatsächlich ist eine integrierte Optimierung dieser Gesamtprozesse derzeit allerdings kaum möglich.

Das Hauptwerkzeug der Untersuchung und Verbesserung der Produktionslogistik ist die ereignisdiskrete Modellierung und Simulation [37]. Ereignisdiskret bedeutet, dass nicht wie bei der kontinuierlichen Simulation zeitliche Veränderungen von Prozessgrößen berechnet werden sondern eine Abfolge diskreter Ereignisse, beispielsweise Start und Ende von Produktions-, Reinigungs- und Transportprozessen. Im Unterschied zur reinen Stückgutproduktion müssen hier allerdings stoffliche Eigenschaften wie die Konzentrationen verschiedener Stoffe in einer Charge oder in einem Tank in ihrer Veränderung im Zuge des Produktionsprozesses oder durch Einfüllen von weiteren Mengen in Tanks mit berücksichtigt werden. Mit Hilfe der ereignisdiskreten Simulation werden

unterschiedliche Szenarien (beispielsweise verschiedene Anzahlen und Kapazitäten von Produktions- und Abfülleinrichtungen oder unterschiedliche Verteilungen der einzelnen Produktionen auf Anlagen oder Standorte) durchgespielt. Bei den Kapazitäts- und Produktionsplanungsentscheidungen werden die Ergebnisse der Simulation verschiedener Szenarien einbezogen, es findet aber nur selten eine Optimierung im strengen Sinne statt.

Der kritische Punkt bei dieser Vorgehensweise ist offensichtlich die Auswahl der betrachteten Szenarien der Nachfrage- und ggf. der Preisentwicklung. Hier kommt eine erhebliche Unsicherheit ins Spiel, die von den Entscheidern mehr oder wenig intuitiv bewertet wird. Es kann festgestellt werden, dass die Handhabung dieser Unsicherheit derzeit den größten Schwachpunkt in der Anwendung der logistischen Optimierung darstellt. Die Prognosen oder Prognoseszenarien werden in der Regel den „Produktionsleuten“ von den Marketingabteilungen als unumstößliche Voraussetzungen der Planung mitgeteilt. Die erhebliche Unsicherheit dieser Daten wird kaum quantifiziert. In der Folge wird unter Umständen mit hohem Aufwand eine Optimierung durchgeführt, die auf Annahmen aufbaut, die nicht Realität werden.

MITTELFRISTIGE PRODUKTIONSPLANUNG

Mittelfristige Produktionsplanung betrifft die Vorausplanung von Produktionsprozessen über einen Zeitraum von einigen Monaten bis hin zu mehreren Jahren. Hier geht es um die konkrete Verteilung von Aufträgen auf Produktionseinheiten und ihre Terminierung. Eine große Rolle spielt hierbei, wie oben schon angesprochen, die Kampagnenbildung. Durch die Produktion möglichst vieler gleichartiger Chargen hintereinander wird Umrüst- und Reinigungsaufwand vermieden. Sehr lange Kampagnen erfordern aber zur Bedarfsdeckung eine große Lagerhaltung, die ihrerseits zu Kosten führt und mit dem Risiko behaftet ist, nicht nachgefragte Produkte unnötig herzustellen, die dann verramscht oder gar teuer entsorgt werden müssen.

Für die mittelfristige Produktionsplanung sind inzwischen leistungsfähige Werkzeuge verfügbar, wie im vorigen Abschnitt beschrieben. Aus mathematischer Sicht handelt es sich um große Optimierungsprobleme mit ganzzahligen oder binären Variablen. In der Regel ist eine Modellierung als lineares gemischt-ganzzahliges Problem (MILP) ausreichend. Nichtlineare Phänomene werden hier in der Regel durch lineare Zusammenhänge approximiert, was im Verhältnis zu den über längere Planungshorizonte vorliegenden Unsicherheiten nur geringe zusätzliche Fehler verursacht. Modelle für die mittelfristige Planung beinhalten im wesentlichen Massenbilanzen für die gelagerten oder umgesetzten Stoffe und Beschränkungen der Ressourcennutzung und der zeitlichen Abfolge. Beschränkungen der Ressourcennutzung können vielfältiger Art sein, neben der Verfügbarkeit von Apparaten spielt die Personalkapazität und die Bereitstellung der benötigten Einsatzstoffe eine wesentliche Rolle. Beschränkungen der zeitlichen Abfolge sind die Reihenfolge der Produktionsschritte, Verzögerungen durch Umrüst- und Reini-

gungsoperationen, notwendige Ruhezeiten oder maximale Haltbarkeiten von Zwischenprodukten, Analysezeiten usw. Die resultierenden Problemgrößen sind bei vereinfachten Modellierungen so, dass sie mit state-of-the-art MILP-Solvern gerade noch mit Rechenzeiten im Stundenbereich gelöst werden können. Diese Rechenzeiten sind für Planungsaufgaben sicherlich akzeptabel, jedoch kaum geeignet, um eine größere Zahl von Alternativen bei der Planung der Supply-Chain optimierungsbasiert zu bewerten. Hierfür sind semi-heuristische Verfahren, die brauchbare zulässige Lösungen in relativ kurzen Zeiten erzeugen können, besser geeignet.

Ein entscheidendes Problem der mittelfristigen Produktionsplanung ist die Verfügbarkeit der benötigten Daten. Immerhin sind hier in der Größenordnung von 10 000 unterschiedliche Produktionsrezepte zur Verfügung zu stellen, in denen für jeden Schritt die benötigten Ressourcen präzise definiert sein müssen. Diese Daten müssen sinnvollerweise aus einer Rezeptdatenbank automatisch in der benötigten Detaillierung generiert werden, wie dies in integrierten Produktionsplanungs- und Steuerungssystemen (ERP-Systeme) heute möglich ist.

ONLINE-PRODUKTIONSSTEUERUNG

Die Ergebnisse der mittelfristigen Produktionsplanung müssen zur Durchführung der Produktion in eine detaillierte Ablaufsteuerung umgesetzt werden. Dies wird auch als Scheduling bezeichnet. Hierbei sind zum einen detailliertere Beschränkungen und Zusammenhänge zu berücksichtigen als bei der mittelfristigen Planung, zum anderen sind die unvermeidlichen Abweichungen vom bei der Planung angenommenen Ablauf und der vorausgesetzten Verfügbarkeit von Ressourcen zu berücksichtigen und soweit wie möglich auszugleichen. Dies erfordert detailliertere Modelle mit einer hohen zeitlichen Auflösung und kurze Rechenzeiten um auf Störungen und Veränderungen wie Eilaufträge rasch reagieren zu können. Der Planungshorizont ist naturgemäß kürzer, da sich eine detaillierte Vorausberechnung in eine fernere Zukunft nicht lohnt. Obwohl auch für diese Aufgabe optimierungsbasierte Werkzeuge verfügbar sind, wird in der Praxis in der Regel eine manuelle Steuerung auf der Grundlage einer Visualisierung des Prozessablaufs inklusive einer Prognose der weiteren Entwicklungen vorgezogen. Dies scheint vor allem auf das generelle Problem der fehlenden Akzeptanz von durch mathematische Optimierung gewonnenen, „undurchsichtigen“ Lösungen zurückzuführen zu sein. Weitere Faktoren sind der hohe Aufwand, der zur detaillierten Modellierung aller zu betrachtenden Randbedingungen und zur Wartung der Daten, die der Anlagenfahrer „im Kopf“ hat, erforderlich ist. Schließlich werden die algorithmischen Probleme mit der Einbeziehung von komplexen Restriktionen und nichtlinearen Zusammenhängen gegenüber der mittelfristigen Planung größer während die zur Verfügung stehende Lösungszeit abnimmt.

3.3 MATHEMATISCHE METHODEN

Mathematische Methoden sind bei der Formulierung und Lösung von Ressourcenplanungsproblemen in der Energie- und Verfahrenstechnik seit einigen Jahren etabliert. Rechnerbasierte Entscheidungshilfen sind, wie in den vorherigen Abschnitten dargestellt, heute aus der Energie- und Verfahrenstechnik nicht mehr wegzudenken. Den „mathematischen Kern“ solcher Entscheidungshilfen machen Modelle und Algorithmen aus verschiedenen Teilgebieten der Optimierung aus. An erster Stelle sind hier die gemischt-ganzzahlige lineare und die nichtlineare Optimierung zu nennen. Mit der stochastischen Optimierung ist eine Disziplin auf dem Vormarsch, die Lösungsvorschläge für das immer unumgänglichere Optimieren bei unvollständiger Information unterbreitet.

GEMISCHT-GANZZAHLIGE LINEARE OPTIMIERUNG

In die Ressourceneinsatzplanung gehen typischerweise sowohl kontinuierlich wählbare Größen ein, wie Verbrauchs- und Produktionsmengen, als auch solche, die einer gewissen Stückelung folgen müssen. Diese Stückelung kann, wie bei unteilbaren Gütern oder Produktionschargen, in der Natur der Sache liegen. Noch bedeutsamer ist sie aber in Bezug auf Ja/Nein-Entscheidungen, die in der Regel zentrale Aspekte von Ressourcenplanungsproblemen ausmachen: Soll ein Kraftwerksblock angefahren werden oder nicht, soll ein Charge auf dieser oder auf jener Anlage produziert werden, soll erst A oder erst B hergestellt werden? Bei der Ressourceneinsatzplanung ist die Behandlung der Komponente „Zeit“ von zentraler Bedeutung. Eine häufig genutzte Methode ist die der Unterteilung des betrachteten Zeithorizonts in geeignete Entscheidungsintervalle (z. B. Stundenraster, Viertelstundenraster oder feiner). Hat man viele solche Intervalle und viele mögliche Alternativen in jedem Intervall, dann führt dies sehr schnell auf Entscheidungsprobleme mit mehreren Tausend ja oft mehreren Hunderttausend oder sogar Millionen von Unbekannten.

Die gemischt-ganzzahlige lineare Optimierung bietet ein flexibles Methodenspektrum zur mathematischen Modellierung und Lösung solcher Entscheidungsprobleme. Beispielsweise bei der Betriebs- und Investitionsplanung in Energieversorgungssystemen oder der Produktionsplanung in der chemischen und pharmazeutischen Industrie haben sich Modelle und Algorithmen der gemischt-ganzzahligen linearen Optimierung mittlerweile recht gut etabliert. Entscheidend dafür war nicht zuletzt die Bereitstellung leistungsfähiger Software², in der neueste mathematische Forschungsergebnisse nutzerfreundlich umgesetzt worden sind. Es gibt allerdings Problemklassen wie das sog. Job-Shop-Scheduling, die Verteilung von Aufträgen auf eine Reihe möglicher Ressourcen, die mittels gemischt-ganzzahliger linearer Programmierung nicht effizient lösbar sind.

² ILOG CPLEX (<http://www.ilog.com>), SCIP (<http://scip.zib.de>), Xpress-MP (<http://www.dashoptimization.com>)

Hier haben sich Methoden des Constraint Programming als effizienter erwiesen. Die Kombination beider Herangehensweisen für Probleme, die sowohl Aspekte der reinen Reihenfolgeplanung als auch der linearen Optimierung aufweisen, ist ein vielversprechendes Arbeitsfeld für die Zukunft.

In der aktuellen mathematischen Forschung ist die gemischt-ganzzahlige lineare Optimierung darüber hinaus für Untersuchungen zur Approximation von Nichtlinearitäten und zur Dekomposition stochastischer Optimierungsprobleme von Bedeutung.

NICHTLINEARE OPTIMIERUNG

Lastflüsse in elektrischen Netzen und Ausbeuten verfahrenstechnischer Prozesse sind Beispiele für nichtlineare Effekte mit entscheidendem Einfluss auf die Ergebnisse der Ressourcenplanung. Reichen lineare Approximationen solcher Phänomene nicht mehr aus, so wird die direkte Behandlung der Nichtlinearitäten unumgänglich. Solange die gleichzeitige Betrachtung kombinatorischer und/oder stochastischer Effekte vermieden wird, ergeben sich im Zuge der mathematischen Modellierung oft hochdimensionale, nichtlineare Optimierungsaufgaben, die mit vorhandener Algorithmik und Software³ häufig innerhalb von akzeptablen Rechenzeiten erfolgreich gelöst werden können. In der Regel meint „lösen“ dabei jedoch die Ermittlung lokaler Optimalpunkte oder lediglich solcher Punkte, die notwendige Optimalitätsbedingungen erfüllen. Praktisch bedeutet dies, dass es durchaus noch bessere Lösungen geben kann (Stichwort: globale Optima).

Viele praktische Optimierungsprobleme enthalten jedoch sowohl nichtlineare Gleichungen als auch diskrete Entscheidungsvariablen, so genannte gemischt-ganzzahlige nichtlineare (MINLP) Probleme. Die Entwicklung von Lösungsverfahren für diese Problemklasse steht noch sehr am Anfang.⁴

STOCHASTISCHE OPTIMIERUNG

Wie verschiedentlich bereits angesprochen, stellen sich bei genauer Betrachtung viele Aufgaben der Ressourcenplanung in der Energie- und Verfahrenstechnik als Optimierungsprobleme unter unvollständiger Information dar: Zum Zeitpunkt der Entscheidungsfindung sind essentielle Modelldaten, wie z. B. Ein- und Ausspeisemengen, Verfügbarkeiten von Anlagen oder Ein- und Verkaufspreise, nur unzulänglich bekannt.

In der Vergangenheit hat man sich oft so beholfen, dass unbekannte Modellgrößen durch Schätzungen (Erwartungs- oder Mittelwerte) ersetzt wurden. Zum Teil fuhr man

³ IPOPT (<https://projects.coin-or.org/Ipopt>), LANCELOT (<http://www.cse.scitech.ac.uk/nag/lancelot/lancelot.shtml>), SNOPT (http://www.sbsi-sol-optimize.com/asp/sol_product_snopt.htm)

⁴ Erste Teilergebnisse (<http://egon.cheme.cmu.edu/ibm/page.htm>) betreffen gemischt-ganzzahlige konvexe Modelle. Nur fallen relevante Ressourcenplanungsprobleme in Energie- und Verfahrenstechnik oft nicht in diese Klasse.

damit recht gut, wie etwa bei der Lastschätzung in elektrischen Versorgungssystemen zu Zeiten regulierter Energiemärkte. Deregulierung und Dezentralisierung haben aber mittlerweile dazu geführt, dass Ressourcenplanung einer solchen Vielfalt an Modellunsicherheit ausgesetzt ist, dass einfache Schätzungen unzulänglich sind und vielmehr das Wechselspiel zwischen Entscheidungsfindung und Informationsgewinn explizit mit in die Optimierungsmodelle einzubeziehen ist.

Als Beispiel sei die Windenergieeinspeisung in elektrische Netze betrachtet. Setzt man im Modell nur eine Schätzung an, so können im Falle von größeren Abweichungen entweder gravierende Betriebsstörungen die Folge sein (z. B. Defizite bei der Lastdeckung oder Überlastung von Übertragungsleitungen) oder man hatte sich im Vorfeld konservativ gegen „alle Eventualitäten“ abgesichert, was sehr teuer sein und Wettbewerbspositionen schwächen kann. Setzt man stattdessen eine Wahrscheinlichkeitsverteilung für die Windeinspeisung an und berücksichtigt, welche Entscheidungen vor und welche nach Beobachtung der realen Windeinspeisung zu treffen sind, so ergibt sich die Möglichkeit, den Informationszuwachs im Modell zu erfassen und weniger konservative aber dennoch praktikable und somit wettbewerbsfähigere Lösungen zu erzielen.

Die stochastische Optimierung bietet flexible Instrumente zur Berücksichtigung von Zufallseinfluss in Optimierungsmodellen. In den Abbildungen 6 und 8 sind risikoneutrale und risikoaverse zweistufige stochastische Optimierungsmodelle mathematisch etwas detaillierter beschrieben. Ansätze der stochastischen Optimierung werden in der Energie- und Verfahrenstechnik zunehmend aufgegriffen. Als Beispiele seien genannt:

- frühe Arbeiten zur Kraftwerkseinsatzplanung bei unvollständiger Information [26, 27, 41, 42],
- Arbeiten zum Batch Scheduling unter Ungewissheit in chemischen Mehrproduktanlagen [2, 4, 43],
- Arbeiten zur Planung der Erschließung von Öl- und Gasfeldern durch Bohrschiffe [5, 45],
- erste Arbeiten zur stochastischen Optimierung in Energieversorgungssystemen mit verteilter Erzeugung [12, 25].

4 PERSPEKTIVEN UND HERAUSFORDERUNGEN

4.1 ENERGIEVERSORGUNG

Der durch die Liberalisierung ausgelöste Paradigmenwechsel in der Energieversorgung führt zu neuen Anforderungen an die Mathematik sowohl bei der Planung von Netzerweiterungen als auch beim Netzbetrieb. Dadurch, dass der Kunde in Zukunft im Mittelpunkt aller Überlegungen steht, werden seine Verbrauchsdaten als neue Optimierungsgrößen neben den bisherigen Größen zur Verfügung stehen. Sein Bemühen wird sich auf die Beschaffung preisgünstiger Energie konzentrieren.

Optimierte effiziente Strombeschaffung. In enger Zusammenarbeit mit dem Handel müssen optimierte Beschaffungsstrategien entwickelt werden, bei denen die Ungewissheit bezüglich zukünftiger Preis- und Lastentwicklung angemessen berücksichtigt werden kann. Mehrstufige stochastische Optimierungsverfahren berücksichtigen dabei die Tatsache, dass die unmittelbare Zukunft mit einer größeren Genauigkeit vorausgesagt werden kann, als weiter entfernt liegende Ereignisse. Mit der Formulierung realistischer Zukunftsszenarien können Entwicklungen abgeschätzt werden, die einen großen Einfluss auf die kurzfristigen Entscheidungen haben. Wegen der sehr großen Anzahl an Akteuren wird die Entwicklung dezentraler resp. hierarchischer Optimierungsverfahren an Bedeutung gewinnen, um so die mögliche Bildung von Teilnetzen mit wirksamen Gegenmaßnahmen verhindern zu können.

Risikooptimierung. Im Bereich des Asset Management sind neue Risiko basierte Verfahren zu entwickeln, für deren Einsatz allerdings eine zuverlässige und ausreichende Datenbank zur Verfügung stehen muss. Die entsprechenden Daten sind in einem aufwändigen aber notwendigen Prozess zu beschaffen, um neue Optimierungsmethoden einsetzen zu können, die auch eine Budgetbegrenzung berücksichtigen können. Diese Verfahren sind in enger Zusammenarbeit mit dem Netzregulator zu entwickeln, da durch die Anreizregulierung der Handlungsrahmen des Netzbetreibers mehr und mehr eingeschränkt wird. Insbesondere die richtige Formulierung einer Qualitätskomponente in der Anreizregulierung wird bei der Formulierung der Zielfunktion für ein optimales Asset Management eine wichtige Rolle spielen. Da es sich auch hier um Unsicherheiten handelt, deren Grenzen allerdings angegeben werden können, werden die Methoden „unknown but bounded“ weiterzuentwickeln sein, um adäquate Lösungen ableiten zu können [38].

Optimierter Betrieb unter Unsicherheit. Von der betrieblichen Seite wird der Netzbetreiber gefordert, sein Netz bis an die technischen Grenzen zu belasten. Damit erhalten Stabilitätsuntersuchungen zukünftig wieder große Aufmerksamkeit, wobei sowohl das Frequenz wie auch das Spannungsverhalten im Detail mit dynamischen Modellen simuliert und untersucht werden muss. Mit optimierungsbasierten neuen Steuerungsmethoden und durch den Einsatz der modernen Leistungselektronik muss der ÜNB und der VNB dafür Sorge tragen, dass jederzeit die bestmögliche Transportkapazität in seinem Netz zur Verfügung steht. Mit dem Einsatz neuer Datenerfassungssysteme wie dem Weitbereichsmonitoring stehen zusätzliche Informationen zur Verfügung, die mit Hilfe der Signalanalyse ausgewertet werden, um den aktuellen Netzzustand optimal und zuverlässig bestimmen zu können. In enger Abstimmung mit dem Handel und den Kunden ist laufend zu überprüfen, wann und wo ggf. Netzengpässe auftreten könnten, um rechtzeitig geeignete, optimierungsbasierte Abwehrmaßnahmen einleiten zu können.

Abb. 6: Risikoneutrales zweistufiges stochastisches Optimierungsmodell

Produktionsplanung unter Ungewissheit. Ziel ist die Bedarfsdeckung zu minimalen Kosten. Zum Zeitpunkt der Planung sind Bedarfe, Kosten oder zukünftige Störungen mit Unsicherheit behaftet. Nach erfolgter Planung und nach Beobachtung der unsicheren Größen können weitere Entscheidungen getroffen werden, z. B. solche, die Abweichungen zwischen Bedarf und Ausstoß der vorher geplanten Produktion kompensieren. Diese Entscheidungen sind unter Einhaltung gewisser Bedingungen wählbar und verursachen Kosten. Sie ergeben sich somit im allgemeinen als Resultat einer nachgelagerten Optimierung. Die Planungen seien mit x bezeichnet und aus einer Menge X zu wählen. Sie verursachen direkte Kosten $c^T x$. Die Kompensationen seien y . Sie mögen auf die Menge Y beschränkt sein. Außerdem gibt es Bedingungen, die x und y verknüpfen. Auf diese Bedingungen, wie auch auf die durch y verursachten Kosten wirke der Zufall. Wir fassen die Bedingungen als $T(\omega)x + W(\omega)y = z(\omega)$ und die Kosten als $q(\omega)^T y$, wobei der Zufallseinfluss durch ω markiert wurde. Die Planungen x heißen Entscheidungen der ersten Stufe. Sie sind „hier und heute“, d. h. unabhängig von

ω , zu treffen, dürfen also die erst später beobachtbaren zufälligen Größen nicht antizipieren. Die Kompensationen y heißen Entscheidungen der zweiten Stufe („recourse“). Sie sind adaptiv je nach Planung x und Beobachtung (oder Szenario) ω wählbar (Abb. 7). Das klassische risikoneutrale zweistufige stochastische Optimierungsproblem kann nun wie folgt formalisiert werden

$$\min\{c^T x + E[\Phi(x, \omega)] : x \in X\}$$

wobei

$$\Phi(x, \omega) := \min\{q(\omega)^T y : T(\omega)x + W(\omega)y = z(\omega), y \in Y\}.$$

Über alle zulässigen Planungen $x \in X$ wird also die Summe aus den direkten Kosten $c^T x$ und dem Erwartungswert der durch x verursachten zukünftigen Kosten minimiert. Bei der Berechnung letzterer wird unterstellt, dass die adaptiven Entscheidungen $y = y(x, \omega)$ bestmöglich getroffen werden.

Abb. 7: Zweistufige stochastische Optimierung

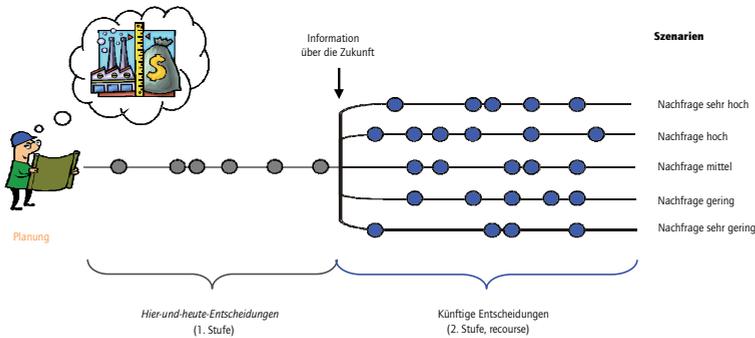


Abb. 8: Abstrakter Ansatz zur zweistufigen stochastischen Optimierung mit Risikoaversion

Gegeben sei das zufallsbehaftete Optimierungsproblem

$$"min" \{c^T x + q(\omega)^T y : T(\omega)x + W(\omega)y = z(\omega), x \in X, y \in Y\}$$

zusammen mit der Forderung, dass (in der ersten Stufe) zunächst x gewählt werden muss, dann die Realisierung der Zufallsgröße $\xi(\omega) := (q, T, W, z)(\omega)$ bekannt wird und schließlich (in der zweiten Stufe) über $y = y(x, \omega)$ entschieden werden kann. Es wird vorausgesetzt, dass die Wahrscheinlichkeitsverteilung von ξ bekannt ist und – für die hiesige Betrachtung – unabhängig von x .

Da zum Zeitpunkt der Wahl von x die Nebenbedingungen des Problems nicht vollständig bekannt sind, ist obige „Minimierung“ noch nicht wohldefiniert. Wir betrachten das Problem unter folgendem Blickwinkel:

$$\begin{aligned} & "min_x" \{c^T x + \underbrace{\min_y \{q(\omega)^T y : W(\omega)y = z(\omega) - T(\omega)x, y \in Y\}}_{\Phi(x, \omega)} : x \in X\} \\ & = "min_x" \{ \underbrace{c^T x + \Phi(x, \omega)}_{f(x, \omega)} : x \in X \} \\ & = "min_{x \in X}" f(x, \omega) \end{aligned}$$

Die Auswahl optimaler x gestaltet sich so zur Auswahl aus der Familie von Zufallsvariablen $(f(x, \omega))_{x \in X}$, wobei x als „Laufindex“ angesehen werden kann. Zur Bewertung der Zufallsvariablen $f(x, \omega)$ mit dem Ziel der Auswahl einer „kleinsten“ können statistische Parameter herangezogen werden. Im risikoneutralen Fall ist dies der Erwartungswert E , bei Einbeziehung von Risikoaversion eine gewichtete Summe $E + \rho \cdot \mathcal{R}$ aus Erwartungswert und einem Risikomaß \mathcal{R} , mit welchem sich unterschiedliche Auffassungen von „Risiko“ formalisieren lassen (siehe Abb. 9).

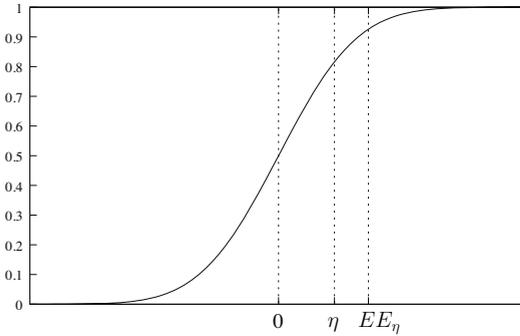
So erhalten wir (wohldefinierte!) Mean-Risk-Optimierungsprobleme vom Typ

$$\min \{ \underbrace{(E + \rho \cdot \mathcal{R})[f(x, \omega)]}_{Q(x)} : x \in X \}.$$

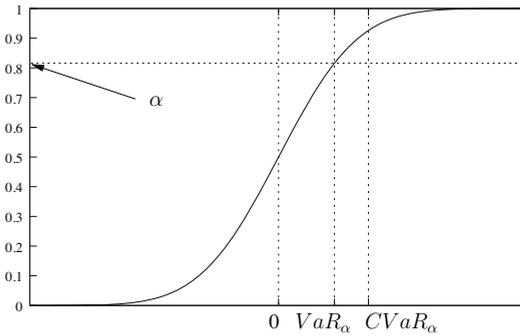
Es wird eine nichtantizipative Erststufenentscheidung x ermittelt, welche die gewichtete Summe aus Erwartungswert und Risiko der Gesamtkosten $f(x, \omega)$ minimiert. Dabei ist berücksichtigt, dass x ohne Kenntnis von $\xi(\omega)$ zu wählen ist und dass nach Festlegung von x und Beobachtung von $\xi(\omega)$ die verbleibenden Entscheidungen y unter Einhaltung der nun vollständig bekannten Nebenbedingungen bestmöglich gewählt werden.

Abb. 9: Beispiele für Risikomaße

Für eine Zufallsvariable $X(\omega)$ und ein Kostenniveau η gibt der *Gemittelte Überschuss* (Expected Excess) $EE_\eta(X) = E[\max\{X(\omega) - \eta, 0\}]$ an, um wie viel der Wert von $X(\omega)$ das Niveau η im Mittel überschreitet:



Für eine Zufallsvariable $X(\omega)$ und ein Wahrscheinlichkeitsniveau α ist der *Value-at-Risk* $VaR_\alpha(X) = \min\{t : P[X(\omega) \leq t] \geq \alpha\}$ die günstigste der $(1-\alpha) \cdot 100\%$ ungünstigsten Realisierungen von $X(\omega)$. Der *Conditional Value-at-Risk* $CVaR_\alpha(X) = E[X(\omega) | X(\omega) \geq VaR_\alpha(X)]^a$ ist der Mittelwert der $(1-\alpha) \cdot 100\%$ ungünstigsten Realisierungen von $X(\omega)$:



^a Für diskret verteiltes $X(\omega)$ ist dieser Ausdruck zu modifizieren.

4.2 VERFAHRENSTECHNISCHE PRODUKTIONSPROZESSE

Problemangepasste Modellierung. Die mathematische Modellierung und optimierungsbasierte Lösung von Ressourcenzuteilungsproblemen in der verfahrenstechnischen Produktion hat in den letzten 10 Jahren gewaltige Fortschritte gemacht. Für viele Teilprobleme gibt es inzwischen Standard-Modellierungen und exemplarische Lösungen, teilweise auch kommerzielle Produkte, die solche Lösungen für den praktischen Einsatz zur Verfügung stellen. Das Geheimnis des Erfolgs in Anwendungen ist hierbei eine problemangepasste Modellierung des zugrundeliegenden Problems [32]. Viele Fragen der Modellierung, beispielsweise welche Beschreibung der Zeit (Intervalle, kontinuierlich, Time-slots) für welches Problem optimal ist, sind immer noch offen [23]. Eine neuere Entwicklung sind disjunktive Formulierungen anstelle der traditionellen big-M-Ansätze, die auf engere Schranken in den gemischt-ganzzahlig linearen Programmen führen [21].

Handhabung von Unsicherheit. Alle beschriebenen Planungs- und Steuerungsprobleme sind gekennzeichnet von erheblichen Unsicherheiten der zugrundeliegenden Daten. Interpretiert man die Daten als Mittelwerte oder wahrscheinlichste Werte der entsprechenden Größen, dann erhält man nicht die im Mittel der auftretenden Fälle optimale Lösung. Praktisch ist es darüber hinaus sehr wesentlich, den Effekt der unvermeidlichen Abweichungen von Modell und Realität (unterschiedliche Nachfrageentwicklungen, Apparate- und Anlagenausfälle, schwankende Ausbeuten usw.) auf die Wirtschaftlichkeit und insbesondere die Einhaltung von Randbedingungen zu berücksichtigen. Bei einem „auf Kante genähten“ Plan können viele Störungen dazu führen, dass der Plan überhaupt nicht mehr ausführbar ist, beispielsweise weil ein benötigter Rohstoff für eine Charge nicht verfügbar ist (etwas zu wenig ist viel zu wenig). Grundsätzlich ist daher anzustreben, die Optimierungen für stochastische Daten durchzuführen, wobei eine Modellierung über verschiedene Szenarios mit unterschiedlichen Wahrscheinlichkeiten derzeit als der gangbarste Weg erscheint. Die systematische und fundierte Generierung eines Satzes von Szenarien der mit möglichst wenig Elementen die auftretenden Fälle gut abdeckt ist ein derzeit aktuelles Forschungsgebiet.

Mensch-Maschine-Interaktion. Dieses Problem tritt vor allem dort auf, wo mit Hilfe der Optimierung konkrete Entscheidungen, was als nächstes wo, mit welchen Ressourcen produziert werden soll, zu treffen sind. Anlagenfahrer, Disponenten usw. haben ein tiefes Misstrauen gegenüber „mechanischen“ Lösungen, deren Zustandekommen sie nicht verstehen, die sie, zu Recht oder zu Unrecht, für ungünstig halten und in die, wegen der mangelnden Berücksichtigung von Unsicherheiten, ihre Erfahrung, was wahrscheinlich schief geht und was nicht, nicht einfließen kann. Die Verbesserung der Optimierung selbst hinsichtlich der Einbeziehung von Unsicherheiten wird ein wesentlicher Beitrag sein, bessere Akzeptanz zu erzielen, allein aber nicht ausreichen. Zukünftig muss im dispositiven Bereich auf eine Kopplung von mathematischer Optimierung und

menschlicher Übersicht und Erfahrung hingearbeitet werden. Dies betrifft sowohl die Kommunikation zwischen Algorithmus und Bediener, beispielsweise durch die anschauliche Darstellung von Alternativen als auch die Algorithmik selbst. Wenn der Nutzer oder die Nutzerin den Verlauf der Optimierung beispielsweise durch Vorgabe von Präferenzen oder durch Fixierungen einzelner wichtiger Entscheidungen beeinflussen kann, wird er (sie) die Ergebnisse hinsichtlich der vielen einzelnen Entscheidungen wesentlich eher akzeptieren und kann damit von aufwändigen Detailplanungen entlastet werden, ohne befürchten zu müssen, seine (ihre) Funktion und Bedeutung komplett zu verlieren.

4.3 MATHEMATISCHE METHODEN

Der sachgerechte und effiziente Umgang mit unvollständiger Information gestaltet sich mehr und mehr zu einer zentralen Herausforderung bei der Ressourcenplanung in der Energie- und Verfahrenstechnik. Die Mathematik leistet zur Bewältigung dieser Herausforderung bereits erste wichtige Beiträge. Das Potential vorhandener mathematischer Erkenntnisse wird aber bei weitem noch nicht ausgeschöpft. Gleichzeitig erfordern die praktischen Gegebenheiten weiterführende mathematische Untersuchungen, die zum Teil bereits angelaufen sind:

Adäquate Modellierung des Zufallseinflusses. Oftmals wird die eingangs fehlende Modellinformation nicht wie im zweistufigen Modell aus Abb. 8 auf einen Schlag verfügbar, sondern peu à peu. Dies führt auf mehrstufige Erweiterungen des genannten Modells, welche deutlich komplexer und bei Beteiligung ganzzahliger Variablen erst im Ansatz theoretisch verstanden sind. Der dem Modell unterliegende, den Informationszuwachs widerspiegelnde stochastische Datenprozess ist ebenfalls komplexer als im zweistufigen Fall und geht im allgemeinen in diskretisierter Form als Szenariobaum ins Modell ein. Die „optimale Bemessung“ dieser Bäume (nicht zu groß, da dies algorithmische Probleme bereitet; nicht zu klein, da dann der Zufall unzureichend erfasst wird) erfolgt oft noch rein empirisch. Dank neuester Ergebnisse mathematischer Forschung gibt es hier nun erste theoretisch fundierte und dadurch numerisch leistungsfähige Methoden, z. B. [16].

Risikomanagement. Die Anforderungen der Finanzmärkte haben zu Entwicklungen beim Risikomanagement geführt, auf die in der Optimierung unter Ungewissheit aufgebaut werden kann. Für die aktuelle Forschung ist die Einbindung ein- und mehrperiodischer Risikomaße in zufallsbehaftete Optimierungsmodelle von besonderer Bedeutung, [28]. Spezielles Augenmerk wird auf solche Risikomaße gelegt, die auch in praxisrelevanten hochdimensionalen Entscheidungsmodellen, z. B. durch Dekomposition, noch numerisch behandelbar bleiben.

Bewegte Horizonte. Ein konzeptionell attraktiver Zugang zur Ressourcenplanung bei sequentiell eingehender Information ist die stochastische Optimierung auf beweglichen Horizonten. Im Ergebnis der Optimierung erhält man implementierbare zeitlich nächste Entscheidungen, bei deren Ermittlung ein gewisser Vorausschauhorizont mit berücksichtigt wurde. Nach Eingang neuer Information werden der Vorausschauhorizont aufdatiert und die Optimierung erneut angestoßen. Die Umsetzung dieses Ansatzes für Probleme realistischer Größenordnung ist eine erstrangige Herausforderung und bedarf grundlegender Untersuchungen zur Struktur und Numerik der entstehenden Optimierungsmodelle. Typischerweise ist dabei die Kopplung mit übergeordneten Planungsebenen zu beachten.

Mensch-Algorithmus-Interaktion. Wie die Einbindung neu eingehender Modellinformation ist auch die Berücksichtigung von Nutzerexpertise im Verlauf der Optimierung ein kritischer Faktor für die „Praxisfestigkeit“ und Akzeptanz mathematisch ermittelter Lösungsvorschläge. Die mathematische Fundierung einer solchen „Mensch-Algorithmus-Interaktion“ und deren methodische Umsetzung sind große Herausforderungen und eröffnen interessante Perspektiven für hochinnovative interdisziplinäre Forschungsvorhaben.

Lineare vs. nichtlineare Modellierung. Oft sind nichtlineare Effekte beim Ressourceneinsatz zu beachten, wie z. B. beim elektrischen Lastfluss oder bei Prozessdynamiken in der Verfahrenstechnik. Die direkte Behandlung solcher Nichtlinearitäten in den typischerweise hochdimensionalen gemischt-ganzzahligen Optimierungsmodellen zur Ressourcenplanung ist oft nicht mehr möglich. Deshalb zielt aktuelle Forschung u.a. auf die Entwicklung algorithmisch vorteilhafter gemischt-ganzzahliger linearer Approximationen für solche Nichtlinearitäten. Spezielle hochdimensionale Optimierungsprobleme mit Nichtlinearität, Stochastik und Ganzzahligkeit werden so rechenbar, [22, 24].

Praxistaugliche Algorithmen und Software. Für diskrete Wahrscheinlichkeitsverteilungen lassen sich zwei- oder mehrstufige stochastische Optimierungsprobleme als hochdimensionale gemischt-ganzzahlige Optimierungsaufgaben mit Blockstruktur schreiben. Selbst fortgeschrittenste kommerzielle Löser für diese Aufgabenklasse scheitern jedoch schnell an der schiereren Größe der aus stochastischen Optimierungsproblemen abgeleiteten Modelle. Dekompositionsverfahren, ein zentrales Thema algorithmischer Forschung, leisten hier Abhilfe. Das Ausgangsproblem wird in einfacher behandelbare Teilaufgaben zerlegt, aus deren Lösungen dann eine (eventuell nur näherungsweise) Lösung des Gesamtproblems generiert wird, [31, 34, 44]. Allerdings gibt es für Dekompositionsverfahren keine „Kochrezepte“. Je nach Spezifik des Problems (zwei- oder mehrstufig, risikoneutral oder risikoavers, ganzzahlige oder stetige Variablen) sind sehr verschiedene mathematische Zugänge zu wählen. Insbesondere für mehrstufige und ganzzahlige Modelle sind noch viele Fragen offen.

Software zur stochastischen Optimierung liegt praktisch nur in Gestalt von Forschungscode vor. Dies betrifft sowohl Dekompositionsverfahren⁵ als auch Modellreduktion durch maßgeschneiderte Abbildung des Zufalls in Szenariobäumen⁶.

Entscheidungsabhängige Verteilungen. Die heute in den oben umrissenen Ansätzen der stochastischen Optimierung strukturell und algorithmisch (manchmal nur ansatzweise) verstandenen Modelle gehen fast ausschließlich davon aus, dass der Zufallseinfluss rein exogen ist. Das heißt, unterliegende Wahrscheinlichkeitsverteilungen hängen nicht von den zu optimierenden Entscheidungen ab. Für eine Reihe von Phänomenen, wie z. B. die Windenergieeinspeisung in ein elektrisches Netz oder den Wärmebedarf eines Versorgungsgebietes, trifft diese Annahme zu. Sie ist aber nicht länger haltbar, wenn z. B. ein Akteur mit Marktmacht am Strommarkt agiert oder verfahrenstechnische Prozesszeiten mit nennenswerten Unsicherheiten behaftet sind oder die Ergiebigkeit von Ölfeldern von der Wahl der ersten Bohrungen, also den Entscheidungsvariablen beeinflusst wird. Die Forschung steckt hier noch in den Kinderschuhen, obgleich das Thema von hoher Praxisrelevanz ist und sich einige vorhandene Resultate, z. B. im Bereich des Stochastic Scheduling [9], als Ansatzpunkte anbieten könnten, [1, 5, 6, 45].

Marktdesign – Gleichgewichtsmodelle. Die Gestaltung transparenter Regeln für die sich ausprägenden liberalisierten Energie- und Rohstoffmärkte ist von großer Bedeutung für eine effiziente Ressourcenplanung in der Energie- und Verfahrenstechnik. Mathematisch spielen dabei Gleichgewichtsbegriffe aus der Spieltheorie und deren Einbindung in Optimierungsmodelle eine wichtige Rolle. Im Fachslang wird von EPECs (Equilibrium Problems with Equilibrium Constraints) und MPECs (Mathematical Programs with Equilibrium Constraints) gesprochen. Als Beispiele aktuellerer Publikationen mit Relevanz für Strommärkte seien [17, 18] genannt. Auch hier sind viele, insbesondere algorithmische, Fragen offen, erst recht wenn es um die Einbeziehung von Stochastik oder ganzzahligen Entscheidungen geht.

5 VISIONEN UND HANDLUNGSEMPFEHLUNGEN

Unsere Vision ist, die Rolle des Produktionsfaktors Mathematik in der Energie- und Verfahrenstechnik so zu stärken, dass den neuen wirtschaftlichen und technologischen Herausforderungen in diesen Bereichen dauerhaft begegnet werden kann. Die Liberalisierung und Globalisierung von Energie- und Rohstoffmärkten sowie die Notwendigkeit res-

⁵ Löser unter GAMS (<http://www.gams.com/solvers/index.htm>), COIN-OR Stochastic Modeling Interface (<https://projects.coin-or.org/Smi>), SLP-IOR (http://www.ior.uzh.ch/Pages/English/Research/StochOpt/index_en.php), MSLiP (<http://myweb.dal.ca/gassmann/MSLIP.html>), ddsip (<http://neos.mcs.anl.gov/neos/solvers/slp:ddsip/LP.html>).

⁶ Szenarioreduktion mit SCENRED (<http://www.gams.com/solvers/solvers.htm#SCENRED>)

sourcenschonenden und umweltverträglichen Wirtschaftens geben dabei entscheidende Rahmenbedingungen vor. Wie im Vorhergehenden dargestellt, leistet die Mathematik bereits jetzt wichtige Beiträge, ist aber noch lange nicht an das Ende ihrer Möglichkeiten gelangt. Realistischerweise ist ein interdisziplinärer „Forschungsmix“ anzustreben, der sich nicht nur auf kurz- bis mittelfristige Aktivitäten mit unmittelbarer Anwendungsrelevanz beschränkt, sondern auch langfristige Vorhaben anwendungsbezogener Grundlagenforschung als zentrale Bestandteile beinhaltet. Energie- und verfahrenstechnische Grundlagenforschung benötigt heute in viel größerem Umfang als früher auch mathematische Grundlagenforschung.

Als beispielhafte Themenbereiche für mittelfristig notwendige Forschungsgebiete seien genannt:

- Optimierungsstrategien für den Betrieb, den Ausbau und die Erneuerung der elektrischen Versorgungsnetze des 21. Jahrhunderts,
- Entscheidungshilfen zur Optimierung des bidirektionalen Informations- und Energiemanagements in innovativen Strukturen der Energieversorgung,
- Problemangepasste Modellierung und effiziente Handhabung von Ungewissheit bei Ressourcenzuteilungsproblemen in der verfahrenstechnischen Produktion,
- Mensch-Maschine-Interaktion bei der Optimierung von Produktionsprozessen.

Aus Sicht der Mathematik kommt dabei besondere Bedeutung dem Ausbau solcher Methoden zur optimalen Ressourcenplanung zu, die nicht nur komplexe technische und wirtschaftliche Zusammenhänge erfassen, sondern in der Lage sind, unvollständige Information über Modelldaten effizient zu integrieren. Insbesondere gilt dies für mehrstufige Entscheidungsprobleme. Hierzu sind wesentliche Fortschritte in der Mathematik nötig, welche sich nur in enger Kooperation mit den Ingenieurwissenschaften, mit Praktikern und IT-Entwicklern erzielen lassen.

Diese Kooperation muss nicht beim Nullpunkt beginnen. Auf Grundlage und mit dem Ziel des Ausbaus bestehender Vernetzungen ließen sich durch das BMBF/BMWi oder die DFG gemeinsam mit der Industrie kurz- bis mittelfristig entsprechende Aktivitäten auf den Weg bringen.

Als erfolgreiche Beispiele in der Vergangenheit seien genannt das DFG-Schwerpunktprogramm 469 *Echtzeitoptimierung großer Systeme* [10] und die erfolgreiche Einbindung mathematischer Forschung im Rahmen der *Netzwerke Grundlagenforschung erneuerbare Energien und rationelle Energieanwendungen*⁷ [35].

In beiden Programmen ist es gelungen, durch die Förderung der Kooperation ingenieurwissenschaftlicher und mathematischer Arbeitsgruppen sowohl die anwendungs-

⁷ <http://www.fz-juelich.de/ptj/netzwerke-grundlagenforschung>

bezogene Forschung als auch die mathematische Grundlagenentwicklung voranzutreiben und langfristige Kooperationen zu etablieren. Dies könnte in einem Schwerpunktprogramm Stochastische Optimierung der DFG ebenso fortgeführt werden wie durch eine bessere Ausstattung entsprechender Förderinitiativen des BMBF.⁸

6 LITERATUR

- [1] E.J. Anderson, A.B. Philpott: Optimal offer construction in electricity markets. *Mathematics of Operations Research* 27 (2002), 82–100.
- [2] J. Balasubramanian, I. Grossmann: Approximation to multistage stochastic optimization in multiperiod batch plant scheduling under demand uncertainty, *Industrial and Engineering Chemistry Research* 43 (2004), 3695–3713.
- [3] J. Brosda: Hierarchische Optimierung für ein zonenübergreifendes Netzengpass-Management, Dissertation, Fakultät für Elektrotechnik und Informationstechnik, Universität Dortmund, 2004.
- [4] S. Engell, A. Märkert, G. Sand, R. Schultz: Aggregated scheduling of a multiproduct batch plant by two-stage stochastic integer programming, *Optimization and Engineering* 5 (2004), 335–359.
- [5] V. Goel, I.E. Grossmann: A Stochastic Programming Approach to Planning of Offshore Gas Field Developments under Uncertainty in Reserves, *Computers & Chemical Engineering* 28 (2004), 1409–1429.
- [6] V. Goel, I. Grossmann: A class of stochastic programs with decision dependent uncertainty, *Mathematical Programming* 108 (2006), 355–394.
- [7] R. Gollmer, M.P. Nowak, W. Römisches, R. Schultz: Unit commitment in power generation – a basic model and some extensions, *Annals of Operations Research* 96 (2000), 167–189.
- [8] M. Göbelt, T. Kasper, C. Sürie: Integrated Short and Midterm Scheduling of Chemical Production Processes – A Case Study. In: *Logistic Optimization of Chemical Production Processes* (S. Engell, Ed.), Wiley-VCH, Weinheim 2008, 239–261.

⁸ Beispielsweise im Förderprogramm *Grundlagenforschung Energie 2020+*, <http://www.fz-juelich.de/ptj/grundlagenforschung-energie/>, indem bei der dort geplanten „Erarbeitung umfassender Optimierungskonzepte“ gezielt *mathematische* Forschungsansätze einbezogen werden.

- [9] A. Grigoriev, M. Sviridenko, M. Uetz: Machine scheduling with resource dependent processing times, *Mathematical Programming* 110 (2007), 209-228.
- [10] M. Grötschel, S.O. Krumke, J. Rambau (Hrsg.): Online Optimization of Large Scale Systems, Springer-Verlag, Berlin, 2001.
- [11] J. Handke: Koordinierte lang- und kurzfristige Kraftwerkseinsatzplanung in thermischen Systemen mit Pumpspeicherkraftwerken, Dissertation, Univ. Dortmund, 1994.
- [12] E. Handschin, F. Neise, H. Neumann, R. Schultz: Optimal operation of dispersed generation under uncertainty using mathematical programming, *International Journal of Electrical Power & Energy Systems* 28 (2006), 618-626.
- [13] I. Harjunkski, G. Beykirch, M. Zuber, H.J. Weidemann: The process "copper": copper plant scheduling and optimization. *ABB Review* 4 (2005), 51-4.
- [14] I. Harjunkski, M. Fahl, H.-W. Borchers: Scheduling and Optimization of a Copper Production Process. In: Logistic Optimization of Chemical Production Processes (S. Engell, Ed.), Wiley-VCH, Weinheim, 2008, 93-109.
- [15] E. Hauptmeier: KWK-Erzeugungsanlagen in zukünftigen Verteilungsnetzen – Potential und Analysen, Dissertation, Fakultät für Elektrotechnik und Informationstechnik, Universität Dortmund, 2007.
- [16] H. Heitsch, W. Römisch: Scenario tree modeling for multistage stochastic programs, *Mathematical Programming* (erscheint), Online First DOI 10.1007/s10107-007-0197-2.
- [17] B.F. Hobbs, J.S. Pang: Nash-Cournot equilibria in electric power markets with piecewise linear demand functions and joint constraints, *Operations Research* 55 (2007), 113-127.
- [18] X. Hu, D. Ralph: Using EPECs to model bilevel games in restructured electricity markets with locational prices, *Operations Research* 55 (2007), 809-827.
- [19] ILOG Inc., Mountain View, CA, USA.
- [20] I. Jürgens: Langfristoptimierung für das risikoorientierte Asset Management von elektrischen Energieversorgungssystemen, Dissertation, Fakultät für Elektrotechnik und Informationstechnik, Universität Dortmund, 2007.

- [21] S. Lee, I.E. Grossmann: New algorithms for generalized disjunctive programming. *Computers & Chemical Engineering* 24 (2000), 2125–2141.
- [22] A. Martin, M. Möller, S. Moritz: Mixed integer models for the stationary case of gas network optimization, *Mathematical Programming* 105 (2006), 563–582.
- [23] C.A. Mendez, I. E. Grossmann, I. Harjunkski: MILP Optimization Models for Short-term Scheduling of Batch Processes. In: *Logistic Optimization of Chemical Production Processes* (S. Engell, Ed.), Wiley-VHC, Weinheim, 2008, 163–184.
- [24] S. Moritz: A mixed integer approach for the transient case of gas network optimization, Dissertation, Fachbereich Mathematik, TU Darmstadt, 2006.
- [25] H. Neumann: Zweistufige stochastische Betriebsoptimierung eines virtuellen Kraftwerks, Dissertation, Fakultät für Elektrotechnik und Informationstechnik, Universität Dortmund, Shaker Verlag, Aachen, 2007.
- [26] M.P. Nowak, W. Römisch: Stochastic Lagrangian relaxation applied to power scheduling in a hydro-thermal system under uncertainty, *Annals of Operations Research* 100 (2000), 251–272.
- [27] M.V.F. Pereira, L.M.V.G. Pinto: Multi-stage stochastic optimization applied to energy planning, *Mathematical Programming* 52 (1991), 359–375.
- [28] G. Ch. Pflug, W. Römisch: *Modeling. Measuring and Managing Risk*, World Scientific, Singapore, 2007.
- [29] C. Plapp, D. Surholt, D. Syring: Planning Large Supply Chain Scenarios with „Quant-based Combinatorial Optimization“. In: *Logistic Optimization of Chemical Production Processes* (S. Engell, Ed.), Wiley-VCH, Weinheim 2008, 59–91.
- [30] C. Rehtanz: *Autonomous Systems and Intelligent Agents in Power Systems and Control and Operation*, Springer-Verlag, 2003.
- [31] A. Ruszczyński, A. Shapiro (eds.): *Handbooks in Operations Research and Management Science*, 10: *Stochastic Programming*, Elsevier, Amsterdam, 2003.
- [32] G. Sand: Engineered Mixed-Integer Programming in Chemical Batch Scheduling. In: *Logistic Optimization of Chemical Production Processes* (S. Engell, Ed.), Wiley-VHC, Weinheim, 2008, 137–161.

- [33] N. Schnurr: Potential multifunktionaler FACTS-Geräte zur Erhöhung von Übertragungskapazität und Mittelzeitstabilität im Europäischen Verbundnetz, Dissertation, Univ. Dortmund, VDE Verlag, 2004.
- [34] R. Schultz: Stochastic programming with integer variables, *Mathematical Programming* 97 (2003), 285–309.
- [35] R. Schultz, H.-J. Wagner (Hrsg.): Innovative Modellierung und Optimierung von Energiesystemen, LIT Verlag, Münster, 2008.
- [36] W. Schulz: Strategien zur effizienten Integration der Windenergie in den deutschen Elektrizitätsmarkt, Dissertation, Fakultät für Elektrotechnik und Informationstechnik, Universität Dortmund, 2007.
- [37] M. Schulz, S. Spieckermann: Logistics Simulation in the Chemical Industry. In: Logistic Optimization of Chemical Production Processes (S. Engell, Ed.), Wiley-VHC, Weinheim, 2008, 21–36.
- [38] F.C. Schuppe: Uncertain Dynamic Systems, Prentice-Hall, 1973.
- [39] G. Spangardt: Mittelfristige risikoorientierte Optimierung von Strombezugsportfolios, Dissertation, Univ. Dortmund, Fraunhofer IRB Verlag, 2003.
- [40] M. Stobbe: Supply Chain and Supply Chain Management. In: Logistic Optimization of Chemical Production Processes (S. Engell, Ed.), Wiley-VHC, Weinheim, 2008, 3–18.
- [41] S. Takriti, J.R. Birge, E. Long: A stochastic model for the unit commitment problem, *IEEE Transactions on Power Systems* 11 (1996), 1497–1508.
- [42] S. Takriti, B. Krasenbrink, L.S.-Y. Wu: Incorporating fuel constraints and electricity spot prices into the stochastic unit commitment problem, *Operations Research* 48 (2000), 268–280.
- [43] J. Till: New hybrid evolutionary algorithms for chemical batch scheduling under uncertainty, Dissertation, Fachbereich Bio- und Chemieingenieurwesen, Universität Dortmund, Shaker Verlag, Aachen, 2007.
- [44] J. Till, G. Sand, M. Urselmann, S. Engell: A Hybrid Evolutionary Algorithm for Solving Two-stage Stochastic Integer Programs in Chemical Batch Scheduling, *Computers & Chemical Engineering* 31 (2007), 630–647.

- [45] S.A. Van den Heever, I.E. Grossmann: An Iterative Aggregation/Disaggregation Approach for the Solution of a Mixed Integer Nonlinear Oilfield Infrastructure Planning Model, *Industrial and Engineering Chemistry Research* 39 (2000), 1955-1971.

> SIMULATIONSBASIERTE OPTIMIERUNG IN DER BAUTECHNIK – NEUE INFORMATISCHE LÖSUNGSKONZEPTE

DIETRICH HARTMANN, MATTHIAS BAITSCH UND VAN VINH NGUYEN

1 EXECUTIVE SUMMARY

Ingenieure sind – aufbauend auf gründlicher Planung, intelligentem Entwurf, solider Konstruktion und sorgfältiger Herstellung – seit jeher bestrebt, „optimale“ technische Systeme zu entwickeln. Auch die Prozesse, die von Ingenieuren eingeleitet und verantwortet werden, sind in der Regel so ausgelegt, dass sie planmäßig, zuverlässig, hochgradig effizient und somit ebenfalls möglichst optimal ablaufen. Bei der Umsetzung ihrer Ideen mithilfe von leistungsfähigen Computermodellen sowie gebrauchstauglicher Software sorgt die zielgerichtete Vernetzung mit der Forschung in Mathematik und Angewandter Informatik dafür, dass sowohl dringende aktuelle Problemstellungen gelöst als auch neue Zukunftsfelder in absehbarer Zeit erfolgreich bearbeitet werden können.

Die Ganzheitlichkeitsaspekte berücksichtigende Optimierung technischer Systeme und Prozesse gewinnt dabei in letzter Zeit immer mehr an Bedeutung! Eine ganzheitliche Sicht bei der Optimierung ist – aus verschiedenen Gründen, etwa im Hinblick auf den Energieverbrauch, die Umweltverträglichkeit, die Einsparung von Kosten unterschiedlichster Art oder die Effizienz im Allgemeinen – von höchstem Interesse für die Volkswirtschaft. Letztendlich entscheiden optimale technische Lösungen über das Wohlergehen, den Wohlstand der Bevölkerung im Lande und deren Lebensqualität. Dies gilt insbesondere auch für die technischen Systeme und Prozesse der Bautechnik bzw. der gebauten Umwelt, womit das gesamte Bauwesen und die Umwelttechnik (Civil, Architectural and Environmental Engineering) eingeschlossen sind.

Die bewusste, alle wesentlichen Gegebenheiten der Realität erfassende und somit ganzheitliche Betrachtungsweise ist erst möglich geworden, seitdem man über leistungsstarke Computersysteme, insbesondere Hochleistungsrechner (High Performance Computing) oder in großer Zahl über Computernetze miteinander verbundener Arbeitsplatzrechner (Distributed Computing bzw. Grid Computing) verfügt. Dass sich die Leistungsfähigkeit der Computer durch fortschreitende Miniaturisierung und verfeinerte Herstellungstechniken ständig verbessert, aber auch die Gebrauchstauglichkeit unserer Computernetze von Jahr zu Jahr erhöht, wird in den Technikwissenschaften nur zu gerne dazu

genutzt, um auch große schwierige Problemstellungen (grand challenge problems), wie sie die ganzheitliche bzw. holistische Optimierung technischer Systeme und/oder Prozesse in Raum und Zeit darstellt, in Angriff zu nehmen. Eine wesentliche Lösungsgrundlage bildet die numerische Simulation über realitätsnahe Computermodelle. Die numerische Simulation hat sich, wie schon seit Jahren bekannt, als die dritte, wesentliche Säule des Erkenntnisgewinns erwiesen, neben Theorie und Experiment. In den Technikwissenschaften jedenfalls ist die numerische Simulation, aufbauend auf mathematisch fundierten Modellen, ein unerlässliches Werkzeug geworden, um eine Vielzahl anstehender komplizierter Probleme zu lösen; sie kann sogar dazu dienen, im Computer vorab neue Theorien hinsichtlich ihrer Gültigkeit bzw. Anwendbarkeit zu untersuchen und dadurch teure Experimente obsolet zu machen.

Simulationsbasiertes Engineering erlaubt es insbesondere, die Konsequenzen (Auswirkungen, Gefahren, Vorteile, Nachteile) bei vorgenommenen Veränderungen oder Eingriffen in Systemen oder Prozessen sowohl quantitativ als auch qualitativ vorherzusagen. Damit ist die numerische Simulation die treibende Kraft und der Motor, um auch Ingenieursysteme und -prozesse kontrolliert und kontrollierbar zu optimieren. Die simulationsbasierte Optimierung ist zwar einerseits vielversprechend, andererseits aber eine echte Herausforderung, da die ihr unterlegten Simulationen im Normalfall mehrfach – je nach Problem mehrere hundert oder sogar tausend Mal und mehr – zu wiederholen sind, bevor ein (möglichst globales) Optimum bestimmt werden kann. Um dies zu gewährleisten, muss der Optimierungsablauf bzw. die Optimierungsmethodik immer eng mit der Simulation verzahnt werden.

Eine weitere gravierende Schwierigkeit besteht darin, dass eine simulationsbasierte Optimierung nahezu durchgängig auf Nichtstandardoptimierungsprobleme führt, die durch unvorhersehbare Nichtlinearität, hohe Dimensionalität, viele lokale Optima und meistens nicht glatte Zulässigkeitsbereiche gekennzeichnet sind. Vielfach sind die als Optimierungskriterium formulierten „Zielfunktionen“ bzw. die bei der Optimierung einzuhaltenden „Restriktionsfunktionen“ keine Funktionen im klassischen mathematischen Sinn, sie lassen sich oft nur in Algorithmusform beschreiben. Kommt noch die zeitliche Veränderung bestimmter Optimierungsgrößen hinzu, oder sind Unschärfen bzw. stochastische Eigenschaften im Simulationsmodell zu berücksichtigen, verschärfen sich die ohnehin schon vorhandenen Schwierigkeiten nochmals. Nur durch neuartige mathematische und informatische Lösungsansätze können dann die sich auftürmenden Komplexitätsbarrieren bewältigt werden. Hierzu werden im vorliegenden Beitrag anhand von konkreten Anwendungsbeispielen aus der eigenen Forschung im Bereich der Bautechnik Vorschläge zur richtigen Vorgehensweise unterbreitet und Handlungsvorschläge für ähnlich geartete Aufgabenbereiche abgeleitet.

2 ERFOLGSGESCHICHTEN

Mehrere, in den letzten Jahren erfolgreich durchgeführte Forschungsprojekte, vor allem auch in regionalen bzw. überregionalen Verbänden, wie etwa von der Deutschen Forschungsgemeinschaft (DFG) geförderten Graduiertenkollegs, Forschergruppen, Schwerpunktprogrammen sowie Sonderforschungsbereichen, sind Beleg dafür, dass aus den Verflechtungen zwischen der Mathematik und den Technikwissenschaften zahlreiche gewinnbringende Resultate hervorgegangen sind. Da technische Lösungen heute ohne den Einsatz von Computersystemen und ohne ein entsprechendes „Computing“ undenkbar sind, ist auch die Informatik bzw. die Ingenieurinformatik – in Wechselwirkung mit der Mathematik und den Technikwissenschaften – entscheidend an diesem Erfolg beteiligt.

Strukturoptimierung großer Tragstrukturen. Die optimale Auslegung von großen Tragwerken (Strukturoptimierung) im Hinblick auf bestimmte Eigenschaften (z. B. Gewicht, Kosten, ein bestimmtes physikalisches Verhalten, Zuverlässigkeit etc.) unter Berücksichtigung einer Vielzahl von Restriktionen (z. B. Nebenbedingungen bezüglich zulässiger Spannungen, Verformungen oder konstruktiver Vorgaben) führt in der Regel auf Softwaremodelle, die leistungsstarke, iterativ arbeitende Optimierungsmethoden erfordern. Bedingt durch die oben erwähnte hohe Dimensionalität und Nichtlinearität des Optimierungsproblems entsteht ein enorm großer Rechen- und Datenverarbeitungsaufwand, sofern man „nur“ sequentiell arbeitende von-Neumann-Rechnerarchitekturen, d. h. normale Arbeitsplatzrechner oder Personal Computer, einsetzen kann. Diesen Nachteil gilt es zu überwinden, da das computergestützte Optimieren unschätzbare Vorteile hat: Die arbeits- und personalintensive Dimensionierung von Tragsystemen, die ihrerseits heute selbstverständlich ebenfalls mit Computermethoden (meistens Finite-Elemente-Methoden) durchgeführt wird, und auch die Konstruktion mithilfe von CAD-Software, kann durch Einsatz von Optimierungsstrategien noch kostensparender durchgeführt werden, da sich das Automatisierungspotential erhöht. Aus diesem Grunde lohnt es sich folglich, alles zu unternehmen, um die bei der Optimierung großer Tragsysteme anfallenden Rechenzeiten, und auch den erhöhten Speicherbedarf, drastisch zu reduzieren. Dies gelingt durch konsequente Parallelisierung des Strukturoptimierungsproblems über geeignete Dekomposition bzw. Partitionierung des Ausgangsproblems in mehrere Subprobleme, die dann zur Abarbeitung auf die Prozessoren eines Parallelrechnersystems verteilt werden.

Erste, wegweisende Lösungsansätze wurden bereits in der DFG-Forschergruppe „Optimierung großer Systeme“ [12] erarbeitet: Das entstehende Dekompositionsproblem, ob und wie man effizienzsteigernd überlappende oder nicht-überlappende Subprobleme erzeugt, und auch die algorithmische Umsetzung der parallelen Optimierung, konnte dabei nur in engem Zusammenwirken zwischen Mathematik, Informatik und Strukturmechanik gelöst werden. Hierzu das nachfolgende Beispiel: Für eine Stadionüberdachung wurde eine Dekomposition in vier nicht überlappende Substrukturen gewählt (siehe Abb. 1), die dann auf unterschiedlichen Prozessoren für unterschiedliche Lastfälle im Kontext zueinander optimiert wurden.

Gestützt auf Dekomposition, Partitionierung und Parallelisierung konnten Tragwerksentwürfe ausfindig gemacht werden, die auf konventionelle Weise – ohne parallele Optimierung – so nicht gefunden worden wären. Das Ergebnis der kooperativen parallelen Strukturoptimierung ist für ein Viertel der Stadionüberdachung in Abb. 2 dargestellt, aus der auch die Abnahme der Optimierungszielfunktion (Gewicht) über die Anzahl der erforderlichen Iterationen bis zum Optimum zu entnehmen ist.

Diskrete bzw. gemischte, kontinuierlich-diskrete Strukturoptimierung. Bei der diskreten bzw. gemischt kontinuierlich-diskreten Strukturoptimierung führt man neben kontinuierlichen, auch diskrete Tragwerksparameter (Optimierungsvariable) ein, die nur ganz bestimmte diskrete Werte annehmen dürfen. Ein typisches Beispiel sind geometrische Abmessungen von standardisierten Querschnittsprofilen (HE-A, HEB, HE-IPB etc.). Für diese Kategorie von Optimierungsproblemen können ebenfalls schon zahlreiche Forschungs-

Abb. 1: 1660-Stab-Raumfachwerk einer Stadionüberdachung

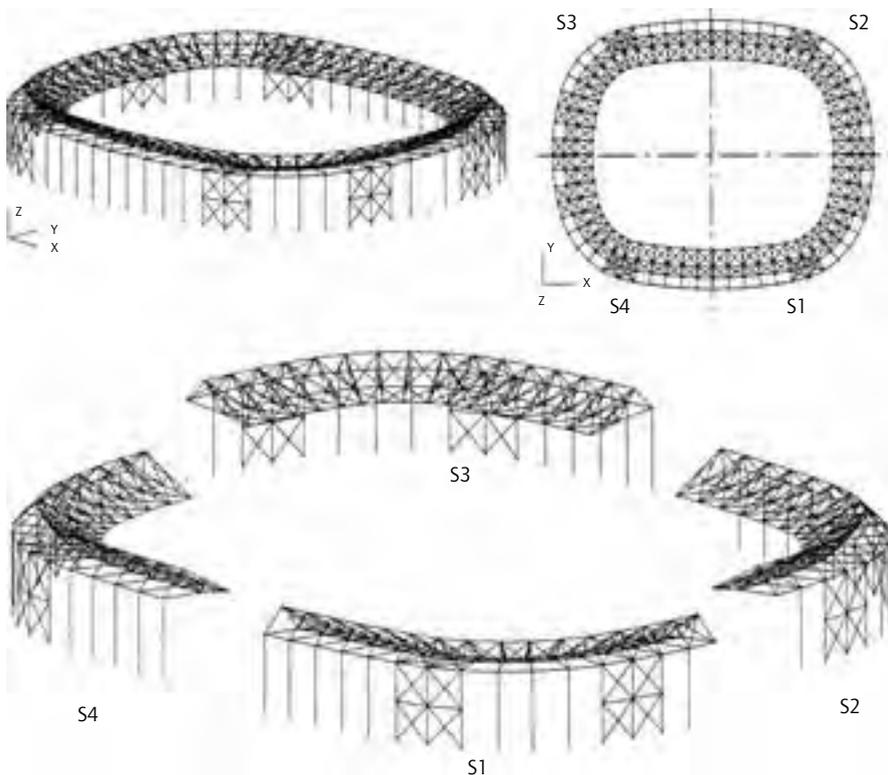
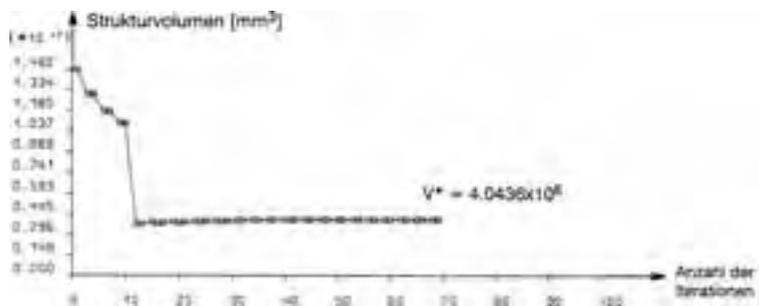
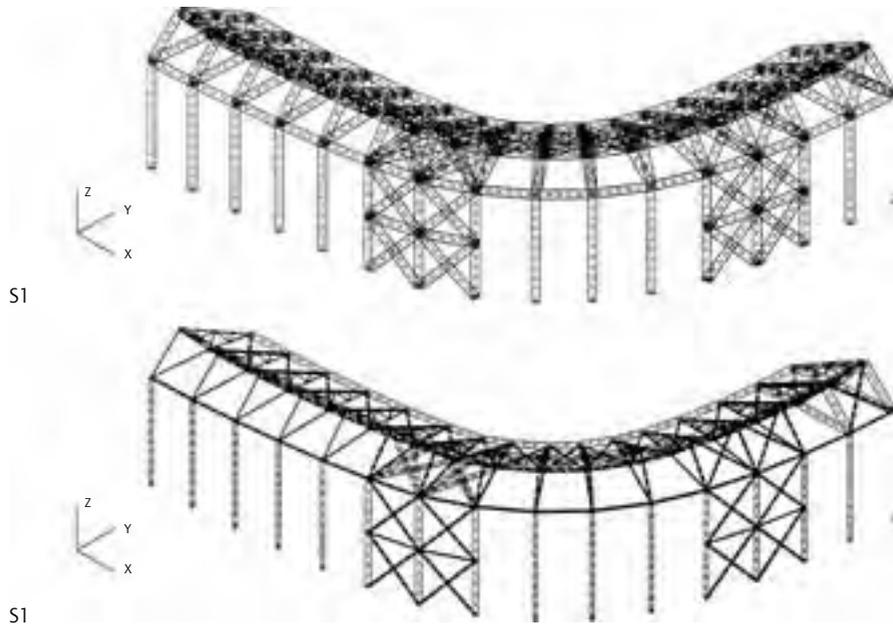


Abb. 2: Räumliches 415-Stab-Substrukturmodell



erfolge mit großem praktischen Nutzen vorgewiesen werden; sie alle belegen, dass die Verzahnung von Mathematik und Technikwissenschaften die beste Gewähr dafür ist, schwierige technische Probleme zu lösen und – bezogen auf Aufgabenstellungen im Bereich der Tragwerksplanung – bislang nicht bekannte Lösungen ausfindig zu machen.

Viele der in der konventionellen numerischen Optimierung populär gewordenen Methoden der Mathematischen Programmierung (ableitungsorientierte Verfahren, wie Newton- und Gradientenmethoden) werden allerdings nicht den Anforderungen der technischen Optimierung mit diskreten bzw. gemischt diskret-kontinuierlichen Optimierungsvariablen gerecht. Die bekannten Methoden liefern lediglich grobe Approximationen der mathematisch korrekten Lösung oder verlangen vielfach tiefe Eingriffe in den Lösungsablauf der verwendeten Verfahren, gegebenenfalls ist sogar die Re-Modellierung des Optimierungsproblems erforderlich.

Um diesen Defiziten entgegenzuwirken, werden bei der Erforschung von gemischt diskreten Optimierungsproblemen – abgesichert durch mathematische Theorie bezüglich Eindeutigkeit, Stabilitäts- und Divergenzverhalten – neben speziell zugeschnittenen Algorithmen seit geraumer Zeit erfolgreich sogenannte Evolutionäre Algorithmen (EA) [1] eingesetzt, in den eigenen Forschungsprojekten zur Strukturoptimierung insbesondere die Familie der Evolutionsstrategien (ES) [5].

Evolutionsstrategien (ES) sind robuste, gut handhabbare, stochastische Suchverfahren, die nach dem Vorbild der biologischen Evolution auf einem kollektiven Lernprozess innerhalb einer Population von simultan existierenden Individuen basieren. Biologische Populationen zeigen nachweislich die Eigenschaft, dass sie im Generationsablauf „evolviert“, aufbauend auf Mutations-, Rekombinations- und Selektionsprozessen, die sämtlich effiziente Optimierungsmechanismen darstellen und eine Selbstoptimierung des Populationsgeschehens auslösen. Die konsequente Übertragung derartiger biologischer Optimierungsprinzipien (wie Mutation, Rekombination, Selektion, Lethalität etc.) in den mathematisch-technischen Raum eröffnet völlig neue, durch klassische mathematische Methoden so nicht realisierbare Perspektiven. Vor allem die Mutation und die Rekombination sorgen dabei für die im technischen Bereich notwendige Variantenvielfalt, die dann – unter Verzicht auf Ableitungsinformationen erster bzw. zweiter Ordnung – durch geeignete Selektion dem Verlauf der Suche nach dem (globalen) Optimum eine Richtung gibt („Diffusion“ der Individuen in Richtung eines Optimums).

Neben der Kopplung von strukturmechanischen Simulationsmodellen mit numerischer Optimierung über ES spielt dabei auch ein neuer, der „Künstlichen Intelligenz Forschung“ entspringender, informatischer Lösungsansatz eine Rolle: Die Methode der Objektorientierung bzw. die objektorientierte Modellierung, die vorhandene Synergieeffekte zwischen Mathematik und Technikwissenschaften um ein Weiteres verstärkt. Aufbauend auf ein prägnantes, präzises und zugleich anschauliches Beschreibungskalkül (Unified Modeling Language oder UML) mit inzwischen weltweit anerkannter Symbolik und Notation, lassen sich selbst hochgradig komplexe, numerische sowie nichtnumerische Problemstellungen der Technikwissenschaften mit den in der Objektorientie-

rung verwendeten Abstraktionsprinzipien (Klassen/Modularisierungs-, Geheimnis-, Vererbungsprinzipien etc.) informatisch korrekt modellieren. Die Semantik der Objekte bzw. Klassen (ausgedrückt durch Eigenschaften, Verhalten, Assoziationen/Relationen, Mengeneigenschaften) orientieren sich dabei an den Begriffen und Konzepten der realen Welt, d. h. der jeweils zugrunde liegenden technischen Problemstellung. Auf dem Weg von der konzeptuellen Modellierung eines „Weltproblems“ bis hin zu seiner Implementierung in Software auf Computern entstehen somit weitaus weniger Informationsverluste als bei anderen Software-Entwicklungsmethoden.

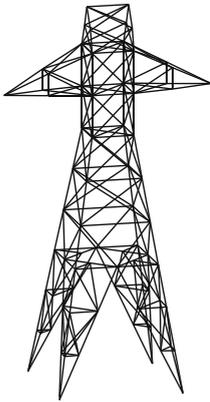
Abb. 3 zeigt Resultate eigener Forschungsarbeiten aus dem Bereich der diskreten bzw. gemischt kontinuierlich-diskreten Strukturoptimierung (siehe bspw. [10]). Hervorzuheben ist, dass sich die gefundenen Resultate wiederum gravierend von bisherigen Lösungen unterscheiden, die bei Vernachlässigung bzw. ohne Berücksichtigung der diskreten Problemstruktur ermittelt wurden.

Lebensdauerorientierte Entwurfskonzepte. Jeder, der mit seinem Auto über eine Brücke fährt, oder sich zum Beispiel in einer Industriehalle mit Lastkränen aufhält, verlässt sich selbstverständlich darauf, dass „nichts zusammenbricht“! Selbst bei betagten Bauwerken verlassen wir uns dabei auf die Stabilität, Haltbarkeit und Zuverlässigkeit der Tragwerke und Tragsysteme, die von Statikern und Konstrukteuren „eingerechnet“ wurden. Statik und Dynamik, Betriebsfestigkeitsanalysen und auch stochastische Effekte sowie Unschärfe erfassende Zuverlässigkeitsberechnungen haben zweifelsohne – zusätzlich gereift durch experimentelle Untersuchungen und die Praxis des Konstruierens – über die letzten Jahrzehnte hinweg einen hohen wissenschaftlichen Stand erreicht. Man braucht jedoch computergestützte Entwurfssysteme und Konstruktionsleitsysteme, mit denen das Tragwerksverhalten – einschließlich der möglicherweise eintretenden Schädigungseinträge – über die Lebens/Nutzungsdauer hinweg simuliert werden kann und die in Bezug auf ein charakteristisches Optimierungskriterium die beste („optimale“) Auslegung eines Tragwerks liefern; dies gilt insbesondere für komplexe, große Tragwerke.

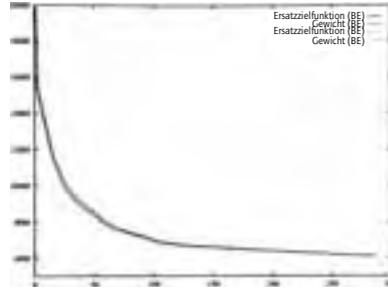
Die Lösung des lebensdauerorientierten Entwurfs- und Konstruktionsproblems unter Berücksichtigung von zufallsbedingten Schädigungen (Deteriorationen) stellt sich allerdings als ein äußerst schwieriges Problem dar: Formal, d. h. aus mathematischer Sicht, lässt sich dieses Problem für den Bezugszeitraum der Lebensdauer in ein äquivalentes, stochastisches Optimierungsproblem mit verschiedenen Raum- und Zeitskalen überführen. Es entsteht dann jedoch ein sowohl mechanisch als auch bezüglich der Optimierung hochgradig nichtlineares Mehrebenen- und Mehrskalproblem (aus Optimierungssicht ein semi-infinites stochastisches Optimierungsproblem). Die Lösung dieses aufwändigen Problems kann nur durch simulationsbasiertes Vorgehen unter gleichzeitiger Verwendung von Hochleistungscomputern erfolgen.

Im Rahmen des Bochumer Sonderforschungsbereichs SFB 398 „Lebensdauerorientierte Entwurfskonzepte“ [18] konnten für diesen „neuen“ Aufgabentyp durch die gezielte Bündelung mathematischer, informatischer und strukturmechanischer Lösungsideen wichtige Lösungswege aufgezeigt werden: Prototypisch behandelt

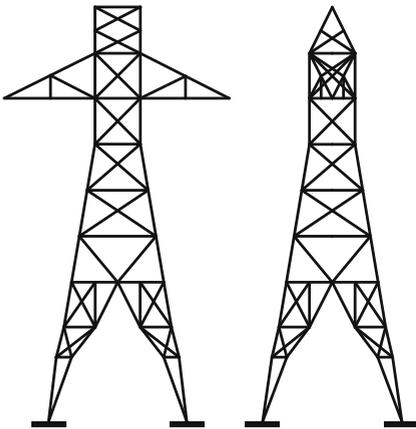
Abb. 3: Optimierung eines Freileitungsmastes mit diskreten Variablen



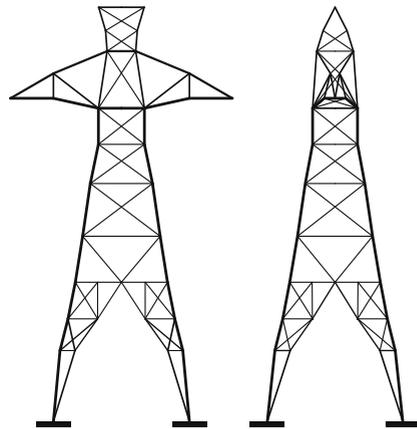
Räumliche Darstellung



Reduktion des Gewichts in der Optimierung



Ausgangsstruktur



Optimierte Struktur

wurden dabei in einem eigenen, zwölf Jahre von der DFG geförderten Forschungsprojekt vor allem Industriehallen (siehe Abb. 4) und Stabbogenbrücken aus Stahl (siehe Abb. 5).

Die Anschlussbleche moderner Stabbogenbrücken, die Hänger und Hauptträger verbinden, sind hochgradig schädigungssensitive Bauteile. Insbesondere windinduzierte Querschwingungen der Hänger führten über so genannte Lock-In-Effekte in der Vergangenheit mehrfach zu Rissen in einigen Anschlussblechen. Im Rahmen des Bochumer SFB 398 wurden derartige Brücken mit einem mehrschichtigen Simulationsmodell im Computer nachgebildet und die Form der Anschlussbleche mit Hilfe von Optimierungsverfahren im Hinblick auf die Lebensdauer verbessert (siehe Abb. 5). Die Simulation der Spannungen im Anschlussblech, die als Grundlage für die Lebensdauerermittlung dienen, erfolgt mit Hilfe der Finite-Elemente-Methode (FEM). Mit der Optimierung wird die Form des Bleches so verbessert, dass die rechnerische Lebensdauer um Größenordnungen gesteigert werden kann (siehe [8]).

Formoptimierung mit der p-version der FEM. Jeder simulationsbasierten Optimierung liegt ein Computermodell zugrunde, mit dem die Auswirkungen der von der Optimierungsstrategie verursachten Änderungen der Entwurfsvariablen ermittelt (simuliert) werden. Für die überwiegende Zahl von Simulationsaufgaben basiert dieses Computermodell auf der Finite-Elemente-Methode (FEM), mit der die gesuchten physikalischen Feldgrößen, z. B. Verformungen, Temperaturverteilung etc., aus Funktionen approximiert werden, die stückweise auf dem geeignet unterteilten Gebiet (dem Finiten Element) definiert sind. Besonders geeignet für die Formoptimierung kontinuierlicher Strukturen ist die p -Version der FEM, in der die Struktur in vergleichsweise große Elemente unterteilt werden kann, wobei diese Unterteilung – gegenüber anderen Versionen der FEM – im Verlauf des Optimierungsprozesses nicht zwingend geändert werden muss (siehe

Abb. 4: Lebensdauerorientierte Optimierung eines Hallenrahmens

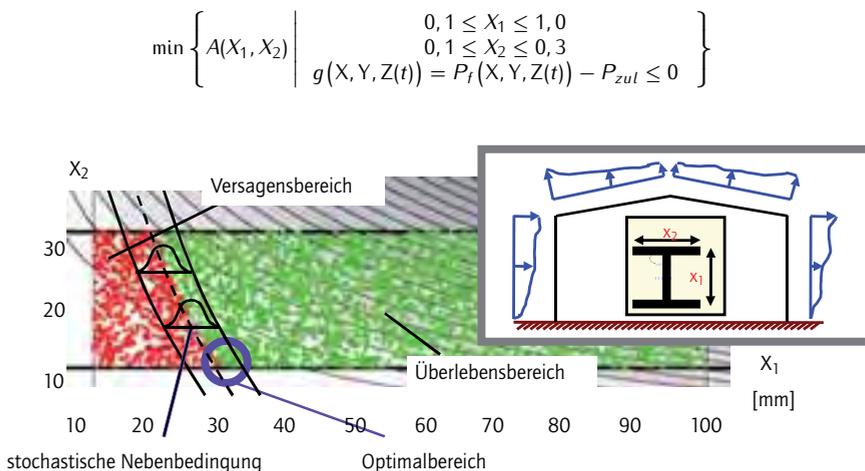


Abb. 6). Der mathematisch-konsistenten Formulierung der p -Version der FEM kommt dabei außerordentlich große Bedeutung zu: Hiermit wird die objektorientierte Implementierung in hoch effiziente Software unterstützt.

Ein Anwendungsbeispiel aus [3] wird in Abb. 7 behandelt: Die Optimierung der Form eines Staudammes im Hinblick auf minimalen Materialaufwand; gleichzeitig dürfen die im Stahlbeton auftretenden Spannungen die zulässige Grenze nicht überschreiten. Mit der Optimierung wurde das Volumen des verbauten Materials um 30 % reduziert. Das Tragwerksmodell basiert dabei auf der p -Version der FEM, wobei verschiedene Techniken des Computer Aided Geometric Design (CAGD) zum Einsatz kommen. Um die Rechenzeit zu reduzieren, wurde die Optimierung parallel auf einem Cluster Computer durchgeführt, der aus 56 Einzelrechnern mit je zwei Prozessoren besteht, die mit einer besonders leistungsfähigen Vernetzung ausgestattet sind. Sowohl die Kopplung der FEM mit Methoden aus dem CAGD als auch die Parallelisierung wurde durch den Einsatz objektorientierter Softwaremethoden erheblich erleichtert, wenn nicht sogar erst möglich gemacht.

Abb. 5: Lebensdauerorientierte Optimierung des Hängeranschlussblechs einer Stabbogenbrücke
 Oben: Zur Schädigung des Anschlussblechs führende Wirkungskette: Der stochastische Lastprozess (Wind) führt zu hochfrequenten Schwingungen der runden Hänger, wodurch im Anschlussblech Risse hervorgerufen werden.
 Unten: Optimierung des Anschlussblechs: Dargestellt sind die initiale und die optimierte Form des Blechs jeweils in einer Draufsicht und mit dem dreidimensional dargestellten Beanspruchungszustand. Deutlich zu sehen ist die Reduktion der in der initialen Form erkennbaren Spannungsspitze.

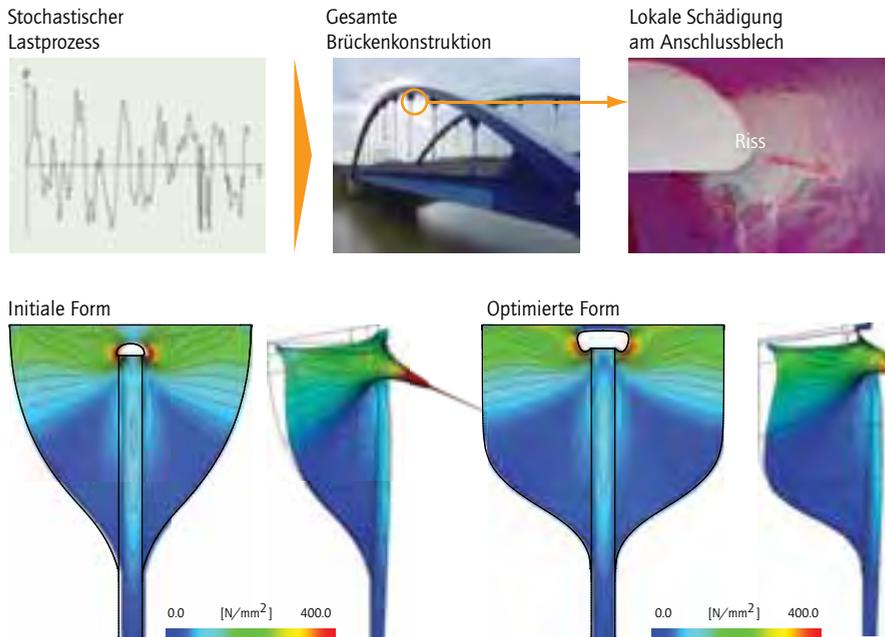


Abb. 6: Optimierung kontinuierlicher Strukturen

Vergleich der h -Version und der p -Version der FEM für ein Formoptimierungsproblem: In der h -Version muss nach jeder Änderung der Geometrie ein neues Netz erzeugt werden. Demgegenüber kann mit der p -Version das Netz beibehalten werden.

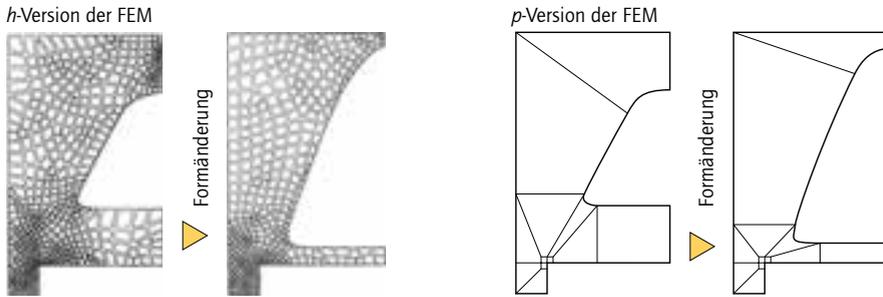
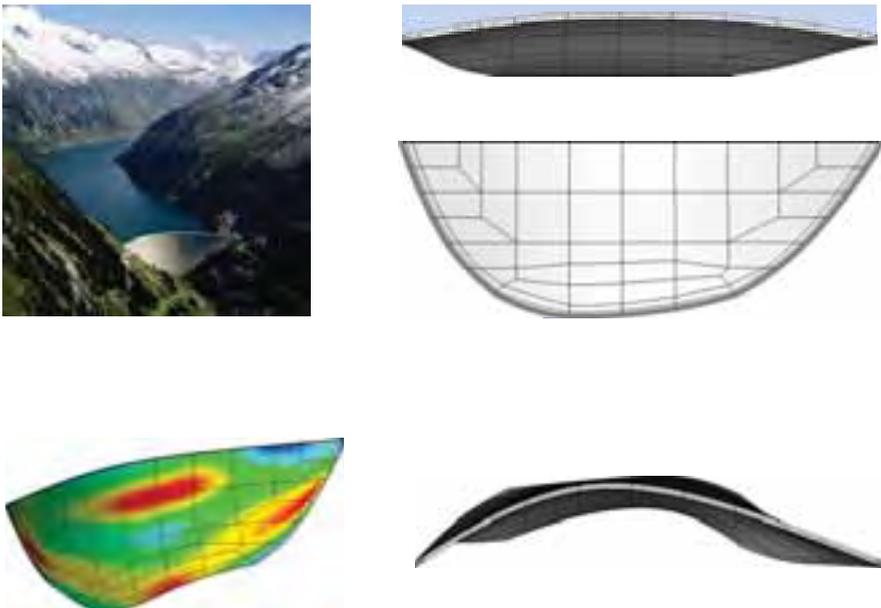


Abb. 7: Optimierung einer Bogenstaumauer

Oben: Bogenstaumauer des Zillergründl-Stausees. Rechts dargestellt ist das Finite-Element-Modell eines vergleichbaren Bauwerks.

Unten: Optimierung der Bogenstaumauer: Links, die simulierte Materialbeanspruchung als Grundlage für die Formoptimierung und rechts die optimierte Form des Tragwerks.

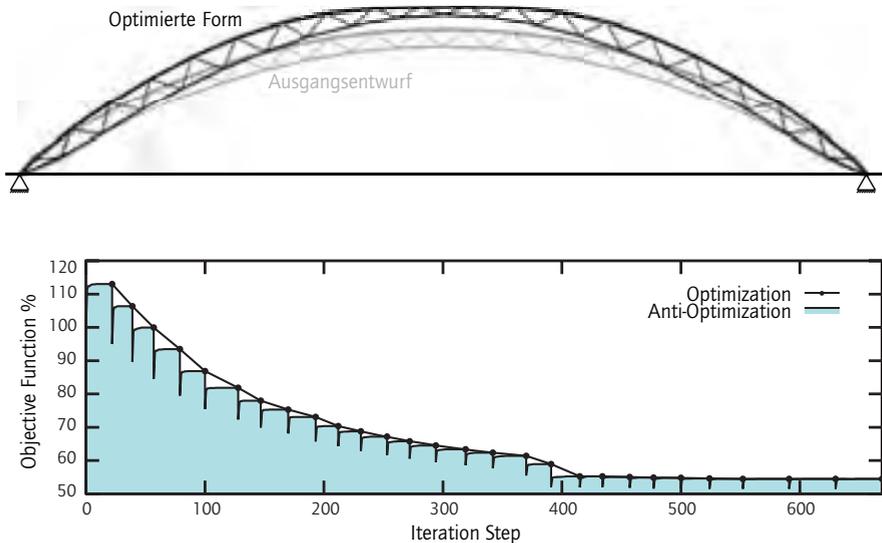


Robuste Optimierung druckbeanspruchter Tragwerke. In der Strukturoptimierung werden die physikalischen Eigenschaften der zu optimierenden Struktur über ein mathematisches Modell (Computermodell) simuliert. Das eingesetzte Optimierungsverfahren findet somit Strukturkonfigurationen, die immer nur im Rahmen des verwendeten Modells „optimal“ sein können und durch rigorose „Ausnutzung“ der Möglichkeiten der Optimierungsstrategie eine „spezialisierte“ Lösung darstellen. Entspricht das später tatsächlich realisierte Tragwerk, etwa infolge Fertigungsungenauigkeiten (Imperfektionen), nicht den im Modell getroffenen Annahmen und Voraussetzungen, wird das wirkliche Tragverhalten von dem durch Optimierung prognostizierten Verhalten abweichen [19]. Da die Lösung des Optimierungsproblems in der Regel zudem auf dem Rand des zulässigen Gebiets liegt, so dass der Entwurf gerade noch zulässig ist, können selbst geringfügigste Abweichungen von der gefundenen „optimalen“ Lösung die Sicherheit optimierter Baukonstruktionen in Frage stellen. Besonders gefährdet sind schlanke, druckbeanspruchte Stahlkonstruktionen, die schon bei kleinsten Änderungen in der Tragwerksgeometrie sehr große Abweichungen im Strukturverhalten nach sich ziehen, also imperfektionsempfindlich sind. Für den Entwurf derartiger Strukturen ist deshalb eine entsprechende Systematik erforderlich, bei der die Unempfindlichkeit gegenüber Imperfektionen in geeigneter Weise als Entwurfsziel berücksichtigt wird.

In der eigenen Forschung [2] wurden deshalb geometrische Imperfektionen gezielt als Bestandteil des mathematischen Tragwerksmodells mit in die Optimierung einbezogen. Das verwendete Optimierungsmodell setzt sich dabei aus vier interagierenden Teilmodellen zusammen: (1) Im Geometriemodell wird der Entwurf ausgehend von einem Basisgeometrieelement beschrieben, das die übergeordnete Form des Tragwerks repräsentiert. Der Einsatz von NURBS ermöglicht dabei eine effiziente Geometriebeschreibung durch wenige, signifikante Knotenpunkte. (2) Die probabilistischen Eigenschaften der geometrischen Imperfektionen werden in einem geeigneten Imperfektionsmodell mit Hilfe des Konzepts der Zufallsfelder realitätsnah beschrieben. Hierbei werden die durch das Zufallsfeld definierten Kovarianzen der zufälligen Abweichungen an diskreten Punkten ausgewertet und in einer Kovarianzmatrix zusammengefasst. Durch eine spektrale Zerlegung der Kovarianzmatrix gelingt es, die Imperfektionsform als Linearkombination deterministischer Basisvektoren (den Eigenvektoren der Kovarianzmatrix) mit zufälligen Amplituden darzustellen. Die zufälligen Amplituden sind dabei unabhängige standardnormalverteilte Zufallszahlen. (3) Für die innerhalb der Optimierung durchzuführende Strukturanalyse wird ein Finite-Elemente-Modell des Tragwerks verwendet. Um das Stabilitätsverhalten des Tragwerks zu erfassen, ist eine geometrisch nichtlineare Berechnung zwingend notwendig. (4) Ein Unsicherheitsmodell wird verwendet, um den Unsicherheiten, die im System aufgrund der geometrischen Imperfektionen vorhanden sind, Rechnung zu tragen. Da sich keine hinreichend genauen Aussagen über die probabilistischen Eigenschaften geometrischer Imperfektionen komplexer Tragsysteme treffen lassen, wird ein konvexes Modell der Unsicherheiten verwendet. Hierbei wird eine konvexe Hülle, in der die zu berücksichtigenden Imperfektionsformen enthalten sind, dadurch vorgegeben, dass die Norm des Imperfektionsvektors kleiner als eine vorgegebene Größe zu sein hat. Es wird diejenige Größe gewählt, bei der die Imperfektionsform innerhalb einer gegebenen Fraktile (bspw. 95 %) aller möglichen Imperfektionsformen liegt. Unter dieser Maßgabe ist es möglich, die ungünstigste Imperfektionsform mit Hilfe von Optimierungsverfahren zu bestimmen („Anti-Optimierung“). Die zugehörige Optimierungsprozedur beinhaltet somit zwei ineinander geschachtelte Schleifen: Eine äußere Optimierungsschleife, in der das Tragwerk optimiert wird, und eine innere Optimierungsschleife, in der die Suche nach der ungünstigsten Imperfektionsform für die benötigte Antwortgröße erfolgt. Im nachfolgend dargestellten Anwendungsbeispiel wird exemplarisch ein aus Einzelstäben zusammengesetzter Bogenträger mit dem entwickelten Modell optimiert (siehe Abb. 8).

Abb. 8: Robuste Optimierung eines druckbeanspruchten Bogenträgers

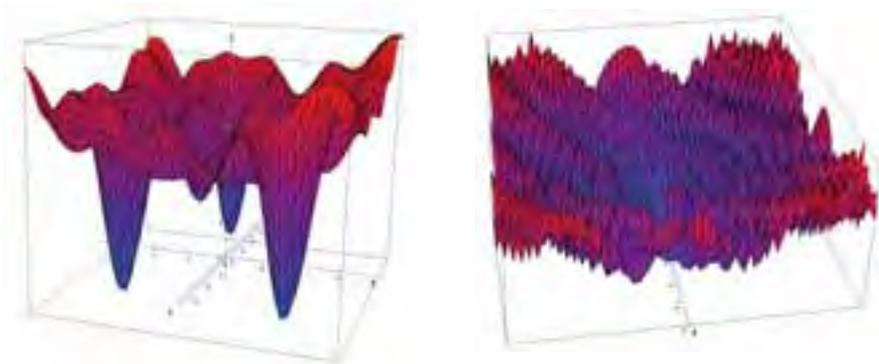
Oben: Vergleich der initialen und der optimierten Form des Bogenträgers. Unten: Optimierung der Systemsteifigkeit: Blau dargestellt ist die in jedem Optimierungsschritt erfolgende Suche nach der ungünstigsten Imperfektionsform



3 OPTIMIERUNG VON NICHTSTANDARDPROBLEMEN MIT MULTIAGENTENSYS-TEMEN (MAS): STATUS QUO

Die Lösung von auf numerischer Simulation aufbauenden Nichtstandardoptimierungsproblemen stellt – wie oben schon ausgeführt – aufgrund der komplizierten topologischen Verhältnisse im Suchraum der Optimierung (Nichtlinearität, Multimodalität, Unstetigkeiten etc.) allerhöchste Anforderungen an den Optimierungsexperten. Bei Problemstellungen mit raumzeitlichen Mehrebenen- bzw. Mehrstufen- und Mehrskalencharakter ist auch der Experte angesichts der mehrschichtigen Komplexität – selbst wenn robuste, allgemein anwendbare Optimierungsstrategien, wie etwa die Evolutionsstrategien (ES) eingesetzt werden – auf eine wissens- und erfahrungsgesteuerte Interaktion mit der Lösungsmethode bzw. lenkende Eingriffe in den Optimierungsablauf angewiesen. Bei der numerischen Simulation ist zudem noch mit zufällig verrauschten Ergebnissen zu rechnen, die zusätzlich bedacht werden müssen. Infolgedessen muss davon ausgegangen werden, dass der durch die jeweiligen Optimierungskriterien („Zielfunktionen“) und Restriktionen entstehende Antwortraum zumeist stark zerklüftet ist und zahlreiche

Abb. 9: Optimierungskriterium mit Mehrfachoptima

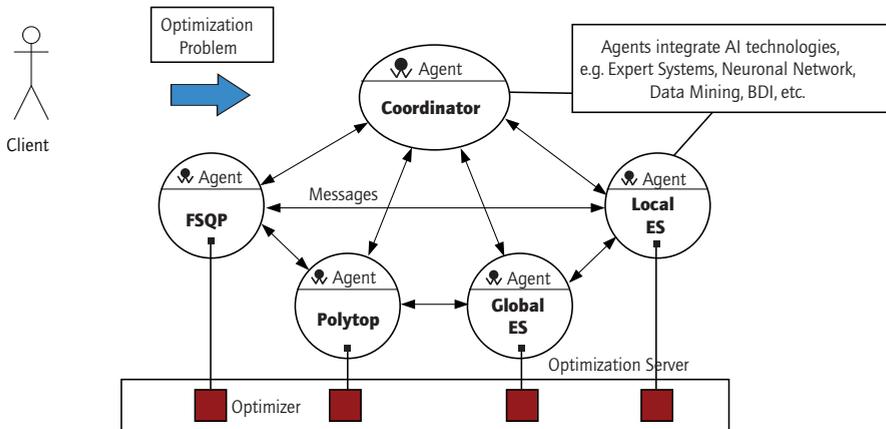


lokale Minima und Maxima enthält. Abb. 9, die ein zwar konstruiertes, aber die Gegebenheiten in der Praxis repräsentativ widerspiegelndes Testbeispiel für Mehrfachoptima bei einem zweidimensionalen Problem darstellt, veranschaulicht die auftretenden Schwierigkeiten. Wie lässt sich für ein derartig komplex zerklüftetes Optimierungsgebiet, dessen Dimension bei technischen Aufgabenstellungen von hier nur zwei schnell auf zehn und mehr Dimensionen anwachsen kann, ein globales Optimum finden? Eine Antwort lautet: Hochleistungsfähige, über das konventionelle Effizienzverhalten hinausgehende, simulationsbasierte Optimierungsstrategien können hier weiterhelfen! Noch wahrscheinlicher ist es aber, dass nicht eine einzige Optimierungsstrategie, sondern eine Mixtur von miteinander konkurrierenden bzw. kooperativ agierenden, unterschiedlichen Optimierungsstrategien den gewünschten Erfolg gewährleistet. Die passende anwendungsspezifische Strategiemixtur zusammenzustellen, ist eine lohnende Aufgabe für Mathematik, Technikwissenschaften und Informatik.

3.1 AGENTENSYSTEME

Das erst seit kurzem in der Ingenieurinformatik erprobte Paradigma der Multiagentensysteme (MAS) bietet dabei ein tragfähiges Theoriegebäude und das zugehörige Rüstzeug, um Nichtstandardprobleme der oben beschriebenen Art zufriedenstellend zu lösen [17], [4]. Auf dem Gebiet der Multiagentensysteme bisher durchgeführte Forschungsarbeiten belegen eindrucksvoll, welche gravierenden Vorzüge sich mit dem Einsatz des Agentenparadigmas verbinden: Die Vorteile gegenüber traditionellen Informatikkonzepten kommen vor allem dann zum Tragen, wenn die Problemlösung auch unvorhersehbare

Abb. 10: Beziehungsgeflecht eines Multiagentenverbundes



Effekte meistern muss und unabdingbar die Koordination, Kooperation, Kommunikation und Kollaboration vieler, fachkompetenter Beteiligter erfordert. Grundlage der Kollaboration ist dabei zumeist eine heterogene, zeitvariante, zum Teil unscharf strukturierte Organisationsstruktur der an der Problemlösung Beteiligten.

Eine derartige Situation liegt im Prinzip auch bei der Lösung der oben beschriebenen, simulationsbasierten Nichtstandardprobleme der technischen Optimierung vor, bei der immer mehrere „Beteiligte“, jetzt in Form miteinander konkurrierender bzw. kooperierender Optimierungsstrategien, gemeinsam – durch ein Beziehungsgeflecht miteinander verbunden – an einer Lösungsfindung mitwirken müssen (vgl. Abb. 10). Die mit zahlreichen Hindernissen verbundene Lösungsfindung nach einem globalem Optimum soll durch die Vorteile des kognitiv motivierten Lösungsansatzes, wie es der multiagentenbasierte Strategienverbund ist, nachhaltig verbessert werden.

3.2 STRATEGIENVERBUND

Bei der Lösung vieler komplexer Optimierungsprobleme aus den Technikwissenschaften, insbesondere bei multimodalen nichtlinearen Optimierungsproblemen mit „nur“ algorithmisch formulierbaren Optimierungskriterien bzw. -restriktionen, hat sich gezeigt, dass die bewusste Kopplung von stochastischen, global arbeitenden Optimierungsmethoden (Evolutionstrategien [5], Genetischen Algorithmen [9], Particle Swarm Methods [6], Ant Colony Concepts [14] etc.) mit deterministischen, lokal arbeitenden Methoden oft Erfolg versprechender ist als der Ansatz einzelner Optimierungsstrategien alleine. Ungeklärt ist allerdings, bei welcher Problemklasse welche Strategiemischung jeweils zu

wählen ist, um robuste, im Hinblick auf die globale Optimalität dennoch effiziente Lösungen zu erhalten. Eine gerade noch gut funktionierende Strategiemixtur kann durch nur leichte Abänderungen in der Problemstruktur rasch ineffizient oder sogar instabil werden. Der mit der simulationsbasierten Optimierung verbundene Aufwand (zur Erinnerung: Simulationen beruhen in der Regel auf strukturmechanisch nichtlinearen Finite-Elemente-Analysen!) bringt es dabei mit sich, dass in der konventionellen, „händischen“ Tragwerksplanung nur sehr wenige Strategiekombinationen bzw. Strategiemischungen untersucht werden können. Der Zeitaufwand, systematisch zu prüfen, welche Mischung von Strategien den besten Erfolg liefert, wäre unvertretbar hoch.

Mithilfe eines agentenbasierten Strategienverbundes jedoch lässt sich die bislang mehr intuitiv vorgenommene Mischung von numerischen Optimierungsstrategien durch einen wesentlich systematischer ablaufenden Ansatz ersetzen. Dies deshalb, weil das Paradigma der MAS beträchtliche Anteile autonomen bzw. emergenten Verhaltens enthält, die zudem gut formalisierbar und damit automatisierbar sind. Hierdurch lässt sich der oben erwähnte, übliche manuelle Aufwand, für neue Strategiekombinationen jeweils neue Optimierungsiterationen einzuleiten, auf die Agenten verlagern. Durch den Agentenansatz wird der Optimierungsexperte somit von stereotypem Interaktionsaufwand befreit. Genau hierin liegt der große Gewinn und Mehrwert gegenüber dem konventionellen Vorgehen einer nicht-agentenbasierten Optimierung. Die Vorteile der Unterstützung durch einen agentenbasierten Strategienverbund kommen dabei besonders bei Mehrebenen- bzw. Mehrstufen und Mehrskalproblemen zum Tragen, bei denen der Lösungsaufwand exponentiell mit der Komplexität zunimmt.

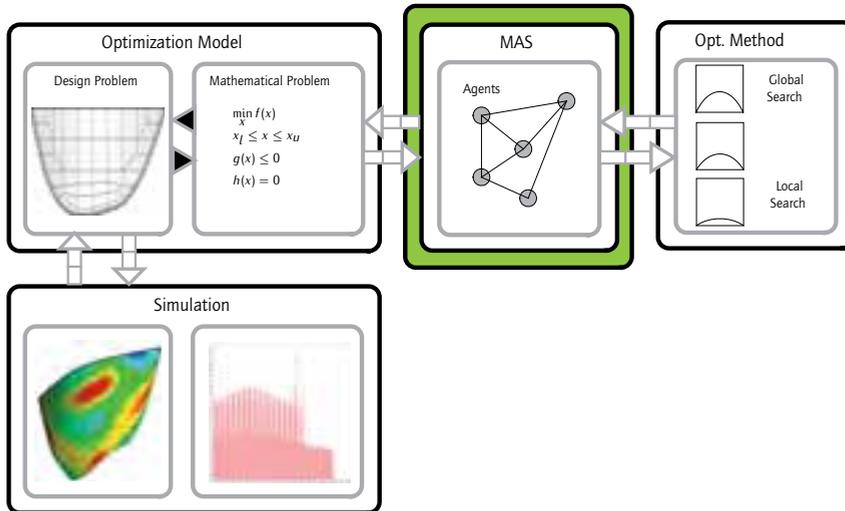
Intelligente und verteilte Optimierungskomponenten, wie sie Agenten darstellen, können im komplexen Lösungsraum mit einer auf die Optimierung zugeschnittenen Wissens- bzw. Regelbasis oder über Künstliche Neuronale Netze (KNN) [16] menschliche Optimierungsexperten ersetzen und ihr Lösungsverhalten ereignisgetrieben an unterschiedliche Nichtstandardoptimierungsprobleme anpassen. Durch die inhärenten Emergenzeffekte des Agentensystems kann man davon ausgehen (oder zumindest hoffen), dass der Strategienverbund auf Agentenbasis wesentlich leistungsfähiger ist als die Summe der Einzelstrategien für sich genommen.

VIER-SÄULEN-KONZEPT

In einem von der DFG geförderten Forschungsprojekt wird (vgl. Abb. 11) das klassische „Drei Säulen“-Konzept der technischen Optimierung, bestehend aus Simulation, Optimierungsmodell und -strategien [7], um eine vierte Säule erweitert: das Multiagentensystem (MAS). Dieses verkörpert quasi den „Künstliche Intelligenz“-Teil des Optimierungskonzeptes und verbindet das technische Optimierungsmodell (Strukturoptimierungsmodell) mit den verwendeten Optimierungsstrategien. Hierdurch gelingt es, eine allgemein gültige, jederzeit einsetzbare Lösungsmethodologie für die simulationsbasierte Optimierung bei technischen Problemstellungen bereitzustellen, unabhängig von einer konkreten Anwendung.

Abb. 11: Drei-Säulen Optimierungskonzept mit Multiagentensystem (MAS)

Das Multiagentensystem (MAS) verkörpert den „Künstliche Intelligenz“-Teil des simulationsbasierten Optimierungskonzeptes. Mithilfe dieses Konzeptes können unterschiedliche Ingenieursprobleme, für die es Computersimulationen gibt, allgemeingültig und intelligent gelöst werden, wie z. B. Verkehrssimulation, Sprengsimulation, Systemidentifikation, optimale Sensorplatzierung etc.



ANALYSEWERKZEUGE, VISUALISIERUNG, STEERING

Bei der bisherigen, nicht auf Agentensysteme zurückgreifenden Vorgehensweise zur Lösung von simulationsbasierten Optimierungsproblemen werden üblicherweise unterschiedliche, nicht integrierte Computerprogramme und -werkzeuge für anfallende Teilaufgaben, wie z. B. Pre-/Postprocessing, numerische Simulation etc., eingesetzt. Der Datenaustausch zwischen diesen Werkzeugen erfolgt in der Regel „Datei-basiert“: Die Ausgabedaten eines vorangehenden Programmes müssen immer erst formatgerecht in Eingabedaten für das im Optimierungsablauf nächst folgende Programm umgewandelt werden. Dies hat ggf. aufwendige manuelle Benutzereingriffe zur Folge, die zeitraubend sind und die automatische Ausführung iterativer, ineinandergreifender Computerberechnungen erschweren.

Eine Möglichkeit, die Nachteile der traditionellen Vorgehensweise zu umgehen, ist der Einsatz von „Computational Steering“, mit dem ein Programm während seiner Ausführung gezielt kontrolliert werden kann und das vor allem die Interaktionen zwischen dem Benutzer und dem Programm unterstützt. In Abb. 12 wird die Softwarearchitektur der institutseigenen Simulations- und Experimentierplattform JASOE (Java Agent Simulation and Optimization Environment) vorgestellt. Diese als „Computational Steering“-

Anwendung konzipierte Plattform stellt zahlreiche Kontroll- und Interaktionsmöglichkeiten bereit, um mit unterschiedlichen Optimierungsstrategien und Agentenverbänden „Experimente“ durchführen zu können. Die „Experimente“ lassen sich dann zur zielgerichteten Steuerung bei der Suche nach dem globalen Optimum verwenden.

Ein Blick auf die zur Computational Steering-Plattform (Abb. 13) gehörende grafische Benutzungsoberfläche zeigt, das zahlreiche, an die natürliche Arbeitsweise eines Optimierungsexperten angepasste Bedienungs- und Kontrollfunktionen realisiert wurden.

Besondere Bedeutung muss der visuellen Beschreibung sowohl des der Optimierung unterlegten technischen Systems als auch den Details des Optimierungsprozesses beigemessen werden. Visuelle Beschreibungen üben seit jeher eine hohe Anziehungskraft auf Menschen aus: Sie unterstützen und erleichtern auf verschiedene Weise das Verständnis für komplexe Systeme und Prozesse. Das gilt auch für die Optimierung mit einem MAS, in dem verschiedene Komponenten und Teilprozesse auf einander abzu-

Abb. 12: Architektur des agentenbasierten Optimierungssystems

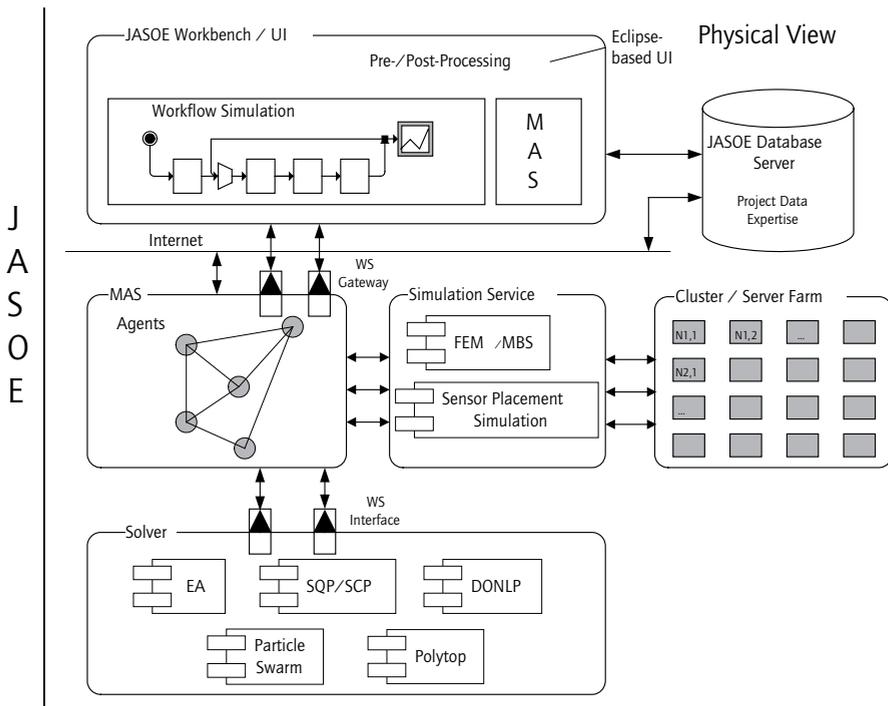
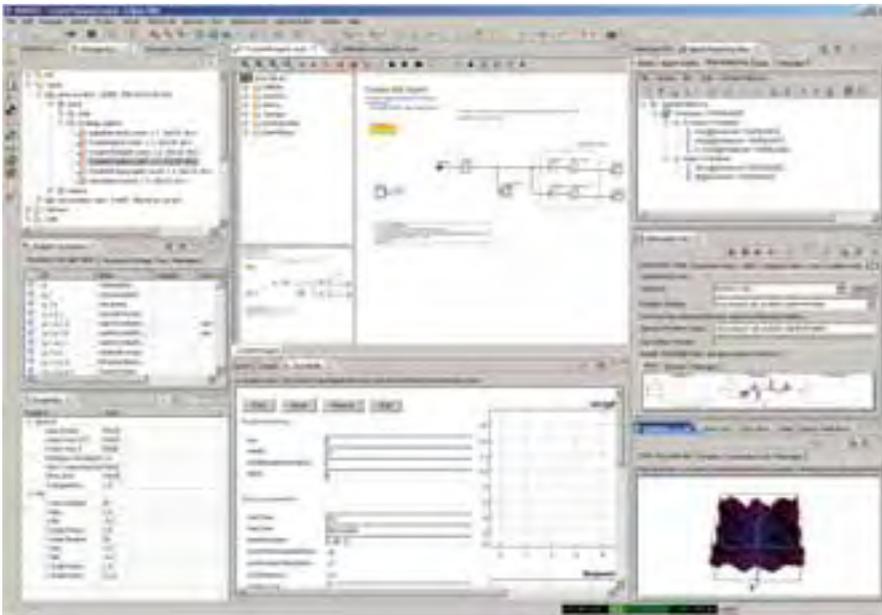


Abb. 13: Computational Steering Workbench für die Optimierung mit MAS



stimmen sind. Visuell aufbereitete Echtzeitinformationen über den Optimierungsprozess in Form von leichtverständlichen Diagrammen, Plots, Nachrichtenvermittlung, Ereignisdarstellungen etc. verschaffen dem Optimierungsexperten einen detaillierten Einblick in das innere Leben des MAS. Rechenberg, einer der Väter der Evolutionsstrategie, wirbt in seinem Buch „Evolutionsstrategie '94“ [15] für die systematisch betriebene Interaktion zwischen Mensch und Optimierungsprozess.

4 STÄRKEN/SCHWÄCHEN-ANALYSE, HERAUSFORDERUNGEN

Auch wenn im vorangehenden Kapitel gezeigt wurde, dass durch die Verknüpfung und Vernetzung von Technikwissenschaften mit Mathematik bzw. Informatik bislang nicht gelöste, technische Problemstellungen behandelbar und – trotz ihre Komplexität – lösbar geworden sind, besteht kein Grund zur Euphorie. Eine nüchterne, sachbezogene Analyse der Stärken, die sich aus den zweifelsohne vorhandenen Synergieeffekten der Zusammenführung der einzelnen Wissenschaftsdisziplinen ergeben, kann gleichzeitig dazu dienen, existierende Schwächen aufzudecken. Deren Behebung stellt nach wie vor eine Herausforderung dar.

4.1 PRAXISGEBENHEITEN

Die in der Ingenieurpraxis zu bewältigenden, großen Problemstellungen sind von einer Komplexität geprägt, deren Konsequenzen der theoretisch arbeitende, mit einer völlig anderen Art von Komplexität umgehende Mathematiker – sowohl von seiner Ausbildung als auch von seiner Denkweise her – nur im Team mit Technikwissenschaftlern erfassen kann. Mit dieser Aussage ist keine Wertung der einen oder anderen Disziplin verbunden. Es soll lediglich zum Ausdruck gebracht werden, dass man einander braucht und miteinander „sprechen können“ muss.

Die Komplexität der Problemlösungen im Bereich der Bautechnik und Umwelttechnik, die im Fokus dieses Beitrages stehen, hängt überwiegend damit zusammen, dass immer mehrere Beteiligte mit unterschiedlichen, teilweise sogar divergierenden Interessen, zeitlich begrenzt an einer gemeinsamen Lösung arbeiten, wobei unterschiedliche Lösungsansätze und Lösungsmethoden zum Einsatz kommen. Dies soll an einem typischen Beispiel deutlich gemacht: Als Tragwerksplaner und Konstrukteur hat man bei der Errichtung eines Bauwerks andere Vorstellungen von der Optimierung als ein mit strukturmechanischen Berechnungen betrauter Wissenschaftler, bspw. aus dem Gebiet der Technischen Mechanik, oder als ein Mathematiker mit numerischer Optimierung als Spezialgebiet. Wiederum anders ist die Situation bei einem Architekten oder einem Bauphysiker, oder einem für die Fundamente des Tragwerks zuständigen Grundbauer/Geotechnikexperten. Dem Strukturmechaniker und auch dem Mathematiker reicht es durchaus aus, als Optimierungskriterium eine Zielfunktion im Sinne der klassischen Mathematik, beispielsweise das Gewicht einer Tragstruktur in Abhängigkeit von Optimierungsvariablen, zu verwenden. Und auch die Nebenbedingungen werden als klassische mathematische Funktionen mehrerer Optimierungsvariablen formuliert. Dafür liegt das Interesse des Strukturmechanikers mehr daran, das nichtlineare Tragverhalten möglichst wirklichkeitsnah zu simulieren, wozu wieder die oben erwähnten, zumeist nichtlineares Tragverhalten erfassenden Finite-Elemente-Programmsysteme eingesetzt werden (vgl. Ausführungen zur p -Version der FEM). Tragwerksplaner dagegen möchten ihre Qualitätsvorstellungen in Form von Bewertungsvorschriften und Restriktionen formulieren können, die keine mathematischen Funktionen im klassischen Sinne sind, sondern nur in Algorithmusform gefasst werden können. Herstellungskosten beispielsweise, die verwickelte Kostenmodelle zur Folge haben, oder konstruktive Randbedingungen und auch fertigungstechnische Vorgaben, führen auf Optimierungskriterien bzw. Restriktionen mit algorithmischem Charakter. Das Gleiche gilt für Optimierungskriterien und -restriktionen, die erst auf der Grundlage komplexer, numerischer Simulation gewonnen werden können. Beispiele aus der eigenen Forschung hierzu sind die lebensdauerorientierte Strukturoptimierung in Raum und Zeit (semi-infinite stochastische Optimierung), inverse Problemstellungen in der Bautechnik (Systemidentifikation zur Lokalisierung von Schädigungen in Tragwerken durch Back-Analysis), die Optimierung von Kollapsvorgängen bei der Präzisionsspannung komplexer Tragwer-

ke (Optimierung über raumzeitliche Mikro-, Meso- und Makrobereiche) oder die optimale Platzierung von Sensoren für das Tragwerksmonitoring (gemischt-diskrete Optimierung mit ereignisbehafteten, zeitvarianten Optimierungsvariablen und -restriktionen).

4.2 NEUE OPTIMIERUNGSSTRATEGIEN/ ALGORITHMISIERUNG

Um die oben beschriebenen Problemstellungen zu lösen, sind zukünftig noch mehr als bisher Kooperationen zwischen der Mathematik, der Ingenieurinformatik und den (Bau-)Technikwissenschaften erforderlich. Nur so gelingt es, bestehende Schwächen der mathematisch orientierten Optimierung abzubauen, in der für den Mathematiker, nicht aber für den Ingenieur relevante Problemklassen behandelt werden. In den Ingenieurwissenschaften und in der Ingenieurpraxis dominieren heute Nichtstandardoptimierungsprobleme, die in der Regel – zumindest bei den komplexen Aufgabenstellungen – durch hochgradige Nichtlinearität, Mehrfachoptima, irreguläre Suchräume, Nichtdifferenzierbarkeit und bei Einbindung von Größen, die mit dem Parameter „Zeit“ behaftet sind, durch Semi-Infinitheit geprägt sind. Zudem sind zunehmend Optimierungsprobleme zu lösen, die nur über die numerische Simulation, z. B. mit Finiten Berechnungsmodellen (Finite-Elemente-Methoden, Rand-Elemente-Methoden, Partikelmethode etc.) oder auf Mehrkörpermodellen beruhenden Programmsystemen gefunden werden können. In Zusammenarbeit zwischen Mathematik, Ingenieurinformatik und Technikwissenschaften müssen folglich bestehende Optimierungsstrategien noch weiter verfeinert und bezüglich ihres Effizienzverhaltens verbessert werden; auch die Erfindung neuer Optimierungsstrategien muss eine Option sein. Die Vorstellungen der Technikwissenschaften müssen dabei von den Mathematikern bzw. Informatikern verstanden werden, die Technikwissenschaften müssen andererseits die mathematische Denkweise und die Qualitätsansprüche der Mathematiker an eine Lösungsfindung (Beweisführung bzgl. Eindeutigkeit, Berechenbarkeit, Konvergenzverhalten etc.) beachten.

4.3 ERWEITERUNG DER MATHEMATISCHEN GRUNDLAGEN FÜR INGENIEURE

Ein weiterer, ernstzunehmender Schwachpunkt bei der Modellierung komplexer Zusammenhänge ist, dass eine wichtige mathematische Grundlage, die Graphentheorie [11], bisher nur unvollständig Eingang in heutige Ingenieurmodelle gefunden hat. Graphentheoretische Konzepte sind aber für viele Anwendungen in der Technik unerlässlich. Ein Beispiel hierfür ist die objektorientierte Modellierung von Kollapsprozessen infolge der oben erwähnten Präzisionssprengung oder infolge terroristischer Attacken bzw. Naturhazards. Hier kommt es darauf an, mit Hilfe von Graphen beim Kollaps auftretende, offene und/oder geschlossene kinematische Ketten aufzudecken, die bestimmend für das Kollapsverhalten sind und bei der Auffindung geeigneter (optimaler) Gegenmaßnahmen gegen unerwünschte Effekte benötigt werden. Weitere Beispiele für den Einsatz der Gra-

phentheorie sind die Zusammenarbeit von Agenten in föderativen Agentensystemen, die Sensorplatzierung und -bestückung von Tragwerken für das Monitoring oder die Abbildung von Arbeitsabläufen (Work Flows) im Engineering mit Hilfe von ggf. farbigen und stochastisch verrauschten Petrinetzen.

Beachtliches Umdenken ist auch bei der mathematischen Modellierung von Unschärfe, Unsicherheit und Unvollständigkeit bezüglich verwendeter Daten und verwendeter Modelle erforderlich. Die Anstrengungen der Mathematik müssen hier weiter intensiviert und zugeschärft werden, nachdem in mehreren Bereichen der Ingenieurwissenschaften (Fuzzy Control, Anwendungen der Fuzzy-Randomness bei der Bemessung von Tragwerken und der Behandlung von Ingenieurprozessen [13]) die Bedeutsamkeit von Unschärfebehandlung im Hinblick auf die Akzeptanz von Lösungen nachgewiesen wurde. Die konsequente Verzahnung von Fuzzy-Randomness mit der Technischen Optimierung, die viele neuartige Probleme aufwirft, befindet sich – abgesehen von einigen Ausnahmen – immer noch in den Anfängen, ist aber zweifelsohne überfällig. Die Sicherheit und Zuverlässigkeit von Lösungen wird hierdurch erheblich verbessert und auch die Akzeptanz von wissenschaftlich untermauerten Lösungen durch die Praxis würde sich nachhaltig erhöhen.

5 VISIONEN UND HANDLUNGSEMPFEHLUNGEN

Die bisher vorliegenden Forschungsergebnisse – aber auch die laufenden Forschungsanstrengungen – berechtigen zu der Annahme, dass demnächst die hinreichend genaue Lösung hochgradig komplexer, simulationsbasierter Optimierungsprobleme aus dem Bereich der Bautechnik und verwandter Gebiete möglich sein wird. Damit eröffnen sich neue Dimensionen für weite Bereiche der Technikwissenschaften: Man könnte sogar von einer neuen Ära für die Arbeit der Ingenieure sprechen. Man ist nicht mehr länger darauf angewiesen, sich „nur“ mit relativen Optimallösungen zufrieden zu geben, die – je nachdem von welcher Anfangslösung eine Optimierungssuche gestartet wurde – „nur“ zu lokalen Verbesserungen der Systeme bzw. Prozesse führen. Dass die bisherige Vorgehensweise in vielen Fällen schon von erheblichem Nutzen gewesen ist, soll dabei nicht bestritten werden!

Die sich im Bereich der Technischen Optimierung anbahnenden, weitreichenden Möglichkeiten hängen mit der konsequenten Anwendung neuer Lösungskonzepte aus der Informatik bzw. Ingenieurinformatik zusammen, insbesondere der Agententechnologie sowie der Objektorientierten Modellierung und Programmierung (OOMP). Diese erlauben es, ein neuartiges „Computational Steering“ für die Technische Optimierung aufzubauen. Hierdurch wird, ohne dass der mit der System- oder Prozessoptimierung betraute Ingenieur seiner übergeordneten Lenkungs- und Eingriffsmöglichkeiten beraubt wird, die Qualität des Entwerfens, Dimensionierens, Konstruierens und der kompetenten Begleitung von Ingenieurprozessen ganz wesentlich verbessert. Die globale Lösung des „Optimierungsproblems“ wird dabei durch einen Verbund bzw. ein problemadäquates

Beziehungsgeflecht von Optimierungsstrategien gewonnen, wobei jede dieser Strategien durch einen Softwareagenten wettbewerblich vertreten wird.

Unabhängig von den Fortschritten im Bereich der Optimierungsalgorithmik und ihrer Umsetzung in Software ist es zudem unabdingbar, dass die im Kern der Optimierungssiteration verwendeten System- bzw. Prozesssimulationsmodelle höchste Effizienz bezüglich Rechenzeit und Speicherbedarf aufweisen. Dies zu erreichen, ist eine wichtige Aufgabe, die am besten durch gezielte Verbundforschung zwischen der Mathematik, den Technikwissenschaften und der Informatik geleistet werden kann.

Besonders gefragt ist allerdings die Mithilfe der Mathematiker, wenn es darum geht, die Globalität der Optimallösung theoretisch fundiert abzusichern. Ober- und Unterschranken für die gesuchte globale Lösung beispielsweise sind von hohem Interesse; diese könnten aus den nur punktförmig, über die jeweilige Simulation errechneten Informationen für das Optimierungskriterium und die -restriktionen mittels guter Approximationen (z. B. über Antwortflächen) angegeben werden. Aber auch die Formulierung von Kooperationsbedingungen für die eingesetzten, unterschiedlichen Optimierungsstrategien (Optimierungsagenten) oder Initialisierungsanweisungen für bestimmte Strategien sowie Abbruchbedingungen stellen offene Fragestellungen dar. Die mathematische Behandlung von Unschärfe und Zeitvarianz im Hinblick auf das Lösungsverhalten (Stabilität, Eindeutigkeit, Konvergenz) bedarf ebenfalls weiterer Forschung. Nur durch enge Zusammenarbeit zwischen Mathematik, Technikwissenschaft und Informatik sind die auch noch in den nächsten Jahren verbleibenden Lücken sowohl in der Theorie als auch bei der praktischen Umsetzung zu schließen.

6 LITERATUR

- [1] T. Bäck. *Evolutionary Algorithms in Theory and Practice*. Oxford University Press, 1996.
- [2] M. Baitsch and D. Hartmann. Optimization of slender structures considering geometrical imperfections. *Inverse Problems in Science and Engineering*, 14(6):623–637, 2006.
- [3] M. Baitsch, Y. Hu, and D. Hartmann. Parallel shape optimization of three-dimensional continua with high-order finite elements. In J. Herskovits, editor, *Proceedings of the International Conference on Engineering Optimization*, Rio de Janeiro, Brazil, June 2008.
- [4] F. Bellifemine, G. Caire, and D. Greenwood. *Developing Multi-Agent Systems with JADE*. Wiley & Sons, 2007.
- [5] H.-G. Beyer. *The Theory of Evolution Strategies*. Springer-Verlag, 2001.

- [6] M. Clerc. *Particle Swarm Optimization*. ISTE Ltd, 2006.
- [7] H. A. Eschenauer, J. Koski, and A. Osyczka. *Multicriteria Design Optimization*. Springer Verlag, Heidelberg, Berlin, New York, 1990.
- [8] M. Gálffy, M. Baitisch, A. Wellmann-Jelic, and D. Hartmann. Lifetime estimation of vertical bridge tie rods exposed to wind-induced vibrations. In C.A. Mota Soares, J. A. C. Martins, H.C. Rodrigues, and J.A.C. Ambrósio, editors, *Proceedings of the IIIrd European Conference on Computational Mechanics Solids, Structures and Coupled Problems in Engineering*, Lisbon, 2006. Springer.
- [9] D. E. Goldberg. *Genetic algorithms in search, optimization, and machine learning*. Addison-Wesley, Reading, MA, 1989.
- [10] H. Grill. *Ein objektorientiertes Programmsystem zur gemischt-diskreten Strukturoptimierung mit verteilten Evolutionsstrategien*. Dissertation, Institut für konstruktiven Ingenieurbau, Ruhr-Universität Bochum, 1997.
- [11] J.L. Gross and J. Yellen. *Graph Theory And Its Applications*. Chapman & Hall/CRC, 2006.
- [12] D. Hartmann and K.-R. Leimbach. Analyse von Strukturoptimierungsmodellen im Hinblick auf ihre parallele Realisierung und Umsetzung, 1996. Abschlussbericht, DFG-Forschergruppe Optimierung in der Strukturmechanik.
- [13] B. Möller and M. Beer. *Fuzzy Randomness*. Springer-Verlag, 2004.
- [14] G. C. Onwubolu and B. V. Babu. *New Optimization Techniques in Engineering*. Springer-Verlag, 2004.
- [15] I. Rechenberg. *Evolutionsstrategie '94*. Frommann-Holzboog, Stuttgart, 1994.
- [16] G. D. Rey and K. F. Wender. *Neuronale Netze: Eine Einführung in die Grundlagen, Anwendungen und Datenauswertung*. Huber, 2008.
- [17] S. Russell and P. Norvig. *Künstliche Intelligenz: Ein moderner Ansatz*. Pearson Studium, 2003.
- [18] SFB-398. Life time oriented design concepts. Monographie zum DFG-Sonderforschungsbereich SFB 398 „Lebensdauerorientierte Entwurfskonzepte“, ca. 600 Seiten, in Druck, erscheint 2008.

- [19] J.M.T. Thompson. Optimization as a generator of structural instability. *International Journal of Mechanical Sciences*, 14:627–629, 1972.

> OBJEKTORIENTIERTE MODELLIERUNG ZUR SIMULATION UND REGELUNG VON ENERGIEWANDLUNGSPROZESSEN

DIRK ABEL

1 EINFÜHRUNG UND ÜBERSICHT

In allen Ingenieursdisziplinen hat die mathematische Modellierung technischer Systeme und Prozesse ihren festen Platz gefunden – nicht zuletzt getrieben durch die Ausführbarkeit auf leistungsfähigen Digitalrechnern, die zur Simulation, zur Analyse und zum Entwurf eingesetzt werden. Beflügelt wird dies durch neuere Entwicklungen im Bereich der Modellierungssprachen und Rechnerwerkzeuge, die mit modernen objektorientierten Konzepten und algebraisch/symbolischen Rechenfähigkeiten aufwarten. Anhand zweier aktueller Forschungsprojekte zur Energiewandlung – stellvertretend für die Bereiche Kraftwerkstechnik und Verbrennungsmotoren – werden der Zugang zur objektorientierten Modellierung und deren Potential gegenüber konventionellen Ansätzen aus der Ingenieurssicht, speziell aus dem Blickwinkel der Regelungstechnik, aufgezeigt.

Die Bedeutung mathematischer Modelle für die Regelungstechnik ist offenkundig; schließlich setzt (fast) jeder Reglerentwurf auf einem Modell des zu regelnden Prozesses auf. Als Abbild des in diesem Bereich anzutreffenden, vorzugsweise rückwirkungs-freien Ursache/Wirkungs-Prinzips hat sich der Wirkungsplan als wichtiges graphisches Hilfsmittel zur Beschreibung dynamischer Systeme mit komplexer Struktur etabliert. Als Benutzeroberfläche findet sich der Wirkungsplan daher in vielen gängigen blockorientierten Simulationsprogrammen wieder (wie z. B. SIMULINK). Seit Ende der 90er Jahre liegt mit MODELICA eine so genannte objektorientierte Beschreibungssprache für physikalische Modelle vor, für die ebenfalls verschiedene graphische Entwicklungs- und Simulationsprogramme (wie z. B. DYMOLA) existieren. MODELICA erlaubt die Beschreibung von übergreifenden Aufgabenstellungen in einem weiten Bereich von Wissensgebieten: Mechanik, Elektrotechnik und Elektronik, Thermodynamik, Hydraulik und Pneumatik, Regelungstechnik und Prozesstechnik.

Objektorientierte Modellierungs- und Simulationswerkzeuge wie DYMOLA eignen sich aufgrund der Modellstruktur und der damit einhergehenden Modularisierbarkeit hervorragend zur Simulation komplexer Systeme, die aus mehreren interagierenden Komponenten bestehen. Die Modellierung der Teilsysteme stützt sich auf algebraische

Basisregeln ab, wobei alle auftretenden Signale in Potential- und Flussgrößen kategorisiert werden. Die Teilsysteme können dann durch vektorielle Schnittstellen, so genannte *Konnektoren*, miteinander verbunden werden. Das wesentliche Merkmal dieser Konnektoren ist darin zu sehen, dass Teilsysteme eines zu simulierenden Systems, so z. B. dynamische Modelle von Komponenten eines zu regelnden Prozesses, durch Schnittstellen miteinander verbunden werden und in Wechselwirkung treten, die eine unmittelbare Analogie zur Realität aufweisen. Dazu gehört auch, dass der Anwender – durch die algebraisch/symbolische Rechenfunktionalität der Werkzeuge – von der bei konventionellen Modellierungs- und Simulationsansätzen unabdingbaren Pflicht entbunden wird, über Ursache/Wirkungsrichtungen zu entscheiden.

2 ANWENDUNG IM KRAFTWERKSBEREICH: PROJEKT OXYCOAL-AC

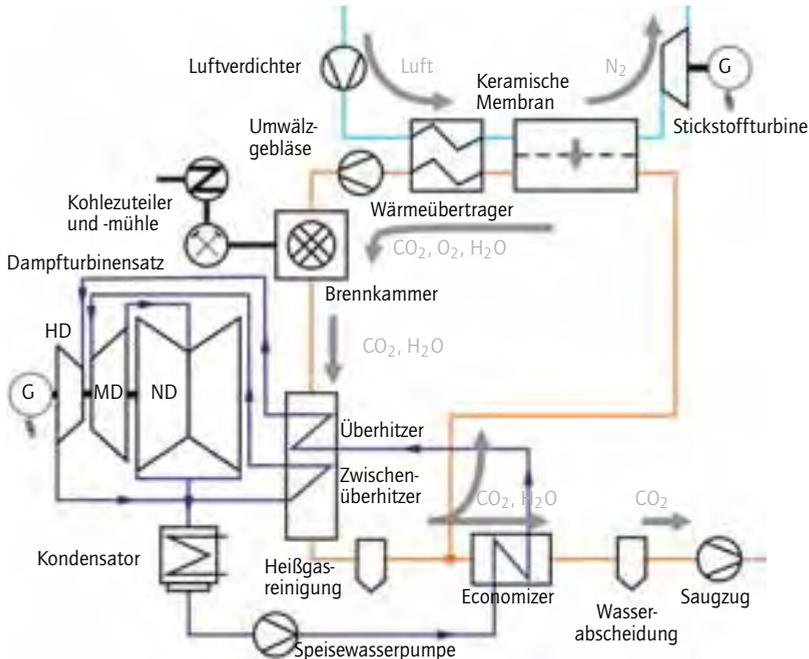
2.1 PROJEKTZIELE UND -ERFORDERNISSE

Das Projekt OXYCOAL-AC ist eines der Projekte des COORETEC-Programms (siehe [5]) des Bundesministeriums für Wirtschaft und Arbeit, das durch die Verpflichtung der Vereinten Nationen durch das Kyoto-Protokoll zur drastischen Reduktion des CO₂-Ausstoßes hervorgerufen wurde. Das Abkommen schreibt verbindliche Zielwerte für den Ausstoß von Treibhausgasen fest, welche die hauptsächliche Ursache der globalen Erwärmung sind.

Im OXYCOAL-AC-Projekt wird ein Prozess zur CO₂-emissionsfreien Kohleverbrennung untersucht, bei dem das entstehende Kohlendioxid zur Speicherung in geologischen Formationen aus dem Prozess abgeschieden wird. Die hierfür notwendige Voraussetzung der wirtschaftlichen Abtrennung des Kohlendioxids aus dem Rauchgas des Kraftwerksprozesses wird durch eine Verbrennung des Brennstoffs unter einer Rauchgas/Sauerstoff-Atmosphäre realisiert. Damit besteht das Verbrennungsgas im Wesentlichen aus Kohlendioxid und Wasser. Eine Prinzipskizze des Verbrennungsprozesses des OXYCOAL-AC-Prozesses zeigt Abb. 1. Die Bereitstellung des Sauerstoffs erfolgt über eine Hochtemperatur-Membrananlage. Eine keramische Membran wird auf der einen Seite (Luftseite) mit stark verdichteter und in einem Wärmeübertrager vorgewärmten Luft und auf der anderen Seite (Rauchgasseite) mit heißem Rauchgas durchströmt. Damit stellt sich ein hohes Sauerstoffpartialdruckgefälle über der Membran ein und es diffundiert Sauerstoff von der Luftseite über die Membran auf die Rauchgasseite. Wie in der Abbildung zu sehen, wird kontinuierlich ein Teil des entstehenden Rauchgases durch ein Umwälzgebläse rezirkuliert, mit Sauerstoff angereichert und für die Verbrennung genutzt, während der andere Teil aus dem Prozess ausgeschleust wird.

Die beschriebene Struktur des OXYCOAL-AC-Prozesses erfordert neue Regelungskonzepte zur Bereitstellung des Verbrennungsgases (siehe auch [7]). Während bei herkömmlichen Kraftwerken der Verbrennungsluftmassenstrom bei einer Veränderung des Brennstoffmassenstroms schnell über die Drehzahl der Luftgebläse angepasst werden kann,

Abb. 1: Skizze des OXYCOAL-Prozesses



liegen hier komplexere Zusammenhänge vor. Es ist zu erwarten, dass Änderungen im Diffusionsmassenstrom über die Membran aufgrund der Wärmeübergänge in Wärmeübertrager und Membran sowie den Totzeiten durch Rohrleitungen nur verhältnismäßig langsam durchgeführt werden können. Zum anderen ist zu beachten, dass die keramischen Werkstoffe der Membran sehr empfindlich auf örtliche wie zeitliche Temperaturänderungen reagieren und somit nur moderate Änderungsraten der Temperatur zulässig sind, um das Material zu schonen. Zur Entwicklung eines geeigneten Regelungskonzepts für das nichtlineare, stark gekoppelte und tozeitbehaftete Mehrgrößensystem ist ein dynamisches Prozessmodell notwendig. Das Modell sollte die Membran sehr detailliert abbilden, damit maximal zulässige Temperaturgradienten im Regelungskonzept berücksichtigt werden können.

2.2 OBJEKTORIENTIERTE MODELLIERUNG VON KRAFTWERKSPROZESSEN

Steigende Anforderungen bezüglich der Sicherheit, Zuverlässigkeit und Effizienz im Kraftwerksumfeld machen eingehende mathematische und physikalische Untersuchun-

gen bei der Entwicklung von Kraftwerken unverzichtbar. Aufgrund der Komplexität der Gleichungssysteme sind solche Analysen ohne entsprechende numerische Modellierungs- und Simulationswerkzeuge nicht zu bewältigen. Während eine Effizienzsteigerung bei Kraftwerksprozessen durch Optimierung der Prozessstruktur auf der Grundlage von stationären Simulationen erfolgt, ist für den späteren sicheren und zuverlässigen Betrieb eine dynamische Abbildung zur optimalen Auslegung von Regelkreisen erforderlich. Für die dynamische Modellierung sind komponentenbasierte Verfahren mit standardisierten Schnittstellen vorteilhaft, um eine vielseitige und universelle Übertragbarkeit der Modelle einzelner Kraftwerkskomponenten auf Prozesse unterschiedlicher Strukturen oder Größenordnungen zu ermöglichen. Damit wird ein hoher Wiederverwendungsgrad einmal entwickelter Modelle erreicht und die Anzahl der zu modellierenden Komponenten wird reduziert.

Die Umsetzung einer komponentenbasierten Modellierung und die Verwendung normierter Schnittstellen für den Informationsaustausch zwischen Teilmodellen ist der Grundgedanke der objektorientierten Beschreibungssprache Modelica. Sie bietet daher einen flexiblen Ansatz, Modelle von Kraftwerkskomponenten zu beschreiben. Modelicas nicht kausales Modellierungskonzept erlaubt die Formulierung von Gleichungen auch in impliziter Form, wie sie aus den physikalischen Zusammenhängen bekannt sind, ohne dass eine Auflösung der Gleichungssysteme nach Ein-, Ausgangs-, oder Zustandsgrößen erforderlich ist und ohne dass eine Signalfflussrichtung vorgegeben werden muss. Damit kann ein und dasselbe Modell mit entsprechender Parametrierung flexibel in unterschiedlichen Verschaltungen und Kontexten eingesetzt werden. Weiterhin ergibt sich daraus für den Anwender ein leicht verständlicher Weg, bestehende Modelle nachzuvollziehen und gegebenenfalls für seine Anwendung zu modifizieren.

Die objektorientierten Prinzipien der Vererbung und der Polymorphie erhöhen zusätzlich die Flexibilität der Modellierung. Der Mechanismus der Polymorphie ermöglicht es, durch Überladen ein Komponentenmodell mit verschiedenen auswählbaren Gleichungssystemen zu versehen. Damit kann beispielsweise mit dem gleichen Modell durch Formulierung von Gleichungssystemen unterschiedlicher Detaillierungsgrade für verschiedene Simulationen in der Genauigkeit und Simulationsdauer variiert werden. Erstellte Modelle können in Bibliotheken strukturiert abgelegt werden. Aufgrund der beschriebenen Vorteile des objektorientierten Ansatzes existieren speziell für die Modellierung von Kraftwerksprozessen bereits frei verfügbare Komponentenbibliotheken. Aufbauend auf der zum Modelica-Sprachstandard gehörenden Media-Bibliothek (siehe [4]) liefert die frei verfügbare ThermoPower-Bibliothek wichtige Grundlagen zur Modellierung von Kraftwerksprozessen (siehe [2]). Das Vorgehen bei der Modellierung des OXYCOAL-AC-Prozesses wird in den folgenden Abschnitten erläutert.

2.3 OBJEKTORIENTIERTE MODELLIERUNG DES OXYCOAL-KRAFTWERKS

Zur Modellierung des in Abschnitt 2.1 beschriebenen Prozesses ist eine Bibliothek mit allen Prozesskomponenten aufgebaut worden. Wie in Abb. 2 dargestellt, basiert die entwickelte OXYCOAL-AC-Bibliothek auf der ThermoPower-Bibliothek, die wiederum auf den Modelica-Standard-Bibliotheken aufbaut, und besteht aus den drei Hauptbereichen pckBasis, pckComponent und pckSimulation. Im Paket pckBasis werden Einheiten, Konstanten, Konnektoren und Basisklassen definiert, das Paket pckComponent umfasst eine Sammlung der Komponentenmodelle und im Paket pckSimulation sind aus den Einzelkomponenten zusammengesetzte Gesamtmodelle zur Simulation des Prozesses enthalten. Die Erstellung der Bibliothek mit allen Komponenten, die zur Abbildung des Gesamtprozesses notwendig sind, ist auf der Grundlage der in der ThermoPower-Bibliothek enthaltenen Standardschnittstellen zum standardisierten Datenaustausch erfolgt. Es kommen Standardschnittstellen zur Beschreibung von Gasen und zur Beschreibung mechanischer Kopplungen zum Einsatz. Die Verwendung der einheitlichen Schnittstellen ermöglicht die Entwicklung des Modells einer Komponente unabhängig vom verwendeten Fluid. Zur einheitlichen Beschreibung einer Fluidströmung in einem Punkt wird der Massenstrom w , der Druck p , die spezifische Enthalpie h und der Vektor der Massenanteile ξ verwendet. Das für die Abbildung des OXYCOAL-AC-Prozesses definierte Fluid besteht dabei aus den sieben Komponenten Stickstoff (N_2), Kohlendioxid (CO_2), Sauerstoff (O_2), Kohlenmonoxid (CO), Wasser (H_2O), Schwefeldioxid (SO_2) und Stickoxid (NO). Für die Beschreibung des Brennstoffstroms ist ein Konnektor definiert worden, der den Massenstrom w , die Temperatur T und die Zusammensetzung ξ der Kohle enthält. Die abgebildete Kohle setzt sich aus den Komponenten Kohlenstoff (C), Wasserstoff (H), Wasser (H_2O), Sauerstoff (O), Schwefel (S) und Stickstoff (N) zusammen. Die Verbindung mechanischer Komponenten wird über den Drehwinkel φ und das Drehmoment M realisiert.

Abb. 2: Struktur der OXYCOAL-AC-Bibliothek



Basierend auf den erarbeiteten Schnittstellen sind sukzessive dynamische Modelle der wichtigsten Prozesskomponenten erstellt worden. Die Modellierung der verschiedenen Komponenten des OXYCOAL-AC-Prozesses ist physikalisch auf der Grundlage von Massen-, Energie- und Momentenbilanzen erfolgt. Im Folgenden werden anhand des Modells der keramischen Membran und des Modells des Wärmeübertragers des in Abschnitt 2.1 vorgestellten Prozesses die Stärken der objektorientierten Modellierung und Simulation aufgezeigt.

KERAMISCHE HOCHTEMPERATURMEMBRAN

Die keramische Membran zur Luftzerlegung stellt das zentrale Element im OXYCOAL-AC-Prozess dar. Der Betrieb der empfindlichen keramischen Komponente ist eine besondere regelungstechnische Herausforderung, da zum einen die Randbedingungen Druck, Temperatur und Massenstrom rauchgas- und luftseitig für einen optimalen Betrieb eingestellt werden müssen und zum anderen gewährleistet sein muss, dass die Membran sicher betrieben werden kann. Ein besonderer Aspekt ist dabei die Vermeidung hoher Temperaturgradienten, da diese zu thermischen Spannungen führen, welche die Membran zerstören können. Die Gradienten können sich entweder durch zeitliche Änderungen der umgebenden Temperaturen (Rauchgas, Luft) oder räumlich, z. B. längs der Strömungsrichtung, ausbilden. Die Problematik, kritische Temperaturgradienten zu vermeiden, macht ein Modell der Membran notwendig, mit welchem diese realitätsnah abgebildet werden können.

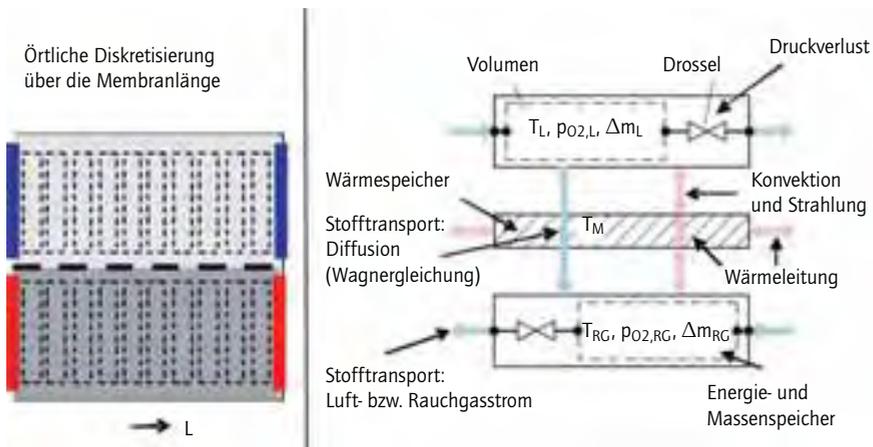
Generell können bei der Modellierung physikalischer Systeme konzentrierte und verteilte Modelle unterschieden werden. In konzentrierten Modellen wird das System über Größen beschrieben, welche unabhängig vom Ort sind und somit eine Beschreibung durch gewöhnliche Differentialgleichungen erlauben. Zur Abbildung eines Systems mit ortsabhängigen Größen sind die beschreibenden Gleichungen partielle Differentialgleichungen. Die Modellierung verteilter Systeme durch partielle Differentialgleichungen ist in der Regel aufwändig, eine analytische Lösung der entsprechenden Gleichungen im Allgemeinen nicht möglich. Einen Ansatz, ein verteiltes System ohne partielle Differentialgleichungen zu modellieren, bietet die Finite-Volumen-Methode (Casella2003b). Dabei wird das durch partielle Differentialgleichungen beschriebene verteilte System in geeigneter Weise örtlich diskretisiert und jedes finite Element durch ein System gewöhnlicher Differentialgleichungen beschrieben.

Zur Abbildung der Temperaturgradienten in Längsrichtung der Membran wurde ein Finite-Volumen-Ansatz gewählt. Durch eine örtliche Diskretisierung über die Membranlänge kann so neben der zeitlichen Temperaturänderung auch die örtliche Temperaturverteilung längs der Membran näherungsweise abgebildet und untersucht werden. Abb. 3 zeigt im linken Teil schematisch die örtliche Diskretisierung des Membranmoduls. Luft (oben) und Rauchgas (unten) werden in dieser Vorstellung durch einen permeablen Membranwerkstoff voneinander getrennt. Das Membranmodul kann durch einen Para-

meter in N diskrete Volumen in Strömungsrichtung unterteilt werden. Für jedes dieser Volumen werden die Erhaltungsgleichungen für Masse und Energie aufgestellt. Die Gesamtmembran wird durch eine automatische Verknüpfung der Einzelvolumina realisiert. Bei der Modellierung der Membran ist durchgängig ein physikalisch motivierter Ansatz verwendet worden, um eine Übertragung auf andere Geometrien und Materialien zu ermöglichen. Die Anzahl der diskreten Volumina ist wählbar. Ebenfalls wählbar ist das zu verwendende Gleichungssystem für ein einzelnes Volumen. Der von Modelica unterstützte Mechanismus der Polymorphie bietet die Möglichkeit des Überladens des im Membranmodul definierten Modells eines finiten Volumens. Durch die Umsetzung einer Auswahl zwischen Gleichungssystemen unterschiedlicher Komplexität wird ein hohes Maß an Flexibilität in Bezug auf Genauigkeit und Rechenaufwand erreicht.

Im rechten Teil von Abb. 3 ist das Modell eines Membranvolumenelements gezeigt. Die Modelle der einzelnen finiten Membranvolumina sind aus verschalteten Subkomponenten aufgebaut. Dies entspricht dem Bibliotheksgedanken und erleichtert die Wiederverwendbarkeit der einzelnen Komponentenbausteine. Das Modell eines Membranvolumenelements enthält rauchgas- wie luftseitig ein Volumen, das als Energie- und Massenspeicher wirkt, sowie eine Drossel, die den Druckverlust in Strömungsrichtung abbildet. Der Membranwerkstoff zwischen den beiden Volumina fungiert als Wärmespeicher. Zur Beschreibung des Wärmetransports im Membranmodul beinhaltet das komplexeste wählbare Gleichungssystem neben Wärmeleitungseffekten in Längs- und Querrichtung auch auf der Grundlage der Strömungsgegebenheiten berechnete konvektive Wärmeübergänge auf der Luft- und der Rauchgasseite sowie Wärmeaustausch in Folge von

Abb. 3: Modellierung des Membranmoduls



Strahlung. Einfachere Modellierungen verzichten beispielsweise auf die Abbildung der Wärmeleitungseffekte in Längsrichtung oder betrachten zur Berechnung des Wärmeübergangs lediglich den konvektiven Anteil und vernachlässigen die Strahlung.

Der Massentransport durch die Membran wird durch die sogenannte Wagnergleichung

$$j'' = \frac{C_1}{d} T e^{\left(\frac{C_2}{T}\right)} \ln \left(\frac{p_{O_2,L}}{p_{O_2,RG}} \right)$$

für die Diffusion ([12]) beschrieben. Dabei beschreibt j'' den flächenbezogenen Sauerstoffstrom durch die Membran, d die Membrandicke, T die Membrantemperatur und $p_{O_2,L}$ und $p_{O_2,RG}$ die Sauerstoffpartialdrücke auf der Luft- und auf der Rauchgasseite. C_1 und C_2 sind Stoffkonstanten zur Beschreibung der Diffusionseigenschaften.

GAS-GAS-WÄRMEÜBERTRAGER

Der Gas-Gas-Wärmeübertrager dient zum zusätzlichen Aufwärmen der verdichteten Luft durch das heiße Rauchgas. Wie schon bei der Membran handelt es sich auch beim Wärmeübertrager um ein verteiltes System, das durch einen Finite-Volumen-Ansatz beschrieben werden kann. Da sich der Wärmeübertrager im Wesentlichen nur dadurch von der Membran unterscheidet, dass kein Stofftransport stattfindet, werden bei der Modellierung des Wärmeübertragers die schon bei der Modellierung der Membran genutzten Subkomponenten zur Beschreibung der Strömungsvorgänge und der Wärmeübergänge verwendet. Abb. 4 zeigt das Dymola-Modell des Wärmeübertragers. Rauchgasseitig wie luftseitig wird die speichernde Wirkung des Wärmeübertragers über die Subkomponente Volumen abgebildet, wobei es sich bei den beiden Volumina um Instanzen derselben Klasse handelt, die unterschiedlich parametrisiert sind. Die beiden Volumina sind durch eine Rohrwand voneinander getrennt. Der Wärmeübergang von der Rauchgasseite über die Rohrwand hin zur Luftseite des Wärmeübertragers wird durch die drei Phänomene Konvektion, Strahlung und Wärmeleitung bestimmt. Das Rauchgas überträgt durch Überlagerung von konvektivem Wärmeübergang und Strahlung in Abhängigkeit von der Temperatur und den Strömungsbedingungen im rauchgasseitigen Volumen einen Wärmestrom an die Rauchgasseite der Rohrwand. Durch Wärmeleitung im Rohr überträgt sich der Wärmestrom hin zur Luftseite der Rohrwand. Hier ergibt sich wiederum durch Konvektion und Strahlung ein Wärmestrom in das luftseitige Volumen.

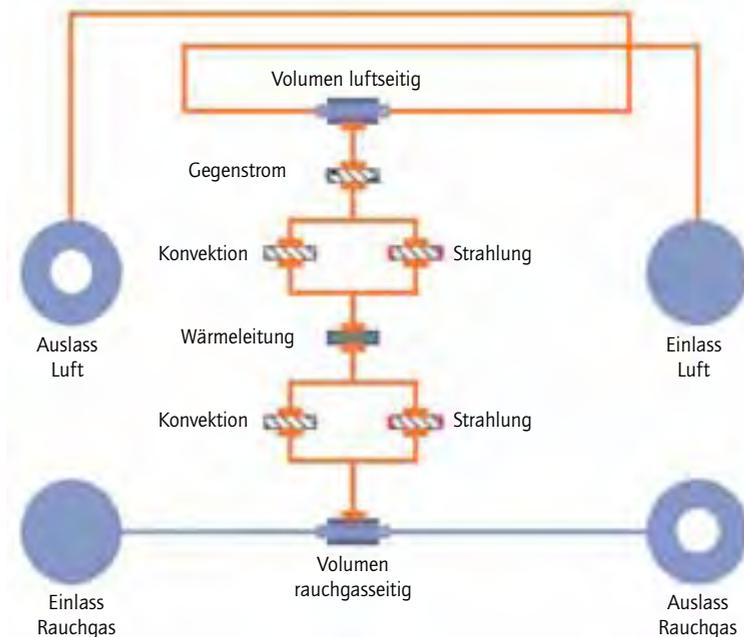
2.4 ERGEBNISSE

Der zuvor beschriebene Oxyfuel-Kraftwerksprozess ist Schwerpunkt eines aktuellen Forschungsprojekts und existiert bisher weder in seiner Gesamtheit, noch in allen seinen Komponenten. Die Auslegung des Gesamtprozesses ist in stationären Prozessrechnungen mit dem Entwicklungswerkzeug *Epsilon* erfolgt (vgl. [8]). In Ermangelung an Daten sind für dynamische Simulationen des zukünftigen OXYCOAL-AC-Kraftwerks zunächst

die Modelle der Komponenten an die Ergebnisse der stationären Prozessberechnungen angepasst worden. Die unbekanntenen Modellparameter sind so bestimmt worden, dass das dynamische Modell im Arbeitspunkt den gleichen stationären Auslegungspunkt bezüglich der Dampfparameter und sonstiger Massenströme, Drücke und Temperaturen aufweist. Hierfür ist die *Calibration-Toolbox* der kommerziellen Entwicklungs- und Simulationsumgebung Dymola verwendet worden. Diese stellt Identifikationsverfahren zur Verfügung, mit denen auswählbare Modellparameter so angepasst werden können, dass bestimmte Prozessgrößen vorgegebenen Trajektorien möglichst gut folgen. Im Falle des Membranmoduls sind beispielsweise die Beiwerte der Drosseln an die Druckverluste aus den stationären Berechnungen angepasst worden.

Im Anschluss an die Identifikation der Komponentenmodelle ist mit einem Modell des Gesamtprozesses eine sogenannte Sensitivitätsanalyse durchgeführt worden. Hierbei werden einzelne Eingangsgrößen, wie zum Beispiel die Drehzahlen der Verdichter und Gebläse, nacheinander sprungförmig oder rampenförmig erhöht und die Reaktionen verschiedener Ausgangsgrößen, wie Massenströme, Temperaturen, Drücke, etc., über

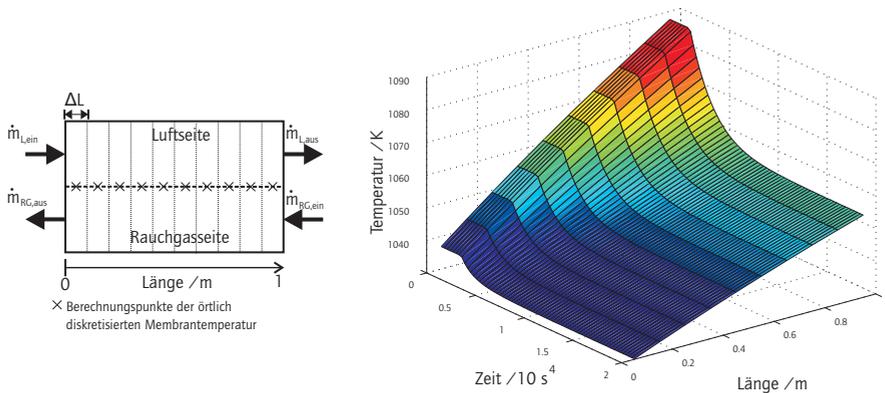
Abb. 4: Dymola-Modell des Wärmeübertragers



der Zeit aufgezeichnet. Aus den Sensitivitätsuntersuchungen können wichtige Informationen über die Kopplungen von Ein- und Ausgangsgrößen und die Zeitkonstanten des Oxyfuel-Kraftwerks abgeleitet werden. Sie stellen daher eine wesentliche Grundlage für die Entwicklung und Auslegung eines Regelungskonzepts dar.

In Abb. 5 sind die zeitlichen Verläufe der Membrantemperaturen der finiten Volumen als Reaktion auf eine rampenförmige Reduktion der Temperatur des Rauchgasmassenstroms am Membraneintritt dargestellt. Das Membranmodul besitzt im Fall der dargestellten Untersuchung eine Fläche von einem Quadratmeter bei einer Länge von einem Meter und ist bei der Modellierung in zehn diskrete Volumina unterteilt worden (siehe Abb. 5, links). Die Temperatur des Rauchgasmassenstroms ist bei $t=2000$ s beginnend innerhalb von 100 Sekunden um 50 Kelvin linear verringert worden. In der Abbildung ist zu erkennen, dass sich zum einen die Temperatur der Membran aufgrund der modellierten Wärmeübergänge und Wärmespeicher zwischen Rauchgas und Membran zeitlich verzögert verringert, und dass sich zum anderen ein Temperaturgefälle entlang der Strömungsrichtung des Rauchgases ausbildet. Durch die gewählte Modellierung können also sowohl zeitliche als auch örtliche Phänomene abgebildet werden. Die Information über die zeitlichen Verläufe der Temperaturen der einzelnen diskreten Membranvolumen nach einer Veränderung der Luft- bzw. Rauchgastemperatur kann bei der Entwicklung eines Regelungskonzept verwendet werden, um mit der Regelung hohe Temperaturgradienten zu vermeiden.

Abb. 5: Örtlicher und zeitlicher Verlauf der Membrantemperatur



2.5 FAZIT

Die objektorientierte Modellierungsmethode mit Modelica bietet einen flexiblen Ansatz zur Abbildung von Kraftwerksprozessen. Insbesondere für die Modellierung des OXYCOAL-AC-Prozesses haben sich der komponentenbasierte Ansatz und die Verwendung bereits vorhandener Modelle aus frei verfügbaren Bibliotheken bewährt. Als besonderer Vorteil ist der Einsatz standardisierter Schnittstellen zu nennen, der eine flexible Verschaltung der Komponentenmodelle zu Gesamtprozessmodellen unterschiedlicher Strukturen erlaubt. Die auf der Modelica-Media Bibliothek basierenden Schnittstellen für Fluide der frei verfügbaren ThermoPower-Bibliothek ermöglichen die Formulierung des Gleichungssystems einer Komponente unabhängig vom verwendeten Fluid und erlauben somit die Wiederverwendung einzelner Komponenten in anderen Kontexten. So können im OXYCOAL-AC-Prozess sowohl der Luftverdichter, als auch das Umwälzgebläse mit dem gleichen Modell eines Verdichters, nur unterschiedlich parametrisiert, abgebildet werden, obwohl sie von verschiedenen Fluiden durchströmt werden. Die Bibliothek der entwickelten Modelle für den Rauchgaskreislauf ist somit nicht nur auf den OXYCOAL-AC-Prozess beschränkt, sondern kann auch darüber hinaus flexibel eingesetzt werden.

Der komponentenbasierte Modellierungsansatz unterstützt die Modellierung verteilter Systeme. Ist einmal das Modell eines finiten Volumens erstellt, kann in einem überlagerten Modell leicht die Verknüpfung einer beliebig parametrisierbaren Anzahl finiter Volumen zu einem Komponentenmodell realisiert werden. Damit, und weiterhin mit der Möglichkeit, das Gleichungssystem einer Komponente zu überladen, können mit den gleichen Modellen Simulationen mit unterschiedlichem Detaillierungsgrad durchgeführt werden.

Die Möglichkeit der flexiblen Verschaltung der Komponenten ermöglicht zusätzlich die Untersuchung verschiedener Prozesskonfigurationen oder den Aufbau von Teilprozessmodellen. Dies ist insbesondere für die Entwicklung von Regelungskonzepten interessant, da auf diese Weise verschiedene Regelungsszenarien wie zum Beispiel der Anfahrvorgang des OXYCOAL-AC-Kraftwerks untersucht werden können.

3 ANWENDUNG IM BEREICH DER VERBRENNUNGSMOTOREN

3.1 MODELLBASIERTE REGELUNGEN IM VERBRENNUNGSMOTOR

Die stetig sinkenden Grenzwerte für Emissionen von Verbrennungsmotoren erfordern in zunehmendem Maße die Verbesserung des Verbrennungsprozesses sowie der damit verbundenen Regelungen. Für den Dieselmotor – der hier als Beispiel dienen soll – stehen insbesondere Stickoxid- (NO_x) und Rußemissionen im Vordergrund. Zur Reduzierung der NO_x-Emissionen hat sich die Abgasrückführung (AGR) bewährt, bei der verbranntes Abgas vor der Turbine entnommen wird und zum Frischluftmassenstrom nach dem Kompressor zurückgeführt wird. Für diese Systeme ist eine Regelung unverzichtbar, da ansonsten der Verbrennungsprozess nicht optimal ablaufen kann und somit die Emis-

sionen sogar steigen können. Ein sehr gut geeigneter Regelungsansatz ist der Modellgestützte Prädiktive Regler, der intern ein einfaches Modell der Regelstrecke verwendet, um so die bestmögliche Stellgröße zu ermitteln. Für die Entwicklung von Reglern hat sich Matlab/Simulink als Standard etabliert, da die signalorientierte Entwicklung dem Regelungstechniker entgegenkommt.

Für eine kostengünstige Entwicklung von Reglern im Allgemeinen ist die Simulation der Regelstrecke enorm wichtig und sie ist daher ein existentieller Bestandteil jeder Reglerentwicklung geworden. Obwohl Simulink auch zur Simulation von Systemen verwendet werden kann, ist die Signalorientierung nicht immer ein Vorteil. Komplexe Systeme werden so schnell unübersichtlich und die Fehleranfälligkeit steigt an. Aus diesem Grund tritt zunehmend die objektorientierte Modellierungssprache Modelica in Erscheinung, die für komplexe Systeme und domänenübergreifende Simulationen ideale Voraussetzungen bietet.

In diesem Abschnitt sollen daher an den beiden genannten Beispielen, der Entwicklung eines reglerinternen Modells für den Modellgestützten Prädiktiven Regler in Simulink und der Entwicklung eines Simulationsmodells für einen Dieselmotor mit der Modellierungssprache Modelica im Simulationstool Dymola, zwei wesentliche Unterschiede zwischen den beiden Modellierungsarten erläutert werden. Die Modelle, die mit Simulink für den Regler erstellt werden, besitzen einen kleineren Detaillierungsgrad als die Modelle, die für die Simulation der Regelstrecke verwendet werden. Der prinzipielle Unterschied ist trotz alledem ersichtlich und wird mit steigender Komplexität noch größer. Der Einsatz von Simulink für die Reglerentwicklung bleibt davon unberührt.

Für diesen Vergleich werden dazu zunächst in Abschnitt 3.2 der schematische Aufbau des Dieselmotors und die Regelungs- bzw. Simulationsaufgabe näher erläutert. Anschließend wird in Abschnitt 3.3 beispielhaft auf die unterschiedlichen Modellierungsarten eingegangen. In Abschnitt 3.4 wird dann ein Fazit der Erfahrungen gezogen.

3.2 AUFBAU DER REGELSTRECKE

Um die stetig steigenden Anforderungen an einen Verbrennungsmotor zu erfüllen, wird die Regelung und Simulation während des Entwicklungsprozesses immer wichtiger. Anhand eines Beispiels der Reglerentwicklung für den Luftpfad eines Dieselmotors soll gezeigt werden, welche Modelle entstehen und welche Tools dazu verwendet werden.

Dazu ist in Abb. 6 der schematische Aufbau des zu modellierenden Dieselmotors dargestellt. Es handelt sich dabei um einen Sechszylinder-Dieselmotor¹ in V-Anordnung mit Common-Rail-Direkteinspritzung. Der Luftpfad des Motors besteht aus einer Hochdruck-Abgasrückführung (AGR) zur Stickoxidminimierung und einem Turbolader mit va-

¹ Der hier beschriebene Motor wird in dem vom BMWI geförderten Projekt ZAMOMO verwendet, in welchem die Verzahnung zwischen den Entwurfsmethodiken für Embedded-Software und Regelungstechnik am Beispiel einer Motorsteuerung untersucht und vorangetrieben werden (siehe [6, 11]).

Um die entwickelten Regler frühzeitig testen zu können, wird zeitgleich ein detaillierteres Simulationsmodell der Regelstrecke erstellt. Mit Hilfe dieses Modells können schon vor Einsatz am realen Motor die Funktionen des Reglers getestet werden. Um das reale Steuergerät in einem Hardware-in-the-Loop-Test (HIL) testen zu können, müssen diese Modelle zusätzlich echtzeitfähig sein. Aus diesem Grund werden meist Mittelwertmodelle für die Luftpfadsimulation eingesetzt, die das Verhalten des Motors über einen Zyklus gemittelt berechnen. Für die Erstellung dieser Modelle wird nun Modelica als objektorientierte Modellierungssprache zur Modellerstellung eingesetzt.

Da für das Reglermodell und für das Simulationsmodell die gleiche Regelstrecke modelliert wird, kann die Modellentwicklung in Simulink sehr gut mit der Entwicklung des Modelica-Modells verglichen werden. Beide Modelle berücksichtigen nur die für die Regelung des Luftpfades relevanten dynamischen Anteile² des Gesamtsystems. Die unterschiedlichen Detaillierungsgrade der Modelle müssen beim Vergleich berücksichtigt werden. So ist auch die Rechenkapazität eines Steuergerätes nicht mit einer echtzeitfähigen Simulationsplattform vergleichbar.

3.3 VERGLEICH ZWISCHEN OBJEKTORIENTIERTER UND SIGNALORIENTIERTER MODELLIERUNG AM BEISPIEL VON LUFTPFADKOMPONENTEN

In diesem Abschnitt werden am Beispiel der Modellierung eines Volumens und eines Turboladers zwei wichtige Unterschiede zwischen der signalorientierten und der objektorientierten Modellierung aufgezeigt. Diese beiden Elemente wurden u. a. ausgesucht, da sie die beiden wesentlichen Elemente repräsentieren, die für das dynamische Verhalten des Luftpfades verantwortlich sind.

VOLUMEN

Ein Volumen wird für Mittelwertmodelle meist nulldimensional in den Größen Druck und Temperatur modelliert. Dynamische Wellenvorgänge in den Röhren und Behältern werden vernachlässigt. Das Volumen ist das Element, welches bei der Simulation eines Luftpfades am häufigsten eingesetzt wird. Die Modellierung sollte daher möglichst wiederverwendbar und einfach parametrierbar sein.

Die vereinfachten Basisgleichungen für Materie und Energie lauten für ein adiabates Volumen:

$$\frac{dm_V}{dt} = \dot{m}_{\text{Ein}} - \dot{m}_{\text{Aus}} = 0$$

$$\frac{dH_V}{dt} = \dot{H}_{\text{Ein}} - \dot{H}_{\text{Aus}} + \underbrace{\dot{W} + \dot{Q}}$$

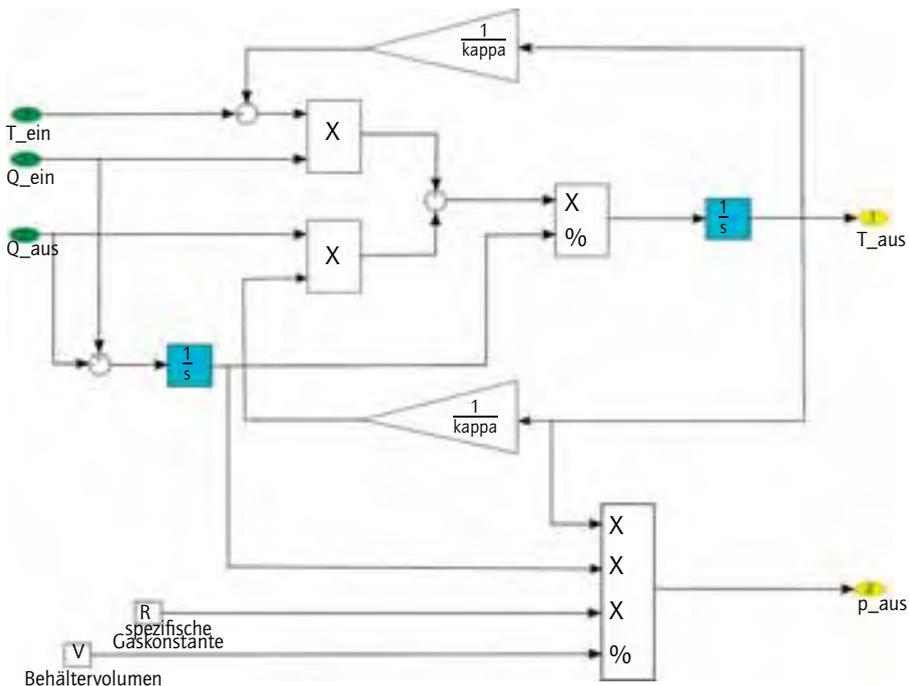
² Unter dem Begriff Dynamik werden im Folgenden auch nur diese Anteile verstanden. Höherfrequente Anteile wie z. B. Resonanzen im Saugrohr werden für die Regelung des Luftpfades nicht berücksichtigt.

Diese Gleichungen – die grundsätzlich für verschiedene Strömungsrichtungen gültig sind – sind den beiden Modellen in Simulink und Modelica gemein und müssen entsprechend umgesetzt werden.

In Simulink ist dies für eine festgelegte Strömungsrichtung noch kein Problem, wie in Abb. 7 ersichtlich ist. Aufgrund der Signalorientierung müssen jedoch bereits drei Signalpfade (ein- und austretender Massenstrom und die Temperatur am Eingang) als Eingangsgrößen und zwei Signalpfade (Temperatur am Austritt und Druck im Volumen) als Ausgangsgrößen mit den umgebenden Komponenten verbunden werden. Insgesamt sind so fünf Signalpfade für dieses einfache Modell des Volumens für jede Instanz der Komponente zu verbinden.

Weiterhin ist an der Umsetzung des Simulink-Modells auch ersichtlich, dass eine Umkehrung der Strömung mit diesem Modell nicht möglich ist. Dazu muss die Temperatur des in das Volumen eintretenden Massenstroms je nach Strömungsrichtung berücksichtigt werden. Dies bedeutet für dieses Modell, dass auch die Ausgangstemperatur als Eingangsgröße zur Verfügung stehen und eine Abfrage über die Strömungsrichtung

Abb. 7: Simulink-Modell des einfachen Volumens



erfolgen muss. Durch diese zusätzlichen Komponenten und Signalpfade lässt sich die ursprünglich einfache Gleichung nicht mehr leicht erkennen.

Im Allgemeinen kann bei einem Volumen zwischen vier Strömungsrichtungen³ unterschieden werden. Noch komplexer wird es, wenn nicht nur Reinstoffsysteme, sondern Mehrstoff- oder Mehrphasensysteme betrachtet werden. Für diese Fälle sind anhand der Signalverläufe des Volumen-Modells die relativ einfachen Grundgleichungen nicht mehr ersichtlich. Zur Dokumentation des Modells reicht die Modell-Datei alleine nicht mehr aus. Es ist erforderlich, eine separate Datei zu pflegen, in der die entsprechenden Gleichungen und Annahmen festgehalten werden. Für die relativ einfachen Gleichungen des reglerinternen Modells ist dies noch mit vertretbarem Aufwand möglich.

In Modelica wird durch die Verwendung von *Konnektoren* und der akausalen Modellierung das Problem der Nachvollziehbarkeit und Wiederverwendbarkeit elegant gelöst. Konnektoren fassen alle Größen zusammen, die über eine Systemgrenze zwischen Komponenten ausgetauscht werden. Dabei wird zwischen Potential- und Flussgrößen unterschieden. Zusätzlich können hier direkt verschiedene Strömungsrichtungen berücksichtigt werden. Als Beispiel sei hier der Konnektor der ThermoPower-Bibliothek (siehe [2]) genannt, der für thermische Systeme die Größen *Massenstrom*, *Druck*, *spezifische Enthalpie* und *Stoffzusammensetzung* definiert. Für die beiden letztgenannten Variablen wird zusätzlich zwischen einer eintretenden und einer austretenden Größe unterschieden.

Für das Volumen werden mit dieser Definition des Konnektors ein Eingangs- (inlet) und ein Ausgangs-Konnektor (outlet) definiert. Aufgrund der Definition des Massenstroms als Flussgröße kann die Gleichung für die Massenerhaltung sehr einfach implementiert werden (siehe dritte Zeile der Basisgleichungen in Abb. 8). Die verschiedenen Strömungsrichtungen der beiden Konnektoren (inlet und outlet) werden automatisch mit richtigem Vorzeichen in die Gleichung eingetragen. Für die Energiegleichung wird jeweils eine Variable für den ein- und austretenden spezifischen Enthalpiestrom definiert, die in Abhängigkeit der Strömungsrichtung besetzt werden. Dies geschieht übersichtlich mit Hilfe von zwei bedingten Abfragen (siehe Abb. 8 rechts), die den Variablen *hi* und *ho* die zugehörige Größe des Konnektors zuweisen. Für die Berücksichtigung von Mehrstoffkomponenten wird zusätzlich die Stoffzusammensetzung benötigt. Diese kann innerhalb der bestehenden bedingten Abfragen implementiert werden. Die Übersichtlichkeit und Nachvollziehbarkeit dieser Lösung ist der Lösung in Simulink deutlich überlegen.

TURBOLADER

Eine zweite wichtige Komponente für die Dynamik des Luftpfades ist der Turbolader. Er entzieht mit Hilfe einer Turbine dem heißen Abgas Energie und führt diese dem

³ Kombination aus jeweils zwei Strömungsrichtungen für den Eingang und Ausgang.

Abb. 8: Ausschnitt des Volumen-Modells der ThermoPower-Bibliothek [2]

Basisgleichungen

```

M = gas.d * V "Gas mass";
E = M*gas.u "Gas internal energy";
der(M) = inlet.w + outlet.w "Mass balance";
der(E) = inlet.w*i + outlet.w*ho "Energy
                                balance";

```

Bedingte Abfragen

```

if inlet.w >= 0
then
  hi = inlet.hBA;
  Xi_i = inlet.XBA;
else
  hi = gas.h;
  Xi_i = gas.Xi;
end if;

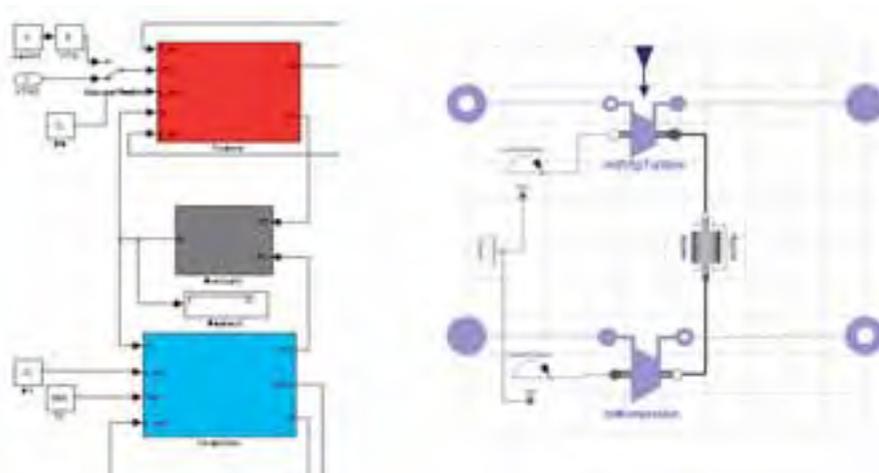
if outlet.w >= 0
then
  ho = outlet.hAB;
  Xi_o = outlet.XA;
else
  ho = gas.h;
  Xi_o = gas.Xi;
end if;

```

Kompressor zu, der damit den Frischluftmassenstrom verdichtet. Bei Dieselmotoren wird häufig ein Turbolader mit variabler Turbinengeometrie (VTG) eingesetzt, mit dessen Hilfe die Turbine besser an den jeweiligen Betriebspunkt des Motors angepasst werden kann. Dieses Verfahren führt zu höheren Wirkungsgraden und der Ladedruck kann mit Hilfe der VTG-Stellung in gewissen Grenzen geregelt werden.

Für die Simulation wird der VTG-Turbolader mit Hilfe von Kennfeldern modelliert, die die Charakteristik von Turbine und Kompressor beschreiben. In Abb. 9 ist der Vergleich der Modellierungen des Turboladers in Simulink und in Modelica bzw. im Simula-

Abb. 9: Vergleich der Turboladermodellierung in Simulink (links) und Dymola (rechts)



tionstool Dymola dargestellt. Auf dieser Ebene fallen zunächst wieder die vielen Signalpfade des Simulink-Modells auf. Die Struktur des Dymola-Modells ist sehr viel übersichtlicher und entspricht direkt der physikalischen Realität.

Neben diesen bereits erwähnten Unterschieden in der Darstellungsform von physikalischen Systemen liegt ein weiterer wesentlicher Unterschied zwischen der signalorientierten und der objektorientierten Modellierung darin, dass die mathematischen Umformungen der Gleichungen nicht mehr von Hand durchgeführt werden müssen. Diese Aufgabe wird bei der objektorientierten Modellierung vom Simulationstool übernommen. Der Vorteil dieser Möglichkeit soll am folgenden Beispiel erläutert werden.

Zur Modellierung des Kompressors werden die üblichen Erhaltungsgleichungen für Materie und Energie sowie einige Kenngrößen für Turbomaschinen verwendet. Zur Beschreibung der Charakteristik des Kompressors werden zwei Kennfelder benötigt. Das erste beschreibt den Wirkungsgrad des Kompressors in Abhängigkeit von einer reduzierten Turboladerdrehzahl und einem reduzierten Frischluftmassenstrom. Das zweite Kennfeld beschreibt je nach Modellierungsart das Druckverhältnis des Kompressors in Abhängigkeit von einem reduzierten Massenstrom und einer reduzierten Turboladerdrehzahl oder den reduzierten Massenstrom in Abhängigkeit von einer reduzierten Turboladerdrehzahl und des Druckverhältnisses. Die beiden Modellierungsarten unterscheiden sich also nur in den Ein- und Ausgangsgrößen. Die Basisgleichungen sind identisch.

Für ein Simulink-Modell des Kompressors bedeutet dies, dass für jede Möglichkeit ein eigenständiges Modell erstellt werden muss. Die Wiederverwendbarkeit bestehender Komponenten, insbesondere der Basisgleichungen, ist im Rahmen der signalorientierten Modellierung nicht gewährleistet. Der Anwender muss je nach Ein- und Ausgangsgrößenwahl die gesamten Gleichungen neu erstellen.

Im Gegensatz dazu können innerhalb der objektorientierten Modellierung die kompletten Gleichungen wieder verwendet werden. Das Einzige, was sich zwischen den beiden Modellierungsarten unterscheidet, sind die Kennfelder. Die Umstellung der Gleichungen nach den entsprechenden Ein- und Ausgangsgrößen wird vom Simulationstool intern vorgenommen. Auf gleiche Weise können auch einfache algebraische Gleichungen durch das Simulationstool aufgelöst werden. Diese automatische Umformung der Gleichungen besitzt jedoch Grenzen. So kann es durchaus von Vorteil sein, wenn der Nutzer bestimmte Formeln entweder selbst auflöst oder eine alternative Berechnungsvorschrift nutzt. Ein Beispiel aus [1] für einen einfachen elektrischen Schaltkreis mit einem nichtlinearen Widerstand soll dies beispielhaft zeigen. Dabei wird der Widerstand über die Gleichung

$$i = a \cdot \tanh(b \cdot u)$$

beschrieben. Im Modell ist die Stromstärke i eine Zustandsgröße, daher muss diese Gleichung nach u aufgelöst werden. Dymola kann diese Gleichung nicht analytisch auflösen und wird daher ein numerisches Verfahren wählen. Eine Alternative ist, eine andere De-

fnitions-gleichung für den Tangens Hyperbolicus zu wählen,

$$\tanh(x) = 1 - \frac{2}{e^{2x} + 1},$$

und die Gleichung direkt in Abhängigkeit der Zustandsgröße Stromstärke aufzustellen.

$$u = \frac{1}{2b} \cdot \log \left(\frac{1 + i/a}{1 - i/a} \right)$$

Ein Vergleich der beiden Implementierungen zeigt, dass die zweite Lösung bei identischem Ergebnis deutlich schneller ist, da das numerische Auflösen der Gleichung entfällt.

An diesem Beispiel wird klar, dass die automatischen Verfahren zum Auflösen der Gleichungen nicht immer die bessere Lösung liefern, sondern der Anwender positiv wie auch negativ auf die Performance eingreifen kann.

3.4 FAZIT

Im Vergleich zwischen der signalorientierten und der objektorientierten Modellierung hat sich gezeigt, dass die objektorientierte Modellierung viele Vorteile gegenüber der signalorientierten Modellierung besitzt. Die Nähe der Struktur der Modelica-Modelle zur physikalischen Realität lassen den Benutzer die Modelle leichter verstehen. Die fehleranfälligen Umformungen von mathematischen Gleichungen für die signalorientierte Modellierung entfallen und geben somit dem Anwender mehr Spielraum. Für die Erstellung der Modelle ist im Gegenzug dazu eine Einarbeitung in die Gedankenwelt der objektorientierten Modellierung notwendig, die sich wie gezeigt deutlich von der signalorientierten Modellierung unterscheidet.

Für die Simulation komplexer Systeme hat sich die objektorientierte Modellerstellung bewährt, auch aufgrund der vielen frei verfügbaren Bibliotheken, die mittlerweile für die meisten Anwendungsdomänen existieren. Diese können von jedem Anwender frei genutzt und für seine eigenen Zwecke angepasst werden.

Für die Reglerentwicklung kann die signalorientierte Modellierung, wie sie in Matlab/Simulink eingesetzt wird, ihre Stärken gegenüber der objektorientierten Modellierung ausspielen. In diesem Bereich ist die objektorientierte Modellierung noch nicht konkurrenzfähig. Ein weiterer Vorteil von Simulink ist, dass dieses Programm mittlerweile zum Quasi-Standard für Co-Simulationen geworden ist. Fast jedes kommerzielle Simulationsprogramm besitzt eine Schnittstelle zu Simulink. Durch Nutzung beider Werkzeuge und deren datentechnischer Kopplung – auf die hier nicht näher eingegangen werden soll – können die Vorteile beider Welten elegant miteinander verbunden werden.

4 RESÜMEE

Nach erfolgten Exkursionen in die Kraftwerks- und Verbrennungsmotorentechnik sollen die dort projektspezifisch aufgezeigten Eigenschaften und Vorteile objektorientierter Modellierungs- und Simulationsansätze noch einmal aus der Vogelperspektive zusammenfassend betrachtet werden. Wesentliches Merkmal ist sicherlich der Verzicht auf Ursache/Wirkungsprinzipien, die nicht zuletzt deswegen so verbreitet sind, weil damit aus einer Gleichung programmtechnisch unmittelbar eine rückwirkungsfreie Zuweisung (in der Gleichung „von rechts nach links“ zu lesen) abgeleitet werden kann. Verweigert sich jedoch die abzubildende Realität einer solchen vereinfachenden rückwirkungsfreien Betrachtungsweise – was oft eher die Regel als die Ausnahme ist – so entstehen ungewollte Schleifen im Wirkungsgefüge, die vielfältige Probleme nach sich ziehen und die Behandlung komplexer Systeme beträchtlich erschweren.

Stellvertretend auch für andere Aspekte seien dagegen als wesentliche Vorteile objektorientierter Modelle deren Universalität und Wiederverwendbarkeit (u.a. in Hinblick auf verschiedene interessierende Ausgangsgrößen) sowie die Übersichtlichkeit der Schnittstellen genannt, womit der Aufbau von modularisierten Modellbibliotheken in ganz erheblichem Maße unterstützt wird. Möglich wird all dies durch erweiterte, nämlich algebraisch/symbolische Rechenfähigkeiten der ausführenden Rechnerwerkzeuge, die dem Anwender einen Gutteil dessen abnehmen, was er zuvor an intellektueller Leistung in die Modellbeschreibung einbringen musste. Besonders bemerkenswert ist die unverkennbare Förderung einer fachübergreifenden Kommunikation, die hierbei von der Mathematik geleistet wird: Vertreter unterschiedlicher Fachdisziplinen können sich wesentlich besser über Modellierungsfragen austauschen, da die Beschreibungsmittel von rechentechnischem Ballast befreit bleiben.

Abschließend verbleibt es mir, den Förderinstanzen der in diesem Beitrag näher betrachteten Forschungsprojekte zu danken. Mein besonderer Dank gilt weiterhin den wissenschaftlichen Mitarbeiterinnen und Mitarbeitern, insbesondere Sarah Dolls, Sebastian Hölemann und Frank Heßeler, die zu diesem Beitrag Inhalte aus ihren aktuell bearbeiteten Forschungsprojekten beigesteuert haben.

5 LITERATUR

- [1] Bachmann, Bernhard: Mathematical Aspects of Object-Oriented Modeling and Simulation, Tutorial auf der 6. Internationalen Modelica Konferenz, Bielefeld 2008.
- [2] Casella, F.; Leva, A.: Modelica open library for power plant simulation: design and validation. In: P. Fritzton (Hrsg.): Proceedings of the 3rd International Modelica Conference 2003, Linköpings, Schweden.

- [3] Casella, F.; Schiavo, F.: Modelling and Simulation of Heat Exchangers in Modelica with Finite Element Methods. In: P. Fritzson (Hrsg.): Proceedings of the 3rd International Modelica Conference. 2003 S. 343–352. Linköping, Schweden.
- [4] Casella F.; Otter, M.; Proelss, K.; Richter, C.; Tummescheit, H.: The Modelica Fluid and Media Library for Modeling of Incompressible and Compressible Thermo-Fluid Pipe Networks. In: Proceedings of the 5th International Modelica Conference 2006. S. 631–640. Wien, Österreich.
- [5] COORETEC: Forschungs- und Entwicklungskonzept für emissionsarme fossil befeuerte Kraftwerke, Bericht der COORETEC-Arbeitsgruppe, 2003.
- [6] Peter Drews, Frank Hesseler, Hoffmann, K.; Abel, D.; Schmitz, D.; Polzer, A., Kowalewski, S.: Entwicklung einer Luftpfadregelung am Dieselmotor unter Berücksichtigung nichtfunktionaler Anforderungen. In: AUTOREG 2008 – Steuerung und Regelung von Fahrzeugen und Motoren. 4. Fachtagung Baden-Baden, 12. und 13. Februar 2008. VDI-Berichte Nr. 2009, ISBN 978-3-18-092009-2.
- [7] Hölemann, S.; Nötges, T.; Abel, D.: Modellgestützte Prädiktive Regelung zur Sauerstoffversorgung eines oxyfuel Prozesses. In: GMA-Kongress 2007, 12.–13. 6. 2007, Baden-Baden, VDI-Berichte 1980, S. 729–738, VDI-Verlag, Düsseldorf.
- [8] Kneer, R.; Abel, D.; Niehuis, R.; Maier, H.R.; Modigell, M.; Peters, N.: Entwicklung eines CO₂-emissionsfreien Kohleverbrennungsprozesses zur Stromerzeugung, VDI-Berichte Nr. 1888 PL03, 2006.
- [9] Rückert, Joachim: Modellgestützte Regelung von Ladedruck- und Abgasrückführrate beim Dieselmotor, Dissertation 2004 RWTH Aachen.
- [10] Richert, Felix: Objektorientierte Modellbildung und Nichtlineare Prädiktive Regelung von Dieselmotoren, Dissertation 2005 RWTH Aachen.
- [11] Schmitz, D.; Drews, P.; Heßeler, F.; Jarke, M.; Kowalewski, S.; Palczynski, J.; Polzer, A.; Reke, M.; Rose, T. (2008): Modellbasierte Anforderungserfassung für softwarebasierte Regelungen. In: Software Engineering 2008 (Herrmann, K.; Brügge, B. (Eds.)). Fachtagung des GI-Fachbereichs Softwaretechnik, 18.–22. 3. 2008 in München. LNI 121 GI 2008, ISBN 978-3-88579-215-4.
- [12] Wagner, C.: Equations for transport in solid oxides and sulfides of transition metals. In: Progress in solid state chemistry Vol. 10 (1), 1975, S. 3–16.

> PLANUNGSWERKZEUGE FÜR ENERGIEEFFIZIENTE GEBÄUDE

DIRK MÜLLER

1 EXECUTIVE SUMMARY

In der Bundesrepublik Deutschland entfallen rund 30 % des Endenergiebedarfs alleine auf die Wärmeversorgung von Gebäuden [1]. Der Hauptanteil dieser Endenergie wird für die Beheizung von Wohngebäuden verwendet, wobei insbesondere Gas und Heizöl als Primärenergieträger eingesetzt werden.

Die Gebäudehülle und die Gebäudetechnik sind daher für die Minderungsmaßnahmen hinsichtlich des Energieeinsatzes und des Schadstoffausstoßes besonders interessant, da oft durch einfache Maßnahmen erhebliche Erfolge erreicht werden können. Die Mathematik leistet einen wichtigen Beitrag beim Planungsprozess eines Gebäudes. Durch unterschiedliche Berechnungswerkzeuge können drei wesentliche Phasen einer energetischen Optimierung begleitet werden: die Reduktion der thermischen Lasten durch eine günstige Ausführung, Ausrichtung und Form der Gebäudehülle, die Einbindung regenerativer Energien und die Betriebsoptimierung.

Mit der Gebäudehülle können viele Funktionen bereitgestellt werden, die für die Energieeffizienz und die Behaglichkeit der Nutzer essentiell sind. Eine ausreichende Dämmung sorgt für geringe Wärmeverluste des Baukörpers und gewährleistet angenehm hohe Wandinnentemperaturen, die gleichzeitig hygienischen Problemen wie der Schimmelpilzbildung entgegenwirken. Hier liefern zwei- und dreidimensionale Berechnungen von Wärmeleitvorgängen mit der Methode der finiten Elemente in Wandecken und geometrisch komplexen Bauteilen, wie beispielsweise einem Fensterrahmen, genaue Daten für die zu erwartenden Wärmeverluste.

Durch den Sonnenschutz vor transparenten Elementen kann eine Überhitzung der Innenräume im Sommer vermieden werden, gleichzeitig soll durch lichtlenkende Elemente Tageslicht in größere Raumtiefen blendfrei geleitet werden. Durch Tageslichtsimulationen kann die Lichtsituation auch in größeren Raumtiefen vorhergesagt und an die Erfordernisse angepasst werden. Hier kommen verschiedene Verfahren der Strahlverfolgung zum Einsatz, mit denen auch die diffuse Streuung des Lichts an Oberflächen und Mehrfachreflexionen berücksichtigt werden können.

An allen Wandelementen kann gezielt Wärme von außen zu- oder abgeführt werden, und durch Fenster oder andere Fassadenöffnungen werden die Innenräume mit

Außenluft versorgt. Diese Transportprozesse für Energie und feuchte Luft können modelliert und mit Hilfe mathematischer Verfahren zur Lösung gekoppelter Differentialgleichungen instationär berechnet werden. So kann der Temperaturverlauf in einem Raum prognostiziert und ein System für die Raumkonditionierung ausgelegt werden.

Moderne Gebäude werden heute oft nicht mehr durch eine einzige Energiequelle beziehungsweise einen Primärenergieträger versorgt. Stattdessen sorgen mehrere Prozesse zur Kälte- und Wärmeerzeugung für behagliche Innenraumtemperaturen. Zunehmend werden auch regenerative Energiequellen eingesetzt, die ergänzend oder in besonderen Fällen allein für eine CO₂ emissionsarme Energieversorgung des Gebäudes sorgen. Das komplexe Zusammenspiel aller Komponenten kann nur mit aufwendigen Berechnungswerkzeugen abgebildet werden, die wiederum auf numerischen Lösungsverfahren für Systeme gekoppelter Differentialgleichungen beruhen.

Und letztendlich wollen wir, die Nutzer, sich in einem Gebäude wohlfühlen. Ob wir eine Umgebung als angenehm und behaglich empfinden, hängt mit unserer direkten Umgebung, unserem Mikroklima zusammen. Die Lücke zwischen unserem Mikroklima und der Gebäudehülle beziehungsweise der Gebäudetechnik kann durch den Einsatz mathematischer Verfahren für die Berechnung von Raumluftströmungen geschlossen werden. Hierfür werden üblicherweise finite Volumenverfahren für die Lösung der Navier-Stokes-Gleichungen eingesetzt. So können bereits in der Planungsphase Vorhersagen zur Behaglichkeit der Nutzer in einem virtuellen Gebäude bereitgestellt werden.

Ohne die Anwendung mathematischer Verfahren ist eine integrale Planung eines Gebäudes heute nicht mehr denkbar. Jedes Gebäude ist ein Prototyp und es können, anders als bei anderen technischen Entwicklungsprozessen, leider keine Versuchsmuster gebaut und vermessen werden. Allein deshalb kommt allen Berechnungswerkzeugen eines virtuellen Planungsprozesses im Bauwesen eine besonders hohe Bedeutung zu.

2 ENTWICKLUNG EINES NEUEN KÜHLSYSTEMS MIT HILFE EINER THERMOHYDRAULISCHEN GEBÄUDESIMULATION

In immer mehr Bürogebäuden werden schon im Frühjahr und im Herbst hohe Innenraumtemperaturen erreicht, welche den Komfort und die Leistungsfähigkeit der Nutzer einschränken. Eine herkömmliche Kühlung oder Klimatisierung mit einer Kompressionskältemaschine verursacht einen hohen Energieverbrauch und Lastspitzen im Stromnetz, die in vielen Ländern schon für einen Ausfall des elektrischen Netzes geführt haben.

Im Rahmen eines Forschungsprojekts, gefördert durch das BMWi [2], konnte mit Industriepartnern ein neuartiges Kühlsystem entwickelt werden, das ohne den Einsatz einer Kältemaschine eine spürbare Absenkung der Raumtemperatur erreichen kann. Die Idee ist, die niedrigeren Temperaturen der Nacht durch den Einsatz eines Latentwärmespeichers tagsüber zur Kühlung zu nutzen. Das System besteht aus einem in die Fassade integrierten Latentwärmespeicher, dessen Phasenwechseltemperatur knapp unterhalb der gewünschten Raumtemperatur liegt, einer wasserdurchströmten Kühldecke im Raum

und einem Wärmeübertrager an der Außenseite der Fassade. Alle Komponenten sind über einen Wasserkreislauf verbunden. Tagsüber wird der Latentwärmespeicher unter Aufnahme von Wärme aus dem Raum aufgeschmolzen, nachts wird der Speicher durch Abgabe von Wärme an die Umgebung erstarrt. Für die erreichte Raumkühlung muss nur die notwendige Pumpenleistung für den Wasserkreislauf bereitgestellt werden.

2.1 EINSATZ DER OBJEKTORIENTIERTEN PROGRAMMIERSPRACHE MODELICA

Für die Entwicklung des neuen Kühlsystems wird die objektorientierte Programmiersprache Modelica eingesetzt, in der alle Komponenten des geplanten Ausbaus abgebildet werden können. Modelica ist eine physikalische Beschreibungssprache, die sich in mancher Hinsicht von bekannten Programmiersprachen unterscheidet. Als Beschreibungssprache enthält sie zwar Konstrukte zur Datenflusssteuerung (wie bedingte Anweisung, Schleife), aber beispielsweise kaum Funktionen zur direkten Kommunikation mit dem Benutzer. Der Grund hierfür liegt in dem genau festgelegten Anwendungsbereich der Sprache – hier geht es nicht um die interaktive Benutzung von Programmen, sondern um die Simulation im Voraus definierter physikalischer Systeme. Die Stärken der Sprache liegen dementsprechend besonders im Umgang mit mathematischen Gleichungen bzw. Gleichungssystemen.

Während herkömmliche Hochsprachen keine echten mathematischen Gleichungen verarbeiten können, sondern nur Zuweisungen kennen, die immer in eine Richtung aufgelöst werden, ist es in Modelica möglich, Gleichungen zu formulieren, die erst bei der Ausführung des Codes umgeformt und gelöst werden. Da sich während einer Simulation durchaus ändern kann, welche Variablen bekannt und welche unbekannt sind, wird auch eine entsprechende Gleichung unterschiedlich umgeformt. Dies ermöglicht eine sehr übersichtliche Beschreibung physikalischer Systeme.

2.2 AUSBAU EINES KOMPLEXEN SIMULATIONSMODELLS

Für die Berechnung des Kühlsystems mit Latentwärmespeichern muss ein komplexes Berechnungsmodell aufgebaut werden, in dem alle wesentlichen physikalischen Effekte berücksichtigt sind: das bauphysikalische Verhalten des Raums, die Wettereinflüsse, die Anlagentechnik und das Regelungskonzept. Im Mittelpunkt des Systems steht der Nutzer, der die inneren Lasten durch die Nutzung vorgibt und gleichzeitig das „Maß der Dinge“ ist, da seine Komfortansprüche befriedigt werden müssen. Da das Kühlsystem mit geringen Differenztemperaturen, kleiner als zehn Kelvin, arbeitet, müssen hohe Anforderungen an die Genauigkeit der Simulation gestellt werden.

Zur Beschreibung des thermischen Verhaltens eines Raums müssen unterschiedliche physikalische Prozesse betrachtet werden. Zunächst kann nach inneren und äußeren Wärmequelle unterschieden werden. Innere Quellen sind alle Wärmequellen (Menschen, Beleuchtung, Maschinen etc.) die sich innerhalb der Gebäudehülle befinden. Äußere

Wärmequellen werden durch die Umschließungsflächen in den Raum transportiert. Dazu zählen die direkte und diffuse Strahlung, der Wärmetransport durch die Umschließungsflächen und der Stofftransport (Luft, Materialien). Im Inneren des Raums müssen neben allen konvektiven Transportprozessen der langwellige Strahlungsaustausch zwischen den Umschließungsflächen und in vielen Fällen die Feuchtequellen berücksichtigt werden. Eine Zusammenstellung der wesentlichen Prozesse zeigt Abb. 2.

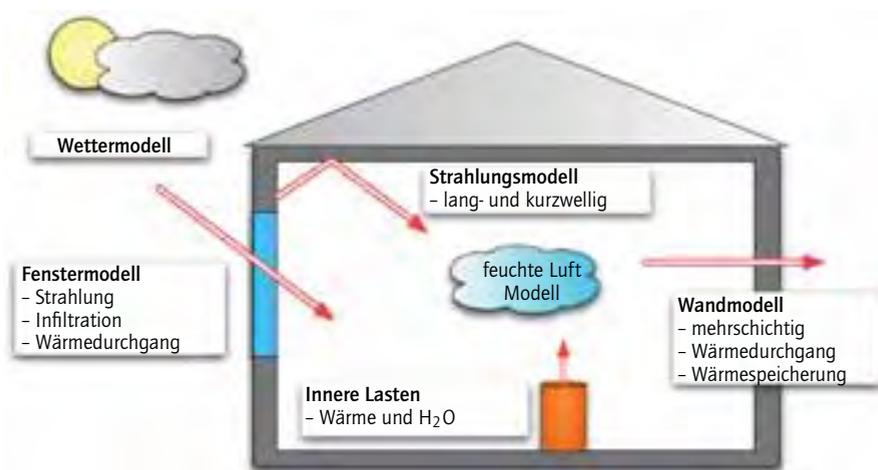
Für die Simulation der Anlagentechnik sind Modelle für Rohrleitungen, Pumpen, Ventile und Wärmeübertrager verfügbar, mit denen die Hydraulik des Kühlsystems abgebildet wird. Zusätzlich werden virtuelle Fühler und Regler eingesetzt, mit denen die Betriebsführung des Kühlsystems optimiert werden kann.

Für die Modellierung des Latentwärmespeichers war die Entwicklung eines Modells der Phasenumwandlung des Latentwärmespeichermaterials notwendig. Alle anderen Komponenten und physikalischen Prozesse konnten mit bekannten Ansätzen der Gebäude- und Anlagensimulation berücksichtigt werden. Viele in der Literatur beschrie-

Abb. 1: Komponenten einer dynamischen Gebäudesimulation



Abb. 2: Physikalische Prozesse in einem Raum

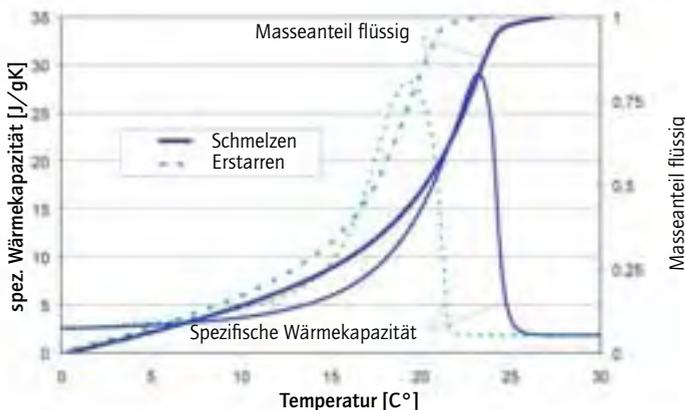


bene Modellierungen des Phasenwechsels basieren in der Regel entweder auf der analytischen Lösung des Phasenwechsels reiner Stoffe oder bei Stoffgemischen mit einem Phasenübergangsbereich auf unterschiedlich starken Vereinfachungen des Phasenwechselprozesses.

So gibt es Ansätze, die eine mittlere Enthalpie des Materials für den Temperaturbereich des Phasenwechsels vorsehen [3] oder lineare Funktionen, wie in [4] beschrieben, verwenden. Dagegen gibt es in der Literatur weniger Ansätze einer detaillierten Modellierung des physikalischen Verhaltens des Phasenwechsels von Stoffgemischen einschließlich möglicher Unterkühlung bzw. Hysterese des Materials. Vorschläge für die Beschreibung des Phasenwechsels über die Enthalpie als temperaturabhängige Funktion wurden im Rahmen dieses Forschungsprojektes veröffentlicht [5] oder auch in [6] beschrieben. Für eine realitätsnahe Abbildung des Phasenwechsels für die Auslegung von Latentwärmespeicher-Systemen und die Abschätzung ihrer Leistungsfähigkeit ist jedoch eine genaue Abbildung des Schmelz- und Erstarrungsvorganges unerlässlich.

Gerade bei der Nutzung natürlicher Wärmesenken für die Auskühlung des Speichers ist die genaue Abbildung des Erstarrungsverhaltens sehr wichtig. Wenige Kelvin Unterschied können die Gesamtleistung des Systems erheblich beeinträchtigen. Das Phasenwechselmodell in Modelica wurde ausgehend von der über DSC-Messungen (Differential Scanning Calorimetry – dynamische Differenzkalorimetrie) ermittelten temperaturabhängigen Enthalpie des Materials entwickelt, siehe Abb. 3. Für die Berechnung wird das Speichervolumen in finite Volumen diskretisiert. Das Modell bildet dabei sowohl die

Abb. 3: Aus DSC-Messungen bestimmte spezifische Wärmekapazität eines Phasenwechselmaterials und der berechnete Masseanteil des flüssigen Materials in Abhängigkeit von der Temperatur



Zustandsabhängigkeit der Stoffwerte Dichte, Wärmeleitfähigkeit und Wärmekapazität als auch die Temperaturabhängigkeit der spezifischen Wärmekapazität im Gebiet des Phasenwechsels ab.

Eine besondere Schwierigkeit ergibt sich aus der Hysterese der beim Schmelzen und Erstarren gemessenen spezifischen Wärmekapazität: Eine eindeutige Zuordnung zwischen Temperatur und spezifischer Wärmekapazität ist nur möglich, wenn die Vorgeschichte des Materials bekannt ist.

2.3 ERGEBNISSE DER BERECHNUNGEN

Mit Hilfe der Simulationen des neuen Kühlsystems können viele Fragestellungen bearbeitet werden, die experimentell nicht oder nur mit sehr großem Aufwand beantwortet werden können. Im Gegensatz zu einer experimentellen Untersuchung an einem Versuchsgebäude kann in der Simulation eine Wetterperiode wiederholt mit unterschiedlichen Systemparametern berechnet werden. Die Ergebnisse liefern einen direkten Vergleich unterschiedlicher Phasenwechsellmaterialien und können für die Optimierung der Betriebsführung eingesetzt werden. Aus Ganzjahressimulationen kann eine mittlere Kühlleistung, die Summenhäufigkeitskurven für die Raumtemperatur und der Energiebedarf der Pumpen ermittelt werden. Vor dem Beginn aufwendiger Messungen steht auf Basis der Berechnungen bereits ein optimiertes Kühlsystem zur Verfügung.

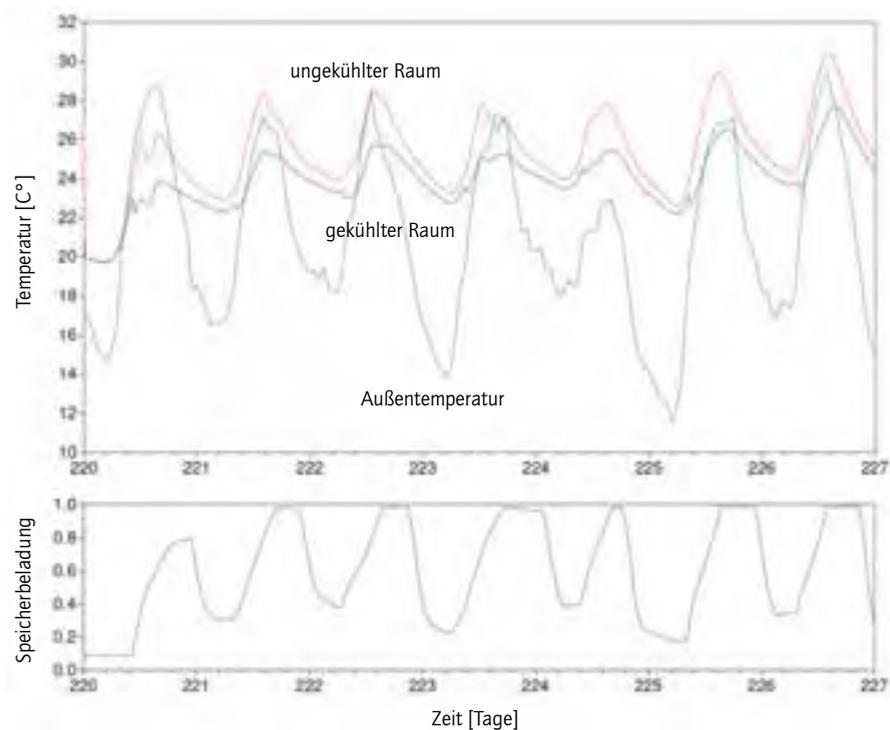
Abb. 4 zeigt die Simulationsergebnisse für den Betrieb eines Kühlungssystems auf Basis des Latentwärmespeichers während einer Hochsommerperiode. Im oberen Teil der Darstellung sind die Raumtemperaturen des gekühlten Raumes, des ungekühlten Referenzraumes und die Temperatur der Umgebungsluft wiedergegeben. Der untere Teil zeigt den Zustand der Speicherbeladung. Deutlich ist hier zu sehen, dass die Kapazität des Latentwärmespeichers nicht an allen Tagen ausreicht.

Die erhaltene Information „Speicherkapazität reicht im Hochsommer nicht aus“ kann auf verschiedene Weise genutzt werden: Soll das System für alle simulierten Fälle ausreichend leistungsfähig dimensioniert werden, können die Simulationsergebnisse zur Optimierung der Speichergröße herangezogen werden. Ist eine weitere Vergrößerung des Speichers hingegen nicht möglich, beispielsweise weil die Größe der Wärmeübertrager an der Fassade baulich begrenzt und somit die Regeneration von größeren Speicherkapazitäten ausgeschlossen ist, können die Erkenntnisse zum optimierten Speichereinsatz genutzt werden. Dazu wäre eine Regelung denkbar, welche die systembedingt begrenzte Ressource Kühlleistung möglichst effizient einsetzt. Dies kann beispielsweise durch Kopplung an ein Raumbenutzungsschema erreicht werden. Speicherkapazität würde in diesem Fall vom Regelungssystem durch Aussetzen der Kühlung solange zurückgehalten, wie ein Raum unbenutzt ist. Dieses Vorgehen würde im Bereich der Büronutzung von Gebäuden bedeuten, dass am Wochenende auch bei hohen Raumtemperaturen nicht gekühlt werden muss und somit zum Wochenanfang die volle Speicherkapazität zur Verfügung steht. Doch auch innerhalb eines Tages kann eine optimierte Regelung

den Einsatz des Latentwärmespeichers verbessern. Durch Einbeziehung von Daten der Wettervorhersage kann erreicht werden, dass an Tagen, an denen die Speicherkapazität voraussichtlich nicht für die Einhaltung der sonst üblichen Temperaturobergrenzen ausreicht, von Beginn des Tages an höhere Temperaturen zugelassen werden. In diesem Fall würde die Raumkühlung später als normal einsetzen, dafür jedoch noch bis zum Dienstschluss der Büronutzer über Speicherkapazität verfügen. Auf diese Weise könnte ein leicht erhöhtes Temperaturniveau gehalten werden, während eine „nicht vorausschauende“ Regelung für den ersten Teil des Tages niedrigere Temperaturen erreichen, dann aber vollständig ausfallen würde. Ein starker Anstieg der Raumtemperatur am Nachmittag wäre die Folge.

Alle genannten Ansätze können auf Basis des mathematischen Modells beantwortet werden. Die Messergebnisse an einem Prototyp des neuen Kühlsystems ohne Kältemaschinen haben gezeigt, dass die Berechnungsergebnisse gut mit den gemessenen

Abb. 4: Berechnungsergebnis für einen Raum mit und einen Raum ohne Kühlsystem mit Latentwärmespeicher (oben), Darstellung der Speicherbeladung (unten)



Daten übereinstimmen. Aufgrund von ungenauen Randbedingungen (Gebäudeumströmung, bauliche Toleranzen, mehrdimensionale Wärmeleitvorgänge) kann nicht erwartet werden, dass die berechneten Temperaturverläufe genau mit den Messdaten übereinstimmen. Differenzen von bis zu einem Kelvin sind üblich und schränken den wesentlichen Vorteil einer Simulation nicht ein: die vergleichende Bewertung unterschiedlicher Systemvarianten.

3 STÄRKEN/SCHWÄCHEN-ANALYSE, HERAUSFORDERUNGEN

Viele wichtige Funktionen der Gebäudehüllen und der Anlagentechnik können mit mathematischen Verfahren abgebildet und bewertet werden. Die Energieeffizienz des Gebäudes wird wesentlich in den ersten Phasen des Planungsprozesses bestimmt und kann durch den gezielten Einsatz von unterschiedlichen Berechnungswerkzeugen verbessert werden.

Obwohl heute bereits viele Werkzeuge für die virtuelle Planung von Gebäuden zur Verfügung stehen, werden die aufwendigeren mathematischen Verfahren in der Praxis nur vereinzelt eingesetzt. Die Zusammenarbeit zwischen den einzelnen Fachplanern der unterschiedlichen Gewerke und dem Architekten ist bei vielen Projekten nicht auf einen virtuellen Entwicklungsprozess ausgerichtet. Häufig gibt es kein Gesamtmodell des Gebäudes, an dem alle Planungsbeteiligten gleichzeitig arbeiten können. Die Planungsaufgaben werden an Teilmodellen durchgeführt, die endgültige Raumgeometrie und die Wandaufbauten werden erst spät im Planungsprozess definiert.

Bei größeren Bauvorhaben werden häufig thermische Gebäudesimulationen für ausgesetzte Räume in den ersten Phasen des Entwurfs eingesetzt. Zusätzlich werden Tageslichtanalysen und in besonderen Fällen Strömungsberechnungen durchgeführt. Leider beschränkt sich die Anwendung dieser Werkzeuge meist auf den frühen Stand eines Entwurfs, alle Änderungen und Modifikationen während des Planungsprozesses werden nicht „online“ verfolgt und bewertet. Dabei können aus guten Berechnungsmodellen bereits viele Erkenntnisse für die oft schwierige Inbetriebnahme der Gebäudetechnik gewonnen werden. Die Berechnungsmodelle eignen sich zudem für eine Optimierung und Überwachung der Gebäudeleittechnik.

Neben einer stetigen Verbesserung der einzelnen Berechnungsverfahren muss insbesondere an den Schnittstellen der einzelnen Gewerke gearbeitet werden. Für den virtuellen Entwicklungsprozess, wie er heute bereits sehr weitgehend in der Automobilindustrie gelebt wird, fehlen im Bauprozess das gewerkübergreifende virtuelle Gebäude und die notwendige Planungskultur.

In Zukunft werden wir deutlich weniger Primärenergie als heute für den Betrieb unserer Gebäude einsetzen können. Für die Reduktion des Primärenergiebedarfs kann sowohl eine Reduktion des Energieflusses durch die Gebäudehülle als auch die Qualität der eingesetzten Energie herangezogen werden. Viele technische Prozesse stellen große Mengen an Abwärme zur Verfügung, die aufgrund des geringen Temperaturniveaus viel-

fach nicht genutzt werden. Durch den Ausbau von Wärmenetzen in Stadtgebieten sowie eine Absenkung der notwendigen Differenztemperaturen in der Anlagentechnik eines Gebäudes können erhebliche Einsparungen für das notwendige Heizen und Kühlen erreicht werden.

Neue Berechnungsverfahren müssen daher nicht nur einen Raum oder ein Gebäude abbilden können. Die Systemgrenze verschiebt sich zu Quartieren, Stadtvierteln und ganzen Versorgungsgebieten. Vorgaben für die Lastprofile und notwendigen Temperaturniveaus müssen daher aus dem Gesamtsystem abgeleitet und in den einzelnen Gebäuden umgesetzt werden.

4 VISIONEN UND HANDLUNGSEMPFEHLUNGEN

Jedes neue Gebäude ist ein Prototyp, und es können keine, wie es bei vielen anderen Entwicklungsprozessen üblich ist, Versuchsmuster gebaut werden. Daher ist die Weiterentwicklung des virtuellen Planungsprozesses essentiell für eine kontinuierliche Verbesserung der Energieeffizienz von Gebäuden. Die Mathematik leistet schon heute große Beiträge zur Entwicklung neuer Komponenten und zur Verbesserung energetischer Konzepte. In Zukunft wird die Bedeutung mathematischer Verfahren weiter steigen und immer mehr Aspekte der Planung können in einen virtuellen Prozess einfließen. Das führt zu sehr komplexen Berechnungsmodellen, die sowohl eine geeignete Datenhaltung und -auswertung als auch geeignete Verfahren zur Abschätzung numerischer Ungenauigkeiten erfordern.

Einer der wichtigsten Punkte für die Weiterentwicklung der Planungswerkzeuge ist sicher die Schnittstelle zwischen den mathematischen Verfahren und dem Menschen. Die kreative Zusammenarbeit zwischen dem Architekten und den Fachplanern darf nicht durch die Vorgaben einer virtuellen Entwicklungsumgebung gestört werden. Vielmehr sollte durch geeignete Visualisierungs- und Bedienkonzepte das gemeinsame Arbeiten an einem Gebäude unterstützt werden. Auch der Bauherr und der spätere Nutzer können durch eine Vereinfachung der Interaktion zwischen Mensch und Gebäudemodell aktiver Teil des Planungsprozesses werden, wenn über dreidimensionale Visualisierungsverfahren erlebbare und vielleicht sogar fühlbare Umgebungen entstehen.

5 LITERATUR

- [1] Umweltbundesamt (Hg.): Nachhaltige Wärmeversorgung – Sachstandsbericht. ISSN 1862-4359. Dessau, 2007.
- [2] Verbundforschungsprojekt „Niedrigenergiesysteme für die Heiz- und Raumlufttechnik“, Finanzierung durch das Bundesministerium für Wirtschaft und Technologie (BMWi), Betreuung durch den Projektträger Jülich (PTJ), Förderkennzeichen 0327370A.

- [3] Lamberg, P.: Approximate analytical model for two-phase solidification problem in a finned phase change material storage. In: Applied Energy 77 (2004), S. 131–152.
- [4] Glück, B.: Einheitliches Näherungsverfahren zur Simulation von Latentwärmespeichern. In: HLH, Lüftung, Klima, Heizung, Sanitär, Gebäudetechnik, Springer Verlag 57 (2006), S. 25–30.
- [5] Tschirner, T.; Haase, T.; Hoh, A.; Matthes, P.; Müller, D.: Simulation of Phase Change Procedure in Heat Storage Devices Using the Object Oriented Modelling Language Modelica. In: Association française du froid (Hrsg.): Proceedings of the 7th Conference on Phase Change Materials and Slurries for Refrigeration and Air Conditioning 2006, Dinan, Frankreich, 2006.
- [6] Günther, E., Mehling, H., Hiebler, S.: Modeling of subcooling and solidification of phase change materials. In: Modelling and Simulation in Material Science and Engineering 15 (2007), S. 879–892.

MEDIZIN

> MEHR MATHEMATIK WAGEN IN DER MEDIZIN

PETER DEUFLHARD, OLAF DÖSSEL, ALFRED K. LOUIS UND STEFAN ZACHOW

1 EXECUTIVE SUMMARY

In diesem Artikel wird an drei Erfolgsmodellen dargestellt, wie das Zusammenwirken von Mathematik und Medizin eine Entwicklung hin zu patientenspezifischen Modellen auf Basis moderner medizinischer Bildgebung angestoßen hat, die in naher Zukunft dynamisch weiter Raum greifen wird. Dabei existiert ein Gleichklang der Interessen von Medizin und Mathematik: Beide Disziplinen wollen die Resultate schnell und zuverlässig. Für die Klinik heißt dies, dass notwendige Rechnungen in möglichst kurzer Zeit, und zwar auf dem PC, ablaufen müssen und dass die Resultate so genau und belastbar sein müssen, dass medizinische Entscheidungen darauf aufbauen können. Für die Mathematik folgt daraus, dass höchste Anforderungen an die Effizienz der verwendeten Algorithmen und die darauf aufbauende Software in Numerik und Visualisierung zu stellen sind. Allerdings ist es noch ein weiter Weg, bis anatomische und medizinisch brauchbare funktionelle Modelle auch nur für die wichtigsten Körperteile und die häufigsten Krankheitsfälle verfügbar sein werden. Führende Universitätskliniken könnten, als Zentren einer interdisziplinären Kooperation von Medizinerinnen, Ingenieuren und Mathematikerinnen, eine Vorreiterrolle dabei übernehmen, mehr Mathematik in der Medizin zu wagen. Dies wäre zweifellos ein wichtiger Schritt in Richtung auf eine *individuelle quantitative Medizin*, bei dem Deutschland die besten Voraussetzungen hätte, die Rolle des „Schrittmachers“ zu übernehmen.

2 MATHEMATIK IN DER MEDIZINISCHEN BILDGEBUNG

2.1 GESCHICHTE EINES ERFOLGES

Der Physiker Röntgen hat mit der Entdeckung seiner „X-Strahlen“ (im Englischen heute noch: *X-rays*) ein Fenster geöffnet, das nicht-invasiv Einblicke in das Körperinnere ermöglicht. Röntgenstrahlen werden beim Durchgang durch unterschiedlich dichte Gewebe, wie Knochen oder Fett, unterschiedlich stark gedämpft. Dies ermöglichte Schattenbilder, die visuelle Information über das Körperinnere liefern. Da jedoch in dieser Art der Bildgebung Knochen die darunter liegenden Gewebe verdecken, waren aussagekräftige Schattenbilder etwa des Gehirns nicht oder nur durch die nicht ungefährliche Zugabe von Kontrastmitteln möglich.

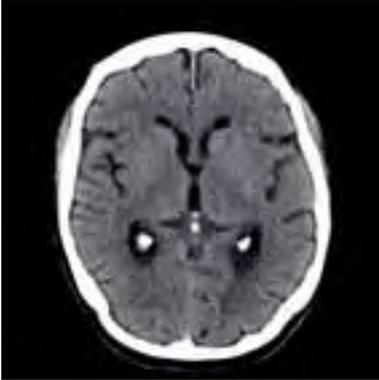
Der konzeptionelle und technische Durchbruch gelang in den 60er Jahren: Die Nobelpreisträger Cormack und Hounsfield schlugen ein Messverfahren vor, bei dem Schattenbilder eines imaginären Schnittes durch den Körper für sehr viele unterschiedliche Richtungen aufgezeichnet wurden. Jedes einzelne dieser Bilder liefert wenig Information. Erst die Ausnutzung der Tatsache, dass alle Schattenbilder den gleichen Bereich des Körpers darstellen, führte zu Bildern des Körperinneren von bis dahin ungekannter räumlicher Auflösung. Dies war der Beginn einer langen Erfolgsgeschichte.

Computertomographie (CT). Die im Deutschen als Computer-Tomographie (kurz: CT) bezeichnete Messtechnik wurde im Englischen ursprünglich genauer als „Computed tomography“ bezeichnet, was die Rolle der Mathematik deutlicher macht. Natürlich geschieht die Berechnung der gesuchten Bildinformation wegen der riesigen Datenmengen mit einem Computer. Die mathematische Beschreibung des komplexen Zusammenhanges zwischen der gemessenen Information, der Dämpfung der Intensität der Röntgenstrahlen – auf etwa einer Million unterschiedlichen Wegen durch den Patienten – und der gesuchten Information über die Dichte¹ führt auf eine Integralgleichung, die im 2D-Fall als Radon-Transformation bezeichnet wird. Darin ergeben sich die Daten als Linienintegrale über die gesuchte Dichte längs der Wege der Röntgenstrahlen. Nach Diskretisierung der Integralgleichung entsteht ein großes lineares Gleichungssystem mit Spezialstruktur. Zu Beginn der Entwicklung hatte der Ingenieur Hounsfield dieses Gleichungssystem iterativ gelöst. Erst Mitte der 70er Jahre erkannte man, dass schon 1917 der österreichische Mathematiker Radon eine Inversionsformel zur Berechnung einer Funktion aus ihren Linienintegralen publiziert hatte.

Von der mathematischen Formel bis zu effizienten Algorithmen war allerdings noch ein weiter Weg, es mussten Fragen nach der Auflösung, der Nichteindeutigkeit bei endlich vielen Daten und der erforderlichen Schrittweiten geklärt werden. Hinzu kommt ein weiteres Phänomen, das typisch ist für alle bildgebenden Verfahren: Es handelt sich um ein sogenanntes *inverses* Problem, bei dem die unvermeidbaren Datenfehler, verursacht etwa durch Photonenrauschen oder Ungenauigkeiten im Detektor, in der Lösung extrem verstärkt werden. Um dies zu vermeiden, muss das Problem „regularisiert“ werden, und zwar derart, dass eine akzeptable Balance zwischen möglichst guter Auflösung im Bild und noch akzeptablem Rauschen durch den Datenfehler erhalten wird. Im Falle der CT erreicht man dies durch Elimination der hochfrequenten Anteile in den Daten. Hier hat sich eine von den Mathematikern Shepp und Logan [17] entwickelte Methode etabliert. Die daraus resultierenden Algorithmen bestehen aus zwei Schritten, einer Filterung der Daten (mittels Fouriermethoden) und einer Rückprojektion der gefilterten Daten auf das Rekonstruktionsgebiet. Beides kann parallel zur Messung durchgeführt werden, was einen dramatischen Gewinn an Rechengeschwindigkeit liefert. Nebenbei sei bemerkt,

¹ Genauer: über den als Gewebedichte interpretierten Röntgenschwächungskoeffizienten.

Abb. 1: Rekonstruktion eines transaxialen Schnittes durch den Kopf (Quelle: Siemens Healthcare, Sektor Computertomographie, SOMATOM Definition AS+)



dass hierbei die Beschleunigung der Verfahren durch die Verbesserung der Algorithmen um ein Vielfaches höher war als der Gewinn durch die Verbesserung der Computer.

Vergessen sollte man sicher nicht die Leistung der Ingenieure, welche ein immer schnelleres Messen der benötigten Daten ermöglichten. Diese Erhöhung der Messgeschwindigkeit führte zu neuen Messanordnungen und als Folge zu neuen Anforderungen an die Mathematik. Von der ursprünglich von Hounsfield eingesetzten parallelen Geometrie, wo immer jeweils ein Strahlenweg gemessen wurde, führte die Entwicklung zur Fächergeometrie, bei der von einer Quellposition aus das gesamte Untersuchungsgebiet mit Strahlen überdeckt wurde. Die Spiralgeometrie, zunächst mit wenigen Detektorzeilen, etablierte sich zu Beginn der 90er Jahre. Heute ist die dreidimensionale Anwendung mit flächigen Detektoren Gegenstand intensiver Forschung; beim zerstörungsfreien Prüfen wird sie schon länger eingesetzt. Die so berechneten Bilder werden anschließend meist weiterverarbeitet, um die Diagnose zu erleichtern. Neu sind Verfahren, bei denen diese Bildverarbeitung in den Rekonstruktionsprozess integriert wird. Der Mathematiker Natterer hat in Deutschland als Erster das Thema Tomographie als Forschungsgebiet aufgegriffen und bahnbrechende Resultate erzielt. Eine umfassende Darstellung findet sich in seinem jüngsten Buch [15].

Magnetresonanztomographie (MRT). Magnetische Resonanzspektroskopie, 1946 unabhängig von Bloch und Purcell entwickelt, liefert Informationen über die chemische Umgebung von Protonen in einem Molekül, indem man es in einem starken Magnetfeld durch geeignete Anregung in Resonanz versetzt. Dies bildet die technische Basis für

Abb. 2: Rekonstruktion und Visualisierung basierend auf MRT-Daten (Quelle: Siemens Healthcare, Sektor Magnetresonanztomographie, MAGNETOM Verio 3T)



die Magnetresonanztomographie (früher: Kernspintomographie). Lauterbur, der 2003 den Nobelpreis für Medizin erhielt, erzielte eine räumliche Auflösung dadurch, dass er das primäre, homogene Magnetfeld durch Gradientenfelder so veränderte, dass die Gebiete konstanter Resonanzfrequenz Ebenen waren. So gelang es, Ebenenintegrale über die Protonenverteilung im Körper zu messen. Das dazu passende mathematische Modell war die Radon-Transformation in 3D [14]. Zu Beginn der 80er Jahre standen keine Arbeitsplatzrechner zur Verfügung, welche die hierbei anfallenden riesigen Datenmengen für dreidimensionale Aufnahmen in akzeptabler Zeit hätten verarbeiten können. Deshalb hat Mansfield, ebenfalls 2003 mit dem Nobelpreis für Medizin ausgezeichnet, die Gradientenfelder und die Anregungen so weiterentwickelt, dass die zweidimensionale Fourier-Transformation zur Inversion eingesetzt werden konnte. Die bei der Röntgen-Tomographie angesprochene Regularisierung wird somit direkt durch die technische Elimination der hohen Frequenzen erreicht. Auf diese Weise hat man also, mit hohem technischen Aufwand, das mathematische Problem so vereinfacht, dass es mit den Computern der damaligen Zeit lösbar wurde.

Ultraschall. Diese Messtechnik wird seit langem in der Medizin eingesetzt. Etabliert sind Geräte, sogenannte B-Scan-Geräte, die sowohl aus Sender als auch aus Empfänger bestehen. Gemessen wird dabei die unterschiedliche Laufzeit von Ultraschallwellen, die an Geweben unterschiedlicher Schallimpedanz und Streueigenschaften reflektiert werden. Daraus kann ein Bild berechnet werden, wenn man annimmt, dass die Schallgeschwindigkeit vom Medium unabhängig ist. Dies ist zwar nicht korrekt; trotzdem entstehen auf diese Weise Bilder mit akzeptabler diagnostischer Information. Berücksichtigt man

zusätzlich Frequenzverschiebungen, so lässt sich über den Doppler-Effekt die Blutgeschwindigkeit darstellen.

2.2 MATHEMATIK ALS INNOVATIONSFAKTOR

Die medizintechnische Industrie in Deutschland behauptet seit vielen Jahren im Bereich der Bildgebung eine international herausragende Führungsposition. In 2002 wurden in Deutschland für ca. 2 Mrd. Euro Waren im Bereich Röntgen und CT produziert. Der Außenhandelsüberschuss Deutschlands bei Röntgen und CT lag bei ca. 1 Mrd. Euro, bei Magnetresonanztomographie Systemen (MRT) betrug er 600 Mio. Euro². Hierfür ist im Wesentlichen das Unternehmen Siemens Medical Solutions verantwortlich. Auch Philips Healthcare hat seine zentralen Forschungszentren für CT und MRT in Deutschland. Die Firma GE hat gerade ein Forschungszentrum in Garching eröffnet, in dem sich eine große Abteilung der medizinischen Bildgebung widmet. Ein wichtiger Standortfaktor sind offenbar gut ausgebildete angewandte Mathematiker und Ingenieure mit soliden Kenntnissen in Mathematik.

2.3 PERSPEKTIVE: NEUE BILDGEBUNGSMETHODEN

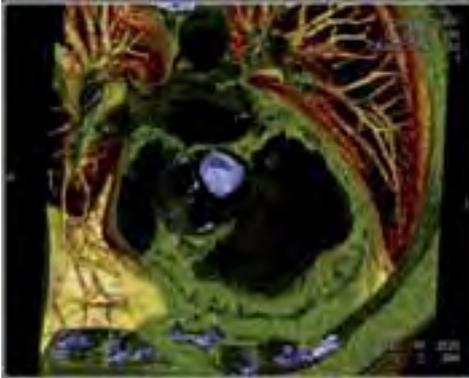
Die Perspektive im Gebiet medizinische Bildgebung ist geprägt von einem rasanten Fortschritt immer weiterer Bildgebungsmethoden.

3D-Tomographie. In der Materialprüfung wird diese Technik seit längerem eingesetzt, in der Regel in Verbindung mit kreisförmiger Abtastgeometrie, also Führung der Röntgenquelle auf einer Kreisbahn um das Objekt und flächigem Detektor. In der Medizin allerdings sind die bei der Spiralabtastung notwendigen variablen Vorschübe, bei denen der Patient während der Messung durch den Scanner geschoben wird, algorithmisch noch nicht zufriedenstellend gelöst. Prinzipiell können Röntgenröhre und Detektor auf beliebigen Trajektorien um den Körper geführt werden. Die Bestimmung optimaler Trajektorien (bezüglich Auflösung und Stabilität) bleibt weiterhin eine Herausforderung an die Mathematik. Die höhere Rechnerleistung lässt auch an den Einsatz anderer Verfahren als diejenigen der gefilterten Rückprojektion denken.

4D-Tomographie. Schon seit langem werden aus Serien von planaren Bildern (CT, MRT etc.) *räumliche Darstellungen* berechnet. Fügt man Serien von Daten unterschiedlicher Messzeitpunkte zusammen, so kommt als vierte Dimension die *Zeit* hinzu. Die so produzierten Filme, zum Beispiel vom schlagenden Herzen, sind beeindruckend: Abb. 3 ist ein Standbild aus einem Film vom schlagenden Herzen. Allerdings sind bis zum standardmäßigen Einsatz solcher Techniken noch viele Verbesserungen nötig. Derzeit erfolgt

² Außenhandelsstatistik der OECD und BMBF-Studie

Abb. 3: Dreidimensionale Visualisierung einer Serie von zweidimensionalen CT-Aufnahmen des Herzens (Quelle: Siemens Healthcare, Sektor, Computertomographie, SOMATOM Definition AS+)



eine Regularisierung nur in den räumlichen Koordinaten, eine zusätzliche Glättung in der Zeit, wie sie zum Beispiel bei der Stromstärkerekonstruktion aus EEG/MEG-Daten unverzichtbar ist, steht noch aus.

Elektronenspin-Resonanz-Tomographie. Bei dieser ebenfalls vierdimensionalen Tomographie ist keine Entkopplung der Dimensionen möglich. Hier tritt nämlich neben den drei Raumdimensionen eine *spektrale* Dimension hinzu. Das zugehörige mathematische Modell ist nun die Radon-Transformation in vier Dimensionen. Dieses Verfahren wird zur Zeit in der Pharmaforschung und bei Tierversuchen eingesetzt. Allerdings können hierbei, wegen der Einschränkung an die verwendbaren Feldstärken, die Daten im Radon-Raum nicht vollständig abgetastet werden, sodass sich das Problem unvollständiger Daten stellt. Obwohl im Idealfall eine eindeutige Rekonstruktion (mathematisch beweisbar) möglich wäre, erschweren Instabilitäten und starke Artefakte das Rekonstruktionsproblem.

Ultraschall-CT. Bei räumlich getrennter Anordnung von Sender und Empfänger ist das beschreibende mathematische Modell ein inverses Streuproblem für die Bestimmung der räumlich variierenden Schallimpedanz und Streueigenschaften. Die Schwierigkeit besteht darin, dass hier, im Gegensatz zum CT, die Wege der Wellen wegen der Streuung von der gesuchten Größe selbst abhängen und somit das Problem hochgradig nichtlinear ist. Eine Linearisierung des Problems mittels Born- oder Rytov-Approximation berücksichtigt nicht die Effekte der Mehrfachstreuung und ist somit nicht hinreichend genau.

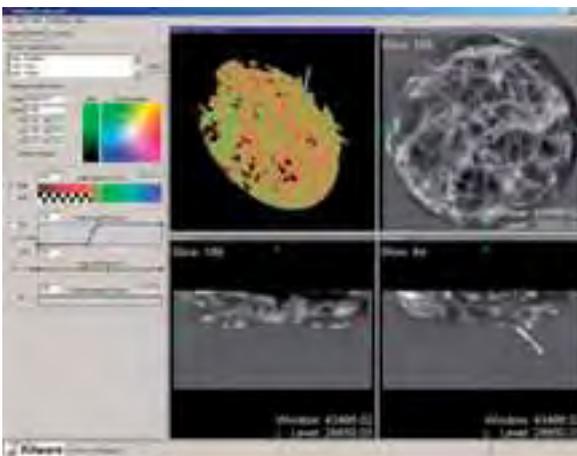
Deshalb stellt die Ultraschall-Tomographie auch heute noch höchste Anforderungen an Mathematik und Algorithmenentwicklung.

Transmissions-Elektronen-Mikroskopie. Zur Visualisierung von Biomolekülen wird eine Reihe recht unterschiedlicher Ansätze verfolgt. Will man nicht über viele gleichartige Proben mitteln, so entsteht dort ebenfalls ein Rekonstruktionsproblem mit unvollständigen Daten. Hinzu kommen bei den kleinen Dimensionen Wellenphänomene der Strahlen, sodass man, ähnlich wie bei der Ultraschall-Tomographie, nichtlineare inverse Streuprobleme zu lösen hat. Glücklicherweise ist hier eine Linearisierung möglich, was die Algorithmenentwicklung wesentlich erleichtert.

Phasenkontrasttomographie. Entsprechendes gilt auch bei dieser neuartigen Technik, bei der komplexwertige Größen rekonstruiert werden. Die Phase liefert auch dann noch Informationen, wenn die Dichteunterschiede im untersuchten Objekt sehr gering sind. Bei Daten von einem Synchrotron hat man gute Resultate erzielt, siehe Abb. 4. Der Einsatz in der Medizintechnik stellt wegen der unterschiedlichen Messgeometrien erhebliche Anforderungen an die Entwicklung der Rekonstruktionsalgorithmen.

Diffusions-Tensor-MRT. Diese Methode liefert als Resultate in jedem Punkt des Rekonstruktionsgebietes einen Tensor, der Auskunft über das diffusive Verhalten von Wassermolekülen im Gewebe gibt, siehe Abb. 2 auf S. 438. Hier sind Rekonstruktion und

Abb. 4: Phasenkontrasttomographie: Phaseninformation (siehe [9])



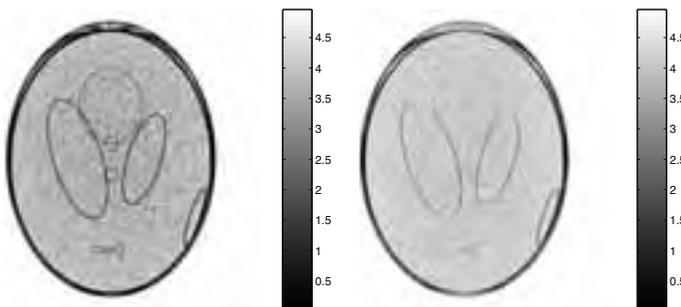
Regularisierung noch getrennt. Erst nach der Berechnung werden Eigenschaften der Tensoren, wie Symmetrie oder positive Definitheit, punktweise erzeugt.

Die Liste der in Entwicklung befindlichen bildgebenden Verfahren ist beileibe nicht vollständig. Verfahren der *Impedanz-Tomographie* werden ebenso untersucht wie die Verwendung von *Licht* für oberflächennahe Untersuchungen. Bei allen genannten Messtechniken ist die technologische Entwicklung so weit fortgeschritten, dass die Lösung der mathematischen Probleme, also Modellbildung, Klärung der erreichbaren Auflösung und Entwicklung effizienter Algorithmen, einen erheblichen Innovationsschub liefern wird.

Bildfusion. Unterschiedliche bildgebende Verfahren liefern unterschiedliche Informationen, deren Zusammenführung dem Arzt mehr Erkenntnisse liefern kann als die einzelnen Bilder. Diese Zusammenführung setzt voraus, dass Aufnahmen (fast) gleicher Schnitte oder dreidimensionale Volumendarstellungen vorliegen. Es müssen die unterschiedlichen Formate aufeinander abgebildet werden. Dies setzt Bilderkennung, etwa durch Segmentierung, voraus, um die geometrischen Transformationen bestimmen zu können. Erste Verfahren existieren schon, sie werden bei der sich stürmisch entwickelnden IMRT, der *intensitätsmodulierten Radio-Therapie*, zum Teil eingesetzt.

Feature-Rekonstruktion. Die Unterstützung des Arztes bei der Interpretation der Rekonstruktionen ist bei der vorhandenen Informationsflut sehr wichtig. Beim zerstörungsfreien Prüfen in der Materialwissenschaft, wo ebenfalls tomographische Methoden eingesetzt werden, versucht man schon seit einiger Zeit eine *automatische* Erkennung klar definierter Eigenschaften, etwa beim Auffinden von Lunkern im Innern des Materials. In der medizinischen Anwendung ist diese Herangehensweise sicher nicht das Ziel; aller-

Abb. 5: Vergleich der Bildverarbeitung. Links: in Rekonstruktion integriert. Rechts: wie bisher üblich getrennt (siehe [13]).



dings ist eine Vorverarbeitung von Bildern oft hilfreich. Zur Zeit werden solche Prozeduren erst im Anschluss an die Rekonstruktion durchgeführt. In [13] wurde ein neuartiges Verfahren vorgestellt, bei dem die Bildverarbeitung schon in den Rekonstruktionsprozess integriert wird. Dadurch ergeben sich deutlich bessere Resultate, wie Abb. 5 zeigt. Fast ohne zusätzlichen Aufwand lassen sich die ursprünglichen und die weiterverarbeiteten Rekonstruktionen gleichzeitig berechnen. Darüber hinaus bietet dieser Weg die Möglichkeit, Bildrekonstruktion und Bildverarbeitung zu vereinen.

3 MATHEMATIK IN DER KARDIOLOGIE UND DER HERZCHIRURGIE

3.1 ERFOLGSGESCHICHTEN: EKG UND BIOSIGNALVERARBEITUNG

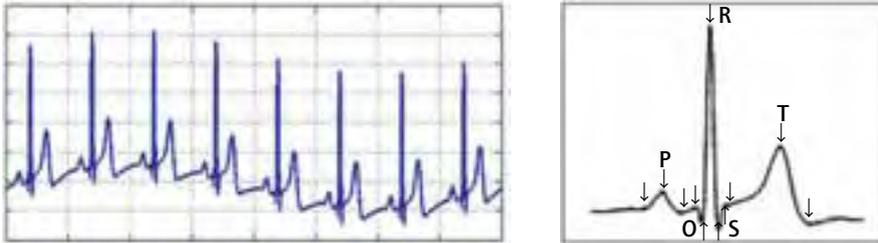
Täglich werden auf der Welt Millionen von Elektrokardiogrammen (EKGs) aufgezeichnet, von denen viele Langzeit-EKGs sind. Ein einziges 24h-EKG enthält schon ungefähr 100 000 Herzschläge. Es ist praktisch nicht möglich, dass Ärzte alle diese Daten nach auffälligen Herzschlägen durchsuchen. Es würde nicht nur kostbare Zeit verlorengehen, sondern es wäre auch zu befürchten, dass die Aufmerksamkeit des Betrachters ziemlich schnell nachlässt und wichtige Ereignisse übersehen werden.

Telemedizin ist eine wichtige Zukunftsbranche – mehrere Unternehmen positionieren sich gerade im Markt. Sportler und ältere Menschen können eine kontinuierliche Überwachung ihrer Gesundheit – 24 Stunden am Tag und 7 Tage die Woche – als Dienstleistung "buchen". Neben dem EKG werden oft noch andere vitale Gesundheitsparameter gemessen, wie Gewicht, Temperatur, Blutsauerstoff-Sättigung etc. Auch hier ist es nicht praktikabel und nicht bezahlbar, dass Ärzte kontinuierlich alle diese Messdaten überwachen.

Abhilfe in dieser Situation bieten Computerprogramme, die den Arzt bei der EKG-Analyse unterstützen. Mit ihrer Hilfe sollen automatisch die wesentlichen Kenngrößen des EKGs gefunden und der Arzt auf wichtige Signalabschnitte hingewiesen werden. Beim Monitoring auf der Intensivstation oder in der Telemedizin soll ein für den Patienten bedrohlicher Zustand schnell und zuverlässig erkannt und ein Alarm ausgelöst werden. Aber welche Algorithmen sind dazu geeignet, diese Analysen mit der bestmöglichen Trefferquote durchzuführen? Die Algorithmen müssen auf alle (!) EKGs anwendbar sein, auch wenn das EKG eines Patienten manchmal sehr ungewöhnlich ist. Und sie müssen sehr robust sein, da oft viele Störsignale die eigentlichen Biosignale überwiegen.

Mathematische Methoden sowie die zugehörige Software sind heute unverzichtbare Bestandteile einer guten Biosignalanalyse. Die Unternehmen der Medizintechnik haben erkannt, dass heute die Wertschöpfung bei EKG- und Monitoring-Systemen nur noch zur Hälfte in der Hardware (z. B. in der Elektronik) entsteht; die andere Hälfte zum Erfolg eines Produktes trägt intelligente Signalauswertung bei. Zur Unterdrückung von Störsignalen wird neben der Filterung im Frequenzbereich zunehmend die Wavelet-Transformation eingesetzt, da sie die bestmögliche Frequenzauflösung bei gleichzeitiger

Abb. 6: EKG-Originaldaten und das Ergebnis einer computerunterstützten Annotation der wichtigsten Zeitpunkte [11]



guter Lokalisierung im Zeitbereich erlaubt. Für die Detektion wichtiger Signalabschnitte im EKG werden mit mathematischen Methoden Merkmale aus dem EKG extrahiert und einem Klassifikator (Nearest Neighbour, Bayes-Maximum Likelihood, Neural Networks usw.) zugewiesen, der dann, beispielsweise, einen QRS-Komplex zuverlässig von einer Extrasystole unterscheiden kann [18]. Auch modellbasierte Techniken wie Kalman-Filter oder Markov-Modelle etc. werden eingesetzt. Abb. 6 zeigt ein typisches EKG, wie es aus einer Langzeitaufzeichnung kommt, sowie die Annotation der wichtigsten Zeitpunkte, wie z. B. R-Peak, P-Welle, T-Welle.

Um frühzeitig Trends im EKG zu erkennen, die auf eine Krankheit hinweisen, werden neue Methoden der *Zeitreihenanalyse* eingesetzt. Mit dieser Methodik sollen etwa stochastische Schwankungen von deterministischen unterschieden oder signifikante Änderungen erkannt werden, lange bevor sie das Auge des Betrachters im EKG erkennen kann. Man sucht heute nach mathematischen Regeln (linear/nichtlinear), mit denen das EKG des nachfolgenden Herzschlags aus dem gegenwärtigen vorhergesagt werden kann, um so aus einer möglichen Abweichung („Regelverstoß“) frühzeitig eine Diagnose ableiten zu können. Damit ist man mitten in der Theorie dynamischer Systeme, mit der neuartige und diagnostisch wichtige Aspekte im EKG erkannt werden können. So beobachtet man meistens stabile Trajektorien im Phasenraum (Attraktoren, normaler Herzrhythmus), die aber manchmal über Bifurkationen ins Chaos übergehen (Herzflimmern [5]).

3.2 MATHEMATIK ALS INNOVATIONSFAKTOR

In der Medizintechnik im Allgemeinen und im Bereich der Kardiologie und Herzchirurgie im Besonderen kommt es immer stärker darauf an, aus sehr vielen unterschiedlichen Informationen komplexe Zusammenhänge zu erkennen und die richtigen Schlüsse daraus zu ziehen. Hierbei ist keinesfalls daran gedacht, den Arzt langfristig durch einen

Computer zu ersetzen. Aber der Computer mit seinen implementierten Algorithmen ist schon heute oft ein wertvoller Ratgeber für den Arzt, und diese Entwicklung wird sich fortsetzen und verstärken.

Zunächst geht es nur darum, Informationen aus den Messdaten hervorzuheben, die mit bloßem Auge nicht gut zu erkennen sind: kleine Veränderungen des EKGs, ungewöhnliche Strukturen in den Bilddaten, charakteristische Abweichungen in den Blutwerten etc. Weiter geht es auch um das quantitative Bestimmen von diagnostisch wichtigen Größen. Oft hängt die Entscheidung für den einen oder den anderen Behandlungspfad davon ab, ob eine Größe kleiner oder größer als ein Schwellenwert ist. Mit mathematischen Methoden können diese Größen auch in gestörten und verrauschten Daten so genau wie möglich gefunden werden. Schließlich können mathematische Methoden eingesetzt werden, um die Therapie des Arztes zu optimieren, indem verschiedene Varianten durchgespielt und nach objektiven und nachvollziehbaren Kriterien bewertet werden.

In Zukunft wird es für einen Hersteller von Medizintechnik entscheidend darauf ankommen, die vielen Daten über den Patienten auf intelligente Weise zu einem Gesamtbild zu integrieren und den Arzt bei seiner Diagnose und Therapieentscheidung zu unterstützen. Nur Unternehmen, die diese Optionen anbieten können, werden im Weltmarkt der Zukunft eine führende Rolle übernehmen oder halten können.

3.3 PERSPEKTIVE: DAS VIRTUELLE HERZ

Integration von Bildgebung, Elektrophysiologie und Laborwerten. Das oben beschriebene EKG ist eine wichtige diagnostische Information für den Arzt, aber bei weitem nicht die einzige: Bilddaten und Laborwerte, wie z. B. Blutwerte, sind unverzichtbare weitere Informationsquellen. Mathematische Methoden spielen eine herausragende Rolle, wenn es darum geht, diese vielen Informationen zu einem Gesamtbild über den Zustand des Patienten zu integrieren. Zu den rein geometrischen Bilddaten, wie sie mit Projektions-Röntgen, CT, Ultraschall oder MRT gewonnen werden, kommen in einem ersten Schritt funktionelle Daten, welche von den bildgebenden Systemen selber akquiriert werden können: Metabolismus (z. B. über PET³), Fluss (z. B. durch Doppler-Ultraschall) oder Perfusion (z. B. durch Röntgen mit Kontrastmitteln). Im nächsten Schritt geht es darum, auch andere Signale zu integrieren, wie z. B. EKG, Blutdruck, Blut-Sauerstoffsättigung, weiterhin Blutwerte und Enzymaktivitäten, und langfristig auch genetische Profile, die eine Prädisposition für eine Erkrankung anzeigen. Ziel ist, alle diese Messdaten in ein Patientenmodell zu integrieren, sodass am Ende wichtige funktionelle Eigenschaften bestimmt werden (für physiologisch Vorgebildete: Elastizität, Kontraktilität, elektrische Depolarisierung, Perfusion, Enzymaktivität etc.).

³ Positron-Elektron-Tomographie

An einem so konstruierten *virtuellen Herzen* kann sichtbar gemacht werden, wie das reale Herz des Patienten von der physiologischen Funktionsnorm abweicht. Dies eröffnet ganz neue Möglichkeiten sowohl der Diagnostik als auch der Therapieplanung.

Therapieplanung. Mathematische Herz- und Kreislauf-Modelle haben vielfältige Anwendungen in Kardiologie und Herzchirurgie.

Mit *elektrophysiologischen* Modellen kann man in Zukunft

- RF-Ablationen zur Therapie von Vorhofflimmern optimieren,
- Pulssequenzen für Herzschrittmacher und die kardiale Resynchronisationstherapie anpassen,
- neue Pharmaka zur Behandlung von Rhythmusstörungen in ihrer zu erwartenden Wirkung abschätzen.

Mit *elastomechanischen* Modellen kann man

- Schnittführungen am Herzen bei Aneurysmen oder bei Ventrikelrekonstruktionen optimieren,
- die elastomechanischen Konsequenzen eines Infarktes besser abschätzen.

Mit *fluidodynamischen* Gefäßmodellen kann man

- die zellulären Ursachen für Gefäßverschlüsse besser verstehen,
- Stents zur Öffnung von Stenosen und zur Behandlung von Aneurysmen verbessern.

Mit *Kreislaufmodellen* kann man

- neue Medikamente zur Behandlung von Bluthochdruck bewerten,
- die Therapie vom Kreislauf-Schock verbessern,
- Herz-Lungen-Maschinen für Operationen am offenen Herzen besser steuern.

Es sei im Folgenden eine der Optionen zur genaueren Beschreibung ausgewählt.

Planung von RF-Ablationen bei Vorhofflimmern. Unter diesem Begriff versteht man die punktgenaue Zerstörung von Zellen (meist: Tumorzellen) durch einen Radio-Frequenz-Strom. Mittels eines CT-geführten Katheterdrahtes wird Gewebe lokal auf über 42°C erhitzt, wobei das Eiweiß in unmittelbarer Umgebung der Drahtspitze koaguliert (und nur dort). Dabei entsteht Narbengewebe.

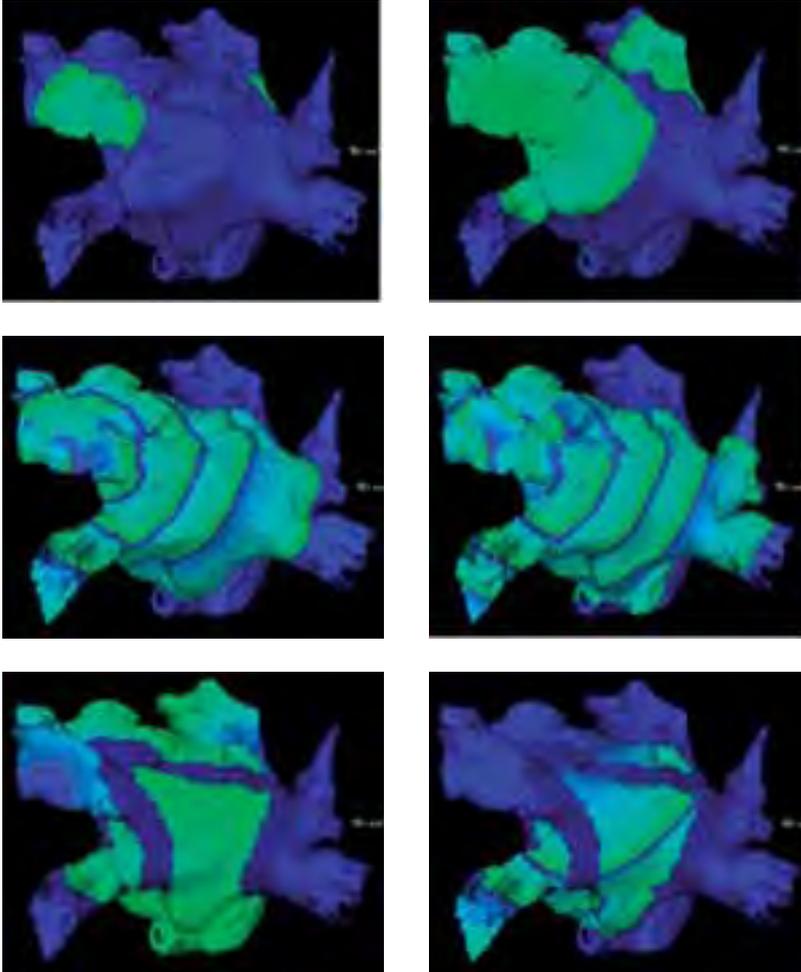
Elektrophysiologische Computermodelle des Herzens beginnen heute bei den einzelnen Ionenkanälen in den Myokardzellen. Deren dynamisches Verhalten wird durch

wenige gewöhnliche, aber „steife“ Differentialgleichungen beschrieben, deren Ratenkonstanten von der Transmembranspannung abhängen [6]. Die räumliche Verknüpfung der Zellen untereinander wird durch partielle Differentialgleichungen beschrieben, z. B. durch das sogenannte „Bidomain-Modell“ [7]. Da am Ende die elektrischen Potentiale im Körper bestimmt werden sollen, sind im Wesentlichen die Gleichungen der elektromagnetischen Feldtheorie zu lösen, und da es sich im Körper um relativ langsame Prozesse handelt, führt das zu einer Poisson-Gleichung, also zu einer elliptischen partiellen Differentialgleichung. Die numerische Simulation des Modells erfordert demnach sowohl eine Diskretisierung der Zeitvariable als auch eine der Raumvariablen. Zur Zeitdiskretisierung wird im Ingenieursbereich bisher die explizite Euler-Diskretisierung bevorzugt [6], während in der mathematischen Community eher steife Integratoren eingesetzt werden (siehe [1] und Zitate dort). Hier ist noch Spielraum für eine wechselseitige Angleichung der Methodik. Die räumliche Diskretisierung erfolgt z. B. mit Finite-Elemente-Methoden, uniform oder adaptiv, oder Finite-Differenzen-Methoden. Bei uniformer Diskretisierung landet man schnell bei einigen Millionen Freiheitsgraden, die möglichst im Sekunden-takt (!) gelöst werden sollten. Auch die sogenannte „Vorwärtsrechnung“, mit der aus vorab bestimmten elektrischen Quellmustern auf dem Herzen die elektrischen Signale an der Körperoberfläche und damit das EKG berechnet werden, erfordert die Lösung einer Poisson-Gleichung. Abb. 7 (obere Reihe) zeigt eine Simulation der Depolarisierung des gesunden menschlichen Vorhofs: Die elektrische Erregung geht vom Sinusknoten, dem „Schrittmacher“ des Herzens, aus und breitet sich dann über den rechten und den linken Vorhof aus.

Der nun folgende Schritt ist die mathematische Modellierung einer Pathologie wie z. B. das Vorhofflimmern. Hierzu müssen dem Vorhofgewebe Eigenschaften zugeordnet werden, die pathologisch sind. Tatsächlich kippt das Computermodell dann in unregelmäßige Muster der elektrischen Erregungsausbreitung um, die den beobachteten Daten bei Vorhofflimmern verblüffend ähnlich sind. Eine heute immer häufiger angewendete Therapie besteht nun darin, mit einer Hochfrequenzsonde kleine Flecken des Vorhofgewebes so weit zu erhitzen, dass sich eine Narbe ausbildet, welche die elektrische Erregung nicht fortleitet (RF-Ablation, Abb. 7, untere Reihe). Eine wichtige offene Frage ist die optimale Strategie bei der RF-Ablation: Wie sollten Bahnen aus Ablationspunkten durch den Vorhof gelegt werden, um das Vorhofflimmern mit möglichst wenigen Narben zuverlässig zu terminieren und ein erneutes Aufflackern zu verhindern? Diese Frage können Computermodelle beantworten: Relativ leicht lassen sich verschiedene Strategien am virtuellen Vorhof ausprobieren. Mit einem „Reset“ kann man immer wieder von der Ausgangsposition aus neu starten. Dies ist am Patienten nicht möglich: Eine Narbe, die in den Vorhof eingebrannt ist, lässt sich nicht wieder zurücknehmen!

Das hier beschriebene Verfahren zur Optimierung der RF-Ablation bei Vorhofflimmern ist noch nicht in die klinische Praxis umgesetzt. Bis zu seiner Validierung sind noch viele Experimente nötig. Besonders schwierig wird es sein, mit Hilfe von Messungen (Bildern und elektrischen Daten) ein Computermodell vom individuellen Patienten zu

Abb. 7: Simulation der elektrischen Depolarisierung des menschlichen Vorhofs, Transmembranspannung zu verschiedenen Zeitpunkten der Ausbreitung. Obere Reihe: gesunder Vorhof, mittlere Reihe: Vorhofflattern, untere Reihe: Test einer Ablationstherapie [16].



erstellen um eine personalisierte Ablationsstrategie zu entwerfen. Ein Engpass auf dem Weg zum Ziel sind aber auch die Algorithmen zur Berechnung der Zellmodelle und der Feldgleichungen: Hier müssen neue mathematische Methoden zur schnelleren Lösung gefunden werden, um mit der Herzdynamik „mitzuhalten“.

4 MATHEMATIK IN DER THERAPIE- UND OPERATIONSPLANUNG

In den letzten Jahren hat die Mathematik eine zunehmend wichtige Rolle in der medizinischen Planung errungen. Augenfällige Erfolge wurden erzielt in der Krebstherapie Hyperthermie, der Mund-Kiefer-Gesichts-Chirurgie (kurz MKG-Chirurgie), der Leberchirurgie und der orthopädischen Chirurgie.

4.1 ERFOLGSGESCHICHTE: MKG-CHIRURGIE

Problemstellung. Ein „schönes“ Gesicht hat nachweislich soziale Vorteile, es öffnet quasi „Tür und Tor“. Umso nachteiliger, wenn ein Gesicht von Geburt an fehlgebildet oder durch Unfall entstellt ist (vgl. Abb. 8). In der MKG-Chirurgie werden knöcherne Strukturen des Schädels operativ korrigiert. Ziel ist eine sowohl funktionelle als auch ästhetische Wiederherstellung.

Erfolg durch Mathematik. Schon seit Anfang der 90er Jahre haben sich Informatiker mit der computergestützten 3D-Planung von MKG-Operationen auf Basis von CT-Daten beschäftigt (siehe [21] und Zitate dort). Mittlerweile spielt auch die Mathematik bei der Planung von hochkomplexen Operationen eine wichtige Rolle. Ihr wichtigstes Instrument sind *Virtuelle Labore* zur Therapie- und Operationsplanung [2]; darunter versteht man umfangreiche Software-Umgebungen (wie z. B. Amira™[19]), in denen patientenspezifische geometrische Modelle mit partiellen Differentialgleichungen zur Elastomechanik, Fluid-Dynamik oder Diffusion sowie schnellen Algorithmen zu deren Lösung und Visualisierung integriert sind. Noch ehe der Chirurg seinen ersten Schnitt ansetzt, können heute verlässliche Vorhersagen über das postoperative Aussehen gemacht werden.

Abb. 8: Patienten mit Kiefer- und Schädelneubildungen [4]



Abb. 9: Vergleich bei zwei Patienten: vor Operation, nach Operation, Überlagerung mit Prognose [22].



Zur Illustration zeigen wir in Abb. 8 eine Reihe von Patienten, deren Operationsplanung in Zusammenarbeit von Kliniken mit dem ZIB durchgeführt wurde. In Abb. 9 ist ein Vergleich der berechneten Vorhersagen mit dem Resultat der tatsächlichen Operation an zwei Beispielen gezeigt [21].

4.2 MATHEMATIK ALS INNOVATIONSFAKTOR

Im Laufe vieler Jahre Zusammenarbeit von Mathematik und Medizin hat sich ein Grundmuster (Paradigma) herausgeschält, das sich wie folgt darstellen lässt (siehe Abb. 10):

- Erstellung des „virtuellen Patienten“ aus Daten des realen Patienten,
- Mathematische Therapie- oder Operationsplanung im „virtuellen Labor“,
- Transfer der Resultate in die Situation des realen Patienten.

Der erste Teilschritt erfordert den Aufbau eines ausreichend genauen Modells des Patienten im Rechner aus medizinischen Bilddaten. Der zweite Teilschritt beinhaltet die schnelle numerische Lösung von partiellen Differentialgleichungen über der realistischen 3D-Geometrie des individuellen Patienten. Der dritte Teilschritt umfasst Registrierungsverfahren und Navigationstechniken zur exakten Umsetzung der Planung. Allen Teilschritten gemeinsam ist die Bedeutung einer effizienten 3D-Visualisierung. Die zugehörige Mathematik und Informatik soll hier nicht weiter vertieft werden; Interessierte seien als Einstieg etwa auf die Arbeiten [4, 21, 22] verwiesen.

Auf dem Weg vom realen zum virtuellen Patienten sind eine Reihe von Teilschritten zu bewältigen, die selbst eine Menge Mathematik enthalten.

Geometrische Modellierung [23]. Die tomografische Bildgebung (vgl. Abschnitt 1.1) liefert einen Stapel von 2D-Schnittbildern. Für eine Therapieplanung sind jedoch unbedingt 3D-Modelle der individuellen Anatomie erforderlich. Deshalb mussten zunächst

Verfahren bereitgestellt werden, um aus diesem Stapel von 2D-Informationen ein verlässliches geometrisches 3D-Patientenmodell zu erzeugen.

Als erste Teilaufgabe fällt hierbei die *Segmentierung* an, bei der aus der reinen Dichte-Information der CT- oder MRT-Bilder durch Zusammenfassung von Teilgebieten (Klassifikation) die organische Substruktur hergeleitet werden soll. Diese Substruktur ist eine der zentralen Voraussetzungen zur Erstellung eines funktionellen Patientenmodells. Bisherige Methoden (z. B. region growing, watershed methods, intelligent scissors, active contours etc.), die von dem 2D-Stapel ausgingen, um das 3D-Modell zu erzeugen, waren zu großen Teilen *interaktiv*, d. h. sie benötigten noch den Eingriff von Experten. Aus diesem Grund sind sie für den Einsatz in der klinischen Praxis zu zeitaufwendig. Eine neue Generation von *vollautomatischen* Methoden (siehe z. B. [10]) benutzt die Methodik der *statistischen 3D-Formanalyse*: Dabei geht man von einem gemittelten 3D-Modell in Kombination mit "wesentlichen Koordinaten" aus, das mittels PCA (principal component analysis) durch „Training“ an bekannten Patientendaten aufgebaut worden ist. Mit jedem segmentierten Datensatz kann die Trainingsmenge erweitert werden. Die Nutzung von Formwissen zur Segmentierung imitiert somit genau den Prozess im Kopf des Radiologen, der ja, durch Ausbildung und Erfahrung, über eine 3D-Vorstellung der Situation verfügt, in die er das jeweils betrachtete 2D-Bild kognitiv einfügt. Einen Einblick in diese Methodik geben wir in Abb. 11.

Als Resultat der Segmentierung erhält man *Flächengitter* aller Gewebegrenzen (außen wie innen). Um eine möglichst geringe Anzahl an Gitterknoten bei gleichzeitigem Erhalt der Approximationsqualität zu erzielen, werden die so erhaltenen Flächengitter krümmungsabhängig ausgedünnt. Diese reduzierten Flächengitter bilden wiederum die Basis für eine Auffüllung zu *Raumgittern* durch Tetraeder (siehe [23] und Zitate dort). Solche 3D-Gittertypen eignen sich besonders gut zur sukzessiven Verfeinerung – eine

Abb. 10: Hyperthermie: Realer und virtueller Patient im Applikator [3]



Abb. 11: Statistische 3D-Formanalyse [10]. Links: Testmenge bekannter Leberdaten unterschiedlicher Patienten. Rechts: Segmentierung neuer Leberdaten (aus Darstellungszwecken nur zwei CT-Schnitte).



Eigenschaft, die im Zusammenspiel mit *adaptiven Mehrgittermethoden* für die schnelle Lösung partieller Differentialgleichungen zentral wichtig ist.

Beispiel: Osteotomieplanung [21]. Unter Osteotomie versteht man das Schneiden von Knochen zum Zwecke der chirurgischen Verlagerung von Knochensegmenten. Dieser Teil der Planung hat in erster Linie eine funktionelle Rehabilitation zum Ziel. Sind dabei mehrere Knochensegmente beteiligt, die in Relation zueinander angeordnet werden müssen oder gibt es unterschiedliche Therapievarianten, dann stellt das zu erwartende ästhetische Ergebnis ein weiteres wichtiges Kriterium dar, das in der Planung berücksichtigt werden sollte. Unterschiedliche Behandlungsvarianten lassen sich am Rechner kostengünstig planen und geben dem Operateur mehr Sicherheit bei komplexen, Knochen verlagernden Eingriffen. Die Simulation der aus einer Knochenverlagerung resultierenden räumlichen Weichgewebeanordnung erlaubt eine zusätzliche Planungsbewertung aus ästhetischer Sicht. Grundlage für eine verlässliche Prognose der Weichgewebeanordnung nach geplanter Verlagerung knöcherner Strukturen ist zum einen ein adäquates geometrisches Modell des Weichgewebevolumens mit allen eingebetteten Strukturen und angrenzenden Knochenrandflächen und zum anderen ein physikalisches Deformationsmodell, das die mechanischen Eigenschaften von biologischem Weichgewebe in guter Näherung beschreibt.

Schon mit dem geometrischen Modell allein kann man wichtige Aufgaben in der Medizin erledigen, etwa

- einen Tumor genau lokalisieren, damit chirurgische Zugangswege oder auch strahlentherapeutische Maßnahmen exakt geplant werden können,
- ein für den Patienten geeignetes Hüft-, Knie- oder Zahnimplantat auswählen und präziser in den Körper implantieren,
- eine Knochenumstellung (Osteotomie) besser planen und computer- bzw. roboterunterstützt durchführen.

Auf dem Weg vom geometrischen zu einem *funktionellen* Modell sind jedoch noch weitere mathematische Teilschritte nötig, die im Folgenden kurz dargestellt werden sollen.

Mathematische Modellierung und Simulation. Ein funktionelles Patientenmodell umfasst, über das geometrische Modell hinaus, zusätzlich eine hinreichend genaue mathematisch-physikalische Beschreibung, meist durch Systeme von partiellen Differentialgleichungen. Zur Illustration seien einige Beispiele genannt: Lamé-Navier-Gleichungen der linearen Elastomechanik und nichtlineare Verallgemeinerungen (Geometrie und Materialeigenschaften) in der Biomechanik, Maxwell-Gleichungen und Bio-Wärmeleitungs-Gleichungen in der Krebstherapie Hyperthermie, Navier-Stokes-Gleichungen zur Analyse von Strömungsverhältnissen bei Plaquebildung in Blutgefäßen und in Aneurysmen. Oft kann man sich auch mit einfacheren, sogenannten reduzierten Modellen begnügen, falls diese eine Beantwortung der gestellten medizinischen Fragen erlauben. Allgemein gilt für mathematische Modelle: sie taugen nur dann etwas, wenn ihre Inputparameter auf ihre Sensitivität hin überprüft worden sind (Sensitivitätsanalyse).

Eine typische Eigenschaft der medizinischen Modelle ist ihre *Multiskalenstruktur*: Die mathematischen Gleichungen stellen in aller Regel Bezüge zwischen mikroskopisch kleinen Raumdimensionen und unseren Alltagsdimensionen her. Dringt man ausreichend tief in die mathematische Beschreibung ein, so erhält man sogar eine ganze Hierarchie von Skalen, die, je nach Fragestellung, in Betracht gezogen werden müssen. Ein anschauliches Beispiel liefert das internationale Projekt *PHYSIOME* [8], das Skalen von nm (Moleküle) über mm (Gewebe) bis zu m (Organe) überstreicht.

Virtuelles Labor. Die im Modell auftretenden partiellen Differentialgleichungen müssen schnell und verlässlich numerisch gelöst und, mit Blick auf die klinische Anwendung, in eine 3D-Visualisierungsumgebung, ein „Virtuelles Labor“, eingebettet werden. Was die Simulation der mathematischen Modelle betrifft, so zeigt sich ein schöner Gleichklang der Zielsetzung von Mathematik und Medizin: Beide Disziplinen wünschen sich die Lösung in 3D, schnell und mit verlässlicher Genauigkeit. Nur so kann eine mathematische Therapie- oder Operationsplanung in der Klinik Fuß fassen und als Basis für verantwortungsvolle medizinische Entscheidungen dienen. An effizienten *Algorithmen* bieten sich hierzu an: (a) *Gebietszerlegungsmethoden* bei festen Gittern im Verbund mit Parallelisierung oder (b) *Mehrgittermethoden* in Kombination mit *Adaptivität* in Raum und Zeit.

Beispiel: Hyperthermie [2]. Forschungen in der regionalen Hyperthermie waren bisher aus ethischen Gründen eingeschränkt auf die Behandlung tiefliegender inoperabler Tumoren, also auf die bis dato hoffnungslosen Fälle. Ziel dieser Behandlungsmethode ist, Krebszellen durch Erwärmung gegen eine Radio- oder Chemotherapie empfindlicher zu machen, ohne dabei gesundes Gewebe durch hohe Temperaturen zu schädigen. In Abb. 10 ist das Setting der Methode gezeigt. Der Patient liegt in einem Applikator mit (hier) acht Antennen, die Radiowellen in den Körper einstrahlen. Die Wärme wird also nichtinvasiv durch elektromagnetische Felder erzeugt. Durch separate Ansteuerung der einzelnen Applikatorantennen kann das Interferenzfeld in gewissem Rahmen auf jeden Patienten individuell angepasst werden. Die wissenschaftliche Frage lautet: Wie sind die Antennen (Amplituden und Phasen) anzusteuern, sodass der Tumor auf eine Temperatur zwischen $42,5^{\circ}$ und 45°C aufgeheizt wird, nicht aber gesundes Gewebe. Die Auswirkung der einzelnen Einstellgrößen auf die therapeutisch wirksame Temperaturverteilung im Körper ist so komplex, dass nur mit numerischer Simulation optimale Therapiepläne gefunden werden können. Das funktionelle Patientenmodell umfasst hier die Maxwell-Gleichungen zur Beschreibung der elektrischen Felder und die Bio-Wärmeleitungs-Gleichung, welche die Temperaturverteilung im Innern des Körpers beschreibt.

Um die Wichtigkeit von Adaptivität im Raum zu illustrieren, zeigen wir in Abb. 12 ein tatsächlich generiertes Gitter für die Lösung der Maxwell-Gleichungen: Mit Adaptivität im Raum erhält man ca. 120 000 Knoten, bei uniformem Raumgitter hätte man dagegen eine geschätzte Zahl von 16 000 000 Knoten benötigt. Die verwendeten Mehrgitteralgorithmen verbrauchen Rechenzeiten proportional zur Anzahl der Knoten; also

Abb. 12: Hyperthermie: Adaptive hierarchische Raumgitter (ZIB)



sind die adaptiven Methoden, in diesem medizinisch relevanten Beispiel und bei vergleichbarer Genauigkeit, etwa um einen Faktor 130 schneller!

Beispiel: Weichgewebesimulation [4]. In der MKG-Chirurgie besteht das mathematische Modell aus den biomechanischen Differentialgleichungen. Sie sind numerisch (mit effizienten Mehrgittermethoden) zu lösen, um eine verlässliche Vorhersage des zukünftigen Aussehens zu ermöglichen – vorausgesetzt, die Operation verläuft erfolgreich und wie geplant. Der jetzige Standard der Vorhersagen ist in Abb. 9 auf S. 450 an zwei Beispielen dargestellt [21]; die bisher erreichte Qualität ist für medizinische Zwecke bereits hervorragend. Neben der verbesserten Operationsvorbereitung führt diese Art von computergestützter Operationsplanung unter Einschluss der Weichgewebeprognose auch zu einer anschaulicheren Patientenaufklärung und damit letztendlich zu einer erhöhten Patientenmotivation (siehe Abb. 13).

Abb. 13: Patientenaufklärung im virtuellen Labor [22]



4.3 PERSPEKTIVE: DER VIRTUELLE PATIENT

Über die beschriebenen medizinischen Anwendungen hinaus hält die Mathematik (unter Einschluss der Informatik) bereits heute eine umfangreiche Methodik bereit, um Therapie- und Operationsplanung zu begleiten. Mittelfristiges Ziel ist, dieser Methodik einen angemessenen Raum im Gesundheitswesen zu verschaffen. Folgende Entwicklungslinien sind absehbar:

Radiologie wird mehr und mehr von der bloßen 2D-Bildinterpretation zur *3D-Modellrekonstruktion* übergehen. Das erfordert umfangreiches „Screening“ individueller Bilddaten mittels automatisierter Segmentierungsverfahren. Die damit verbundene Zunahme patientenspezifischer Daten wird die folgende zweigleisige Entwicklung nach sich ziehen: (a) den Aufbau zentraler medizinischer Datenbanken in großen Krankenhäusern (wie etwa der Charité) und (b) die flächendeckende Einführung persönlicher Datenträger in der Verfügung der Individuen (individuelle elektronische Patientenakte). Google-med ist eine denkbare Form der Speicherung solcher Daten, wird sich jedoch wegen der nationalen Unterschiede im Gesundheitswesen und der nicht garantierbaren Sicherheit individueller Daten modifizieren müssen.

In zentralen Archiven (PACS, S-PACS) werden neben den patientenspezifischen Daten allgemeine *geometrische Anatomiemodelle* auf Basis statistischer Formmodelle vorzuhalten sein, d. h. über die lokale Bevölkerung gemittelte anatomische Formen inklusive wesentlicher Koordinaten, welche die Variabilität der zugrundegelegten Patientenumenge beschreiben. Hinzu kommen werden *mathematische Funktionsmodelle*, d. h. mathematische Beschreibungen der Funktion von Organen oder des Bewegungsapparates durch weiter zu differenzierende Multiskalenmodelle. Erst wenn für ein Organ ein genaues mathematisches Modell vorliegt, dann ist seine Funktion im gesunden wie im kranken Zustand auch wirklich genau genug verstanden.

Die rechtlichen und ordnungspolitischen *Rahmenbedingungen* werden zu klären sein. Radiologen werden sicher weiterhin die rechtliche Verantwortung für die Korrektheit der Interpretation von medizinischen Bilddaten und daraus abgeleiteten anatomischen Modellen behalten. Allerdings müssen Krankenkassen modellgestützte Planungen auf Basis geometrischer 3D-Modelle und mathematischer Funktionsmodelle in ihren Katalog als abrechenbare Leistung hereinnehmen. Diese neue Art von Planung dient neben der medizinischen Indikation auch der anschaulichen Patientenaufklärung, der Ausbildung, der Dokumentation und der Qualitätssicherung.

5 VISIONEN UND HANDLUNGSEMPFEHLUNGEN

In diesem Artikel wurde an ausgewählten Beispielen gezeigt, wie sich durch das Zusammenwirken von Mathematik und Medizin eine dynamische Entwicklung hin zu patientenspezifischen Modellen („virtuelles Herz“ in Abschnitt 2, „virtueller Patient“ in Abschnitt 3) auf Basis moderner medizinischer Bildgebung (Abschnitt 1) vollzogen hat.

Sie wird ganz sicher in naher Zukunft noch weiter Raum greifen. Allerdings ist noch ein weiter Weg zurückzulegen, bis anatomische und medizinisch brauchbare funktionelle Modelle auch nur für die wichtigsten Körperteile und die häufigsten Krankheitsfälle verfügbar sein werden. Hier wird noch auf längere Zeit die Unterstützung der entsprechenden Forschung durch öffentliche Mittel nötig sein. Es wäre gewiss lohnend, diese mathematisch-medizinische Forschung in Zentren oder auch, wegen unserer föderalen Verfasstheit, in bundesweiten Netzwerken zu bündeln. Dazu sind gesundheitspolitische und forschungspolitische Rahmenbedingungen in enger Absprache mit ausgewählten Medizinerinnen, Ingenieuren und Mathematikern anzupassen. Führende Universitätskliniken könnten beispielhaft das Vorhaben unterstützen, mehr Mathematik in der Medizin zu wagen. Dies wäre zweifellos ein wichtiger Schritt in Richtung auf eine *individuelle quantitative Medizin*, bei dem Deutschland die besten Voraussetzungen hätte, den internationalen „Schrittmacher“ zu spielen.

6 LITERATUR

- [1] P. Colli Franzone, P. Deuffhard, B. Erdmann, J. Lang, L. F. Pavarino: *Adaptivity in Space and Time for Reaction-Diffusion Systems in Electrocardiology*. SIAM J. SISC., Vol. 28, No. 3, pp. 942–962, 2006.
- [2] P. Deuffhard: *Differential Equations in Technology and Medicine: Computational Concepts, Adaptive Algorithms, and Virtual Labs*. In: R. Burkhard, P. Deuffhard, A. Jameson, J.-L. Lions, G. Strang (eds.): *Computational Mathematics Driven by Industrial Problems*. Springer Lecture Notes in Mathematics, vol. 1739, pp. 70–125, 2000.
- [3] P. Deuffhard, R. Hochmuth: *Multiscale analysis of thermoregulation in the human microvascular system*. Math. Meth. Appl. Sci., vol. 27, pp. 971–989, 2004.
- [4] P. Deuffhard, M. Weiser, S. Zachow: *Mathematics in Facial Surgery*. Notices of the AMS, vol. 53, no. 9, pp. 1012–1016, 2006.
- [5] O. Dössel: *Kausalität bei der Entstehung, der Diagnose und der Therapie von Krankheiten – aus dem Blickwinkel des Ingenieurs*. Kausalität in der Technik, Berlin-Brandenburg. Akad. Wiss., pp. 69–80, 2006.
- [6] O. Dössel, D. Farina, M. Mohr, M. Reumann, G. Seemann: *Modelling and imaging electrophysiology and contraction of the heart*. In.: T. M. Buzug, D. Holz, S. Weber, J. Bongartz, M. Kohl-Bareis and U. Hartmann (eds). *Advances in Medical Engineering*, Springer, pp. 3–16, 2007.

- [7] C. S. Henriquez, A. L. Muzikant, C. K. Smoak: *Anisotropy, fiber curvature, and bath loading effects on activation in thin and thick cardiac tissue preparations: simulations in a three-dimensional bidomain model*. J. Cardiovascular Electrophysiology, pp. 424–444, 1996.
- [8] P. J. Hunter, T. K. Borg: *Integration from proteins to organs: the Physiome project*. Nature Rev., Mol. Cell Biology, vol. 4. pp. 237–243, 2003.
- [9] P. Jonas, A. K. Louis: *Phase contrast tomography using holographic measurements*. Inverse Problems, vol. 20, pp. 75–102, 2004.
- [10] D. Kainmüller, Th. Lange, H. Lamecker: *Shape constrained automatic segmentation of the liver based on a heuristic intensity model*. In: T. Heimann, M. Styner, B. van Ginneken (eds.). Proc. MICCAI Workshop 3D Segmentation in the Clinic: A Grand Challenge, pp. 109–116, 2007.
- [11] A. Khawaja, O. Dössel: *Predicting the QRS complex and detecting small changes using principal component analysis*. Biomed. Technik, pp. 11–17, 2007.
- [12] A. K. Louis: *Medical Imaging: State of the Art and Future Development*. Inverse Problems, vol. 8, pp. 709–738, 1992.
- [13] A. K. Louis: *Combining Image Reconstruction and Image Analysis with an Application to Two-dimensional Tomography*. SIAM J. Imaging Sciences, Vol. 1, No. 2. pp. 188–208, 2008.
- [14] R. B. Marr, C. Chen, P. C. Lauterbur: *On two approaches to 3D reconstruction in NMR zeugmatography*. In: G.T. Herman, F. Natterer (eds.): Mathematical aspects of computerized tomography, Springer LNMI pp. 225–240, 1981.
- [15] F. Natterer, F. Wübbeling: *Mathematical methods in image reconstruction*. SIAM, 2001.
- [16] M. Reumann, J. Bohnert, B. Osswald, S. Hagl, O. Dössel: *Multiple wavelets, rotors, and snakes in atrial fibrillation – a computer simulation study*. J. Electrocardiology, pp. 328–334, 2007.
- [17] L. A. Shepp, B. F. Logan: *The Fourier reconstruction of a head section*. IEEE Trans. Nucl. Sci. NS-21. pp. 21–43, 1974.
- [18] L. Sörnmo, P. Laguna: *Bioelectrical signal processing in cardiac and neurological applications*. Elsevier Inc., 2005.

- [19] D. Stalling, M. Westerhoff, H.-C. Hege et al.: *Amira: A Highly Interactive System for Visual Data Analysis*. In: C. D. Hansen, C. R. Johnson (eds.): *The Visualization Handbook*. Chapter 38, pp. 749 - 767, 2005. URL: amira.zib.de
- [20] K. H. ten Tusscher, A. V. Panfilov: *Alternans and spiral breakup in a human ventricular tissue model*. *Am. J. Heart Circulation Physiology*, pp. 1088–1100, 2006.
- [21] S. Zachow: *Computergestützte 3D-Osteotomieplanung in der Mund-Kiefer-Gesichtschirurgie unter Berücksichtigung der räumlichen Weichgewebeanordnung*. Medizininformatik, Verlag Dr. Hut, 2005.
- [22] S. Zachow, H.-C. Hege, P. Deuflhard: *Computer Assisted Planning in Cranio-Maxillofacial Surgery*. *J. Comput. Inform. Technology – Special Issue on Computer-Based Craniofacial Modelling and Reconstruction*, pp. 53–64, 2006.
- [23] S. Zachow, M. Zilske, H.-C. Hege: *3D Reconstruction of Individual Anatomy from Medical Image Data*. In: *Proc. ANSYS Conference & CADFEM Users' Meeting, Dresden*. Siehe auch: Report 07–41, ZIB, 2007.

> WIRKSTOFFE, MEDIKAMENTE UND MATHEMATISCHE BILDVERARBEITUNG

GÜNTER J. BAUER, DIRK A. LORENZ, PETER MAASS, HARTWIG PRECKEL
UND DENNIS TREDE

1 EXECUTIVE SUMMARY: VOM BILD ZUM MEDIKAMENT

Die Entwicklung neuer Medikamente ist langwierig und teuer. Der erste Schritt ist hierbei die Suche nach neuen Wirkstoffkandidaten, die für die Behandlung bislang schwer therapierbarer Krankheiten geeignet sind. Hierfür stehen der Pharma- und Biotechnologieindustrie riesige Substanzbibliotheken zur Verfügung. In diesen Bibliotheken werden die unterschiedlichsten Substanzen gesammelt, die entweder synthetisch hergestellt oder aus Pilzen, Bakterienkulturen und anderen Lebewesen gewonnen werden können.

Welche Wechselwirkungen die ausgewählten Wirkstoffkandidaten in komplexen biologischen Systemen auslösen, muss allerdings in Labortests jeweils neu analysiert werden. In diesen so genannten Screening-Experimenten werden millionenfach Wechselwirkungen ausgewertet: Gegenwärtig einsatzfähige Screening-Systeme analysieren mehr als 100 000 Substanzen pro Tag, die jeweils mit mehreren hundert Zellen in kleinsten Reaktionsgefäßen (Mikrotiterplatten) zusammengebracht werden.

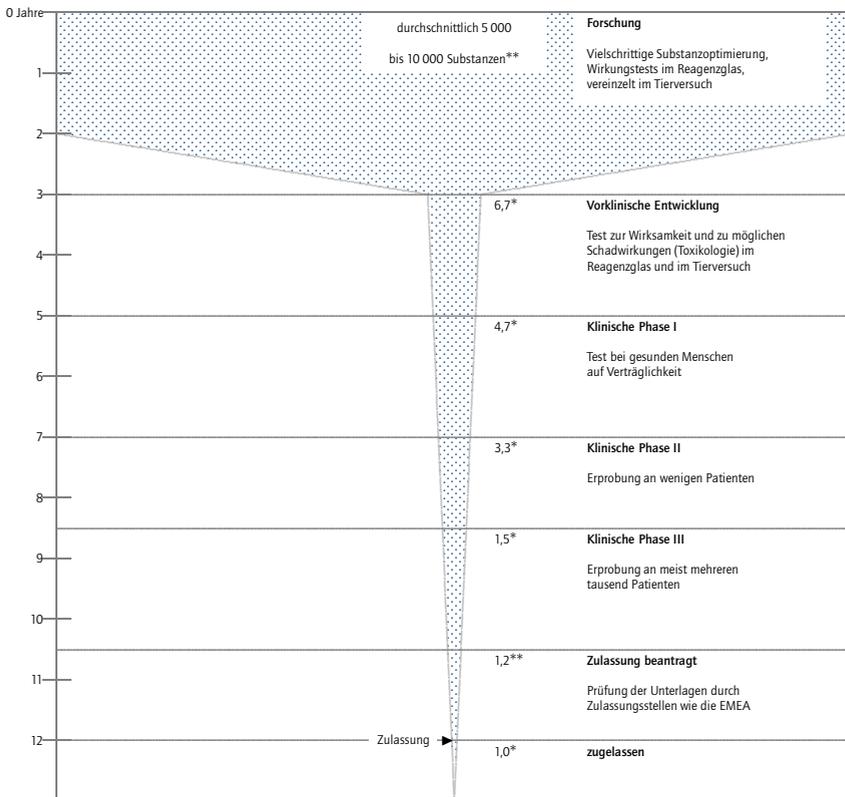
Die Wirkung der einzelnen Substanzen wird dabei farblich kodiert (Fluoreszenz-Marker, Polarisations-Eigenschaften) und mit hochauflösenden optischen Systemen erfasst. Die Aussagekraft dieser Experimente steht und fällt mit der Güte der automatisierten Analyseroutinen für diese gigantischen Datenmengen. Eine zusätzliche Schwierigkeit entsteht dadurch, dass bei der Suche nach geeigneten Wirkstoffen immer komplexere biologisch-chemische Interaktionen auf zellulärer Ebene ausgenutzt werden. Eine standardisierte Auswertung für mehrere Experimente ist damit nicht mehr möglich, die Auswertungsroutinen müssen auf jedes Experiment neu angepasst werden.

Diese Probleme erfordern optimierte, mathematische Methoden, sie sind durch reine biologisch-technische Entwicklungen nicht lösbar. Die revolutionären Entwicklungen auf der Seite der theoretisch-mathematischen Bildverarbeitung der letzten zehn Jahre führten bereits in einigen Anwendungsszenarien zu einer um Größenordnungen verbesserten quantitativen Auswertung von Screening-Experimenten. Das Potential mathematischer Methoden im Bereich der pharmazeutischen Wirkstoffsuche ist allerdings längst nicht ausgeschöpft, die Umsetzung theoretischer Ergebnisse in einsatzfähige Algorithmen

men steht erst am Anfang. Der nächste Qualitätssprung in der Screening-Technologie wird deshalb weniger von einer technischen Weiterentwicklung als vielmehr von verbesserten quantitativen Methoden der mathematischen Bildauswertung erwartet.

Die Bildverarbeitung ist also Anfang und Grundlage der gesamten Entwicklungskette, an deren Ende ein neues Medikament entsteht. Die wesentlichen Herausforderungen der nächsten Jahre in diesem Bereich sind:

Abb. 1: Die Entwicklungsstufen der Arzneimittelforschung: Längst nicht jeder Wirkstoffkandidat wird an komplexen biologischen Systemen ausgetestet. Von anfänglich durchschnittlich 5 000 bis 10 000 Substanzen bleiben nach der Durchführung von Screening-Experimenten nur noch knapp fünf zur Erprobung am Menschen (Klinische Phase I) übrig, Reproduktion nach [13].



* nach J. Di Masi *et al.*, Journal of Health Economics 22 (2003), 151-185
 ** Schätzung des VFA

- Mathematische Modellierung quantitativer Qualitätskriterien für Screening-Experimente,
- Entwicklung von adaptiven Bildverarbeitungsrouitinen zur Bestimmung kritischer Indikatoren in komplexen Zellkulturen.

Diese Herausforderungen müssen von integrierten Teams aus Biologen, Chemikern, Mathematikern und Informatikern angegangen werden. Verbundprojekte zwischen Forschern und ausgewählten Industriepartnern sind Voraussetzung für die zukünftige Wettbewerbsfähigkeit dieser Firmen, wie z. B. Herstellern von Analysegeräten für den chemischen, medizinischen und biochemischen Markt oder Pharmaunternehmen. In Deutschland und der Schweiz sind dies z. B. die PerkinElmer Cellular Technologies Germany GmbH oder die Bruker BioSciences Corporation, und auf der Seite der Pharmaunternehmen z. B. die Bayer Schering Pharma AG, die F. Hoffmann-La Roche AG, die Jenapharm GmbH & Co. KG oder die Schwabe AG.

2 ERFOLGSGESCHICHTEN

Die Suche nach geeigneten Wirkstoffen in der medizinischen Chemie erfolgt über Screening-Experimente. Darunter versteht man die automatische Erforschung und das Testen der Aktivität von Molekülen in Testvolumen von wenigen Mikrolitern. Die Analyse der Wechselwirkung zwischen Wirkstoff und biologischer Probe erfolgt auf Grund eines optischen Signals bzw. eines Bildes, z. B. einer optischen Absorption, der Fluoreszenzintensität oder der Polarisation. Heutige Screening-Systeme analysieren den zeitlichen Verlauf der Wechselwirkung mit hohen und höchsten Durchsatzraten, so genannte High-Throughput- (1000–100 000 Substanzen pro Tag) und Ultra-High-Throughput-Systeme (mehr als 100 000 Substanzen). Eine Weiterentwicklung des klassischen (Ultra)-High-Throughput-Screenings ist das so genannte High-Content-High-Throughput-Screening. Hier werden die optischen Signale mit einer hohen, zellulären Auflösung erfasst und die Wechselwirkung in subzellulären Strukturen untersucht. Abb. 2 zeigt das High-Content-High-Throughput-Screening-System Opera aus dem Hause PerkinElmer Cellular Technologies.

Die Analyse von High-Content-Daten, d. h. die Analyse von Wechselwirkungen in subzellulären Strukturen wie z. B. Rezeptoren, wird häufig erst durch Verwenden fluoreszierender Marker möglich, die in die zu untersuchenden Strukturen gebracht werden. Durch das Verwenden unterschiedlicher Marker können simultan unterschiedliche Strukturen aufgezeichnet werden. Abb. 3 zeigt Zellaufnahmen eines High-Content-High-Throughput-Screenings. Hier findet durch Stimulation der Zellen eine Translokation eines grün gefärbten Proteins statt. Während es sich links homogen in der Zelle verteilt, ist es rechts in der stimulierten Zelle in der Nähe des Zellkerns konzentriert. Die Zellkerne sind rot markiert.

Die Wechselwirkung der biologischen Proben wird in so genannten Mikrotiterplatten untersucht (siehe Abb. 4), die unterschiedlich viele Probelöcher (sog. Wells) enthalten können. Neben der Identifizierung der aktiven Substanzen, d. h. einer qualitativen Aussage ob eine Wechselwirkung stattfindet oder nicht, wird bei der Durchführung des Experiments mit verschiedenen Dosen auch eine quantitative Aussage der Wechselwirkung (Dose-Response-Kurve) erreicht [5].

Die biochemische Wechselwirkung wird bei High-Content-High-Throughput-Screening-Experimenten für jede Zelle einzeln mit Hilfe von Bildverarbeitungsmethoden untersucht. Zur Beschreibung der Gesamtwirkung der zugeführten Substanz wird dann über alle Zellen in einem Well gemittelt. Die zentralen Aufgaben bei der Auswertung solcher Screening-Experimente sind:

Abb. 2: Screening Experimente werden in hoch automatisierten Plattformen zur Analyse der Wechselwirkungen von Wirkstoffen und Zellen durchgeführt [1, 4].



Abb. 3: Translokation eines grün gefärbten Proteins. Links unstimulierte Zellen, rechts stimulierte Zellen.

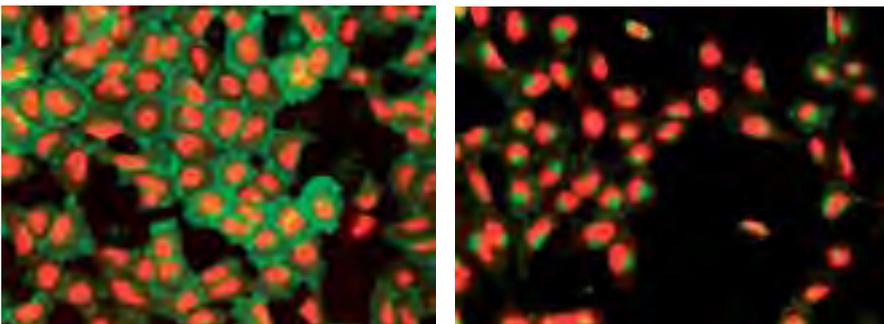
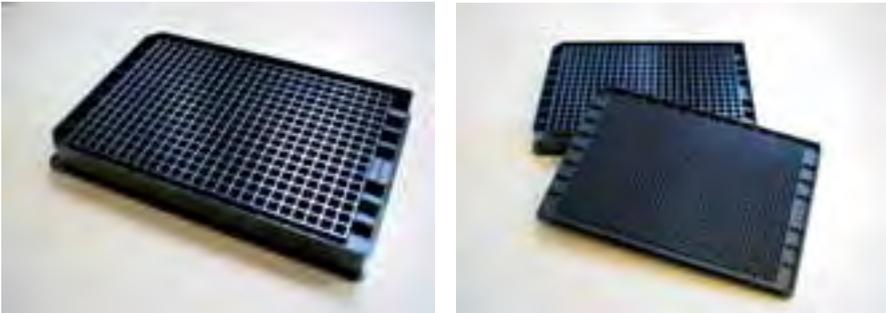


Abb. 4: Die Analyse der Proben erfolgt in Mikrotiterplatten mit Testvolumen von wenigen Mikrolitern.



1. Die Analyse der biochemischen Wechselwirkung zwischen synthetischem Wirkstoff und der biologischen Probe erfolgt auf Grund eines optischen Signals, das als digitales Foto oder als zeitlich aufgelöster Film gespeichert wird. Die Auswertung dieser millionenfach anfallenden digitalen Bilder erfordert die Entwicklung zuverlässiger, adaptiver Methoden der *mathematischen Bildverarbeitung*.
2. Nachdem jedes Bild analysiert wurde, wird unter Einsatz *statistischer Methoden* überprüft, ob und wie stark die Wechselwirkung zwischen synthetischem Wirkstoff und der biologischen Probe ist. Die Aussage muss anschließend statistisch belegt werden (Dose-Response-Kurve) bevor mit nachfolgenden Schritten zur Entwicklung des neuen Arzneimittels fortgefahren werden kann.

Für die Lösung dieser Aufgaben steht ein umfangreiches standardisiertes Arsenal bewährter Methoden der Statistik und Mathematik zur Verfügung, das im Folgenden illustriert werden soll.

2.1 AUTOMATISCHE ANALYSE VON ZELLBILDERN IM HIGH-CONTENT-HIGH-THROUGHPUT-SCREENING

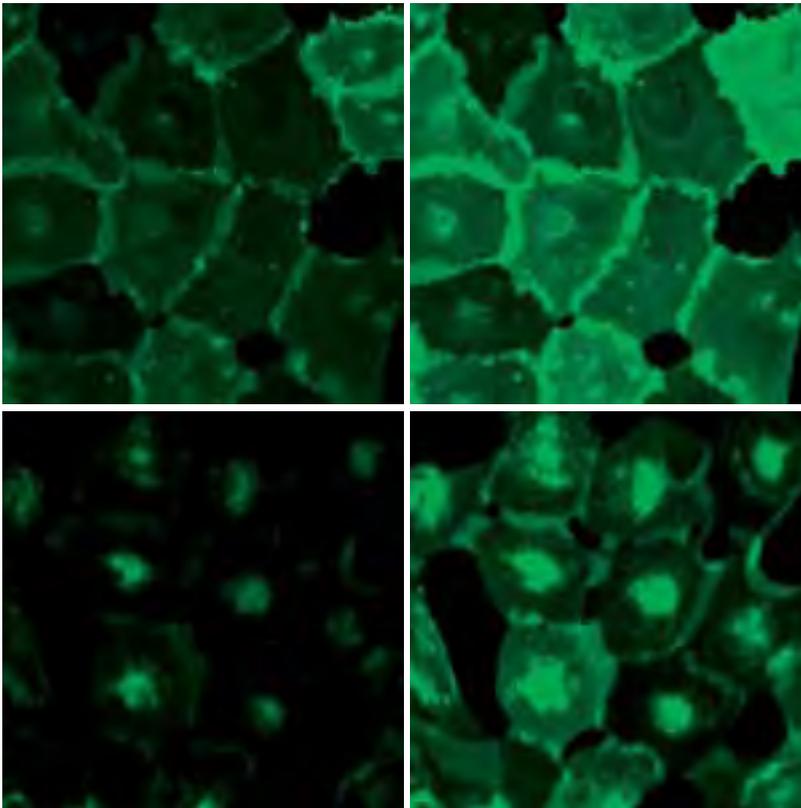
Da beim High-Content-High-Throughput-Screening sehr große Datenmengen anfallen, ist eine manuelle Auswertung der Bilddaten unmöglich. Zur automatischen Auswertung von Zellbildern stehen eine Reihe von Bildverarbeitungssystemen zur Verfügung. Dies ist z. B. die Open-Source-Software CellProfiler [6], und bei PerkinElmer übernimmt die rechnergestützte Verarbeitung der digitalen Bilder die Software Acapella [3]. Das Vorgehen bei der Analyse geschieht dabei immer nach folgendem Schema:

Der erste Schritt der Verarbeitung ist die Vorverarbeitung. Hier wird die Bildaufnahme so modifiziert, dass sie für die nachfolgenden Operationen geeignet ist. Dabei sollte

der wesentliche Informationsgehalt nicht verändert werden. Eine solche Verbesserung der Bilddaten kann z. B. eine Minderung des Rauschanteils oder eine andere Korrektur einer nicht optimalen Bilderfassung sein. Abb. 5 zeigt eine Verstärkung des Kontrastes zweier Beispielbilder.

Nach dieser Aufwertung des Bildes kommt die Segmentierung oder Objekterkennung. Das Ziel dieses Schritts ist die Trennung der interessierenden Objekte von den übrigen Bildteilen. Konkret heißt das z. B. die Trennung von Zellen und Hintergrund oder die Zerlegung einer Zelle in kleine subzelluläre Strukturen. Solche Segmentierungsverfahren können sehr unterschiedliche Komplexität haben. Abb. 6 zeigt eine Segmentierung von Zellkernen und ganzen Zellen.

Abb. 5: Aufwertung von Zellbildern durch Kontrastverstärkung des grünen Kanals. Oben unstimulierte Zellen, unten stimulierte Zellen. Links Originalbilder, rechts vorverarbeitete Bilder.



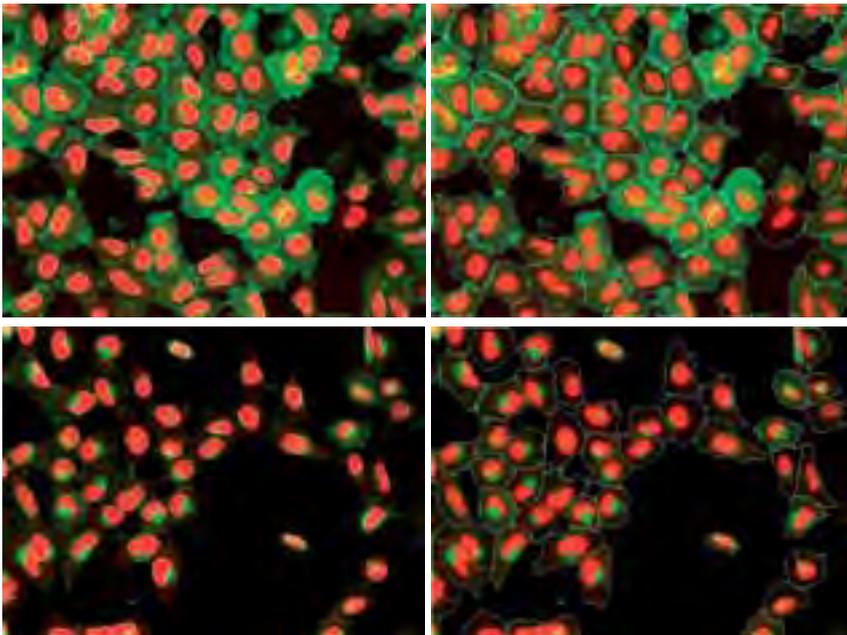
Nachdem nun die Zellen und die zur Analyse interessierenden subzellulären Strukturen im Bild detektiert sind, können auf diesen jeweils Merkmale extrahiert werden (z. B. mittlere Helligkeit, Fläche oder Eigenschaften der Form). Das Ergebnis ist ein Merkmalsvektor pro Zelle, der die Grundlage der Klassifikation bildet: Mit Hilfe der errechneten Merkmale können für jede Zelle einzeln Aussagen zur Quantität der Wechselwirkung gemacht werden. Durch Mitteln über alle Zellen des Wells ergibt sich dann die mittlere Wechselwirkung zwischen synthetischem Wirkstoff und der biologischen Probe und man erhält so für jedes Well Paare

(Dosis des Wirkstoffs, mittlere Stimulation der Zellen).

2.2 STATISTISCHE AUSWERTUNG DER DATEN

Um aktive Substanzen zu identifizieren reicht es aus, Bilder von Zellen ohne Substanzhinzugabe und Bilder von Zellen mit sehr hoch dosierter Substanzhinzugabe zu vergleichen. Zur Beschreibung der *Qualität* der Wirkung, d. h. zur Beschreibung ob die untersuchte Substanz auch bei mehrmaliger Durchführung zuverlässig wirkt, wird in [15] mit dem

Abb. 6: Segmentierung der Zellkerne (links) und der ganzen Zellen (rechts)



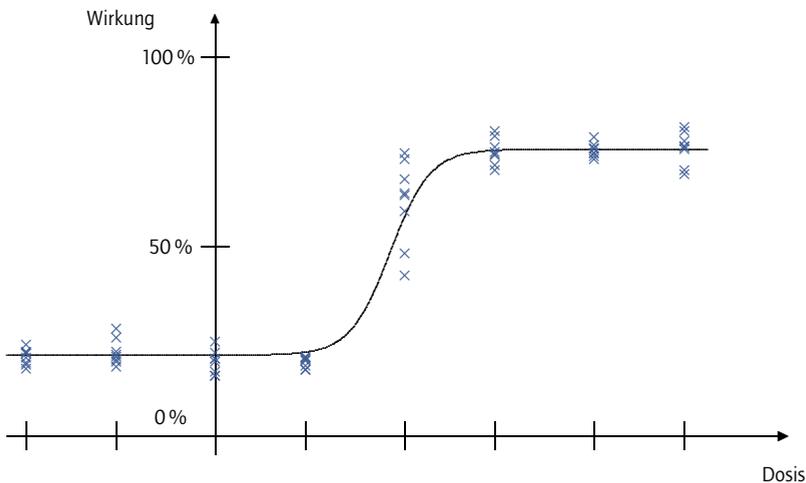
Z' -Wert ein statistisches Kennzeichen vorgestellt. Dieser Wert hängt von Signalen minimaler und maximaler Intensität ab und wird definiert als

$$Z' := 1 - 3 \frac{\sigma_{\text{pos}} + \sigma_{\text{neg}}}{|\mu_{\text{pos}} - \mu_{\text{neg}}|},$$

wobei μ_{pos} , μ_{neg} und σ_{pos} , σ_{neg} die Mittelwerte bzw. Standardabweichungen über die Bilder mit maximaler bzw. minimaler Wirkung sind. Die Wirkung des Wirkstoffs ist um so größer, je näher Z' an Eins liegt.

Zur Entwicklung eines neuen Medikaments ist auch ein Maß zur Beschreibung der *Quantität* der Wirkung nötig. Zur Berechnung eines quantitativen Zusammenhangs von Dosis und Wirkung müssen auch Zellbilder mit mittleren Dosen der Substanz ausgewertet werden. Mit Hilfe nichtlinearer Regression kann eine Ausgleichskurve berechnet werden, vgl. [9–11]. In Abb. 7 sind in blau die Ergebnisse der Bildanalyse und in schwarz eine S-förmige Dose-Response-Kurve dargestellt.

Abb. 7: Beispiel einer S-förmigen Ausgleichskurve (Dose-Response-Kurve)



3 STATUS QUO: MATHEMATISCHE METHODEN FÜR DIE HIGH-CONTENT-HIGH-THROUGHPUT-ANALYSE

Die im vorigen Abschnitt beschriebenen mathematischen Methoden der Bildverarbeitung und Statistik sind zumindest auf der Forschungsseite schon lange bekannt und etabliert. Da jedoch immer komplexere biologisch-chemische Wechselwirkungen auf zellulärer Ebene bei der Medikamentenentwicklung ausgenutzt werden, nimmt die Komplexität der messbaren Effekte in gleichem Maße zu. Es werden immer mehr verfeinerte und optimierte Analysealgorithmen benötigt. Dies ist die Domäne der mathematischen Bildverarbeitung.

Die revolutionären Forschungsergebnisse der letzten Jahren auf Seiten der mathematischen Bildverarbeitung haben allerdings größtenteils in den Lebenswissenschaften noch nicht Einzug gehalten. Es gibt zahlreiche mathematische Methoden, die sich für die Anwendung in der High-Content-High-Throughput-Analyse förmlich aufdrängen.

Im Folgenden soll an zwei Beispielen das erfolgreiche Zusammenspiel von neuesten mathematischen Methoden, biologisch-chemischem Fachwissen und moderner Screening-Technologie demonstriert werden.

1. Die gesamte Prozesskette zur Auswertung des Screening-Bildmaterials besteht aus mehreren Einzelschritten. Jeder dieser Einzelschritte wird über einen eigenen Parametersatz gesteuert. Die bisher übliche getrennte Optimierung dieser Einzelschritte reicht nicht mehr für die Beantwortung der komplexer gewordenen Fragestellungen aus. Es ist also eine globale Optimierung der gesamten Prozesskette erforderlich. Mit der Mathematik steht nun zunächst eine Sprache zur Verfügung, die es ermöglicht, die Güte der Parameter der gesamten Prozesskette zu bewerten und anschließend mit Methoden der *Parameteroptimierung* zu bestimmen.
2. Die Voraussetzung für die Auswertung von Zellversuchen mit Hilfe von Bildverarbeitungsmethoden ist das fehlerfreie Erkennen der einzelnen Zellen, die *Zellsegmentierung*. Häufig werden dafür sehr einfache Methoden (Schwellenwertverfahren) genutzt, die eine hohe Ausfallrate aufweisen. Dies führt dazu, dass das vorhandene Bildmaterial nur teilweise genutzt werden kann. Moderne Verfahren der morphologischen Bildverarbeitung, die durch Gleichungen der Physik motiviert werden, liefern demgegenüber um mehrere Größenordnungen bessere Ergebnisse.

3.1 PARAMETEROPTIMIERUNG UND INVERSE PROBLEME

Die Ergebnisse der Bildverarbeitung hängen entscheidend von geeignet zu wählenden Parametern ab, die in Abhängigkeit von äußeren Einflüssen wie z. B. der Art der Mikrotiterplatte gewählt werden. Dies sind z. B. Schwellenwerte oder statistische Maße, die

in die Beschreibung der Aktivität einer Zelle eingehen. Die Parameter werden häufig von Hand auf Grund biologischer Erfahrungswerte eingestellt. Weiter gibt es Parameterscanner, die in einer anzugebenden endlichen diskreten Teilmenge aller möglichen Parameter nach einer guten Einstellung sucht (Grid-Search). Solch eine Suche ist offensichtlich sehr aufwändig und besitzt keinen Anspruch auf Optimalität.

Wir haben jüngst in interdisziplinärer Kooperation vom Zentrum für Technomathematik der Universität Bremen und der PerkinElmer Cellular Technologies Germany GmbH eine Methode entwickelt, die diese Parameter so einstellt, dass die Analyse bestmöglich geschieht. Die Idee dieser Parameteroptimierung für die High-Content-High-Throughput-Analyse ist, die einstellbaren Modellparameter mit Hilfe der Ergebnisse der statistischen Auswertung durch eine Rückführung iterativ zu verbessern. Dies kann sogar geschehen, ohne dass die Parameter auf Grund biologischer Begründungen voreingestellt wurden (siehe Abb. 8). Für die Optimierung kann z. B. der Z' -Wert genutzt werden: Die unstimulierten und stimulierten Zellen unterscheiden sich umso mehr, je näher Z' der Eins kommt. Für die Parameteroptimierung betrachten wir Z' als mathematische Funktion der Parameter, dessen Maximum berechnet werden soll.

In [12] wurde die Parameteroptimierung auf einige Beispielversuche angewandt und hat bei allen Versuchen eine deutliche Verbesserung der Analyse hervorgebracht. Abb. 9 zeigt die Ergebnisse der Optimierung eines Beispiels: Auf der linken nichtoptimierten Dose-Response-Kurve ist der Unterschied zwischen unstimulierter und stimulierter Zelle mit ca. 20 % nicht sonderlich deutlich. In der rechten Kurve sind unstimulierte

Abb. 8: Verarbeitungsschritte bei der High-Content-High-Throughput-Analyse und Konzept für die Parameteroptimierung

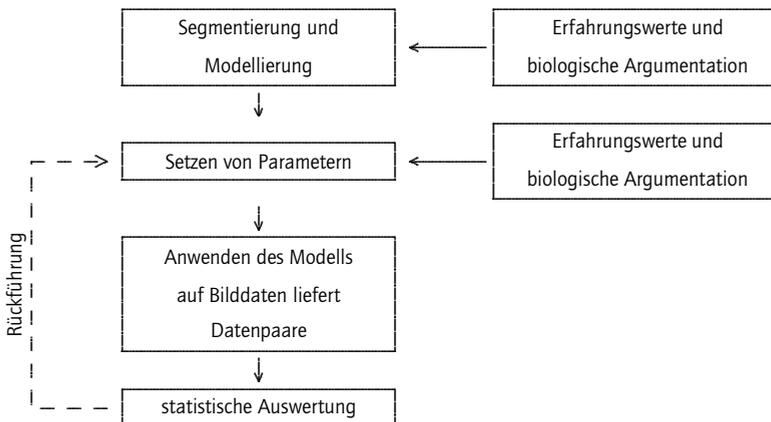
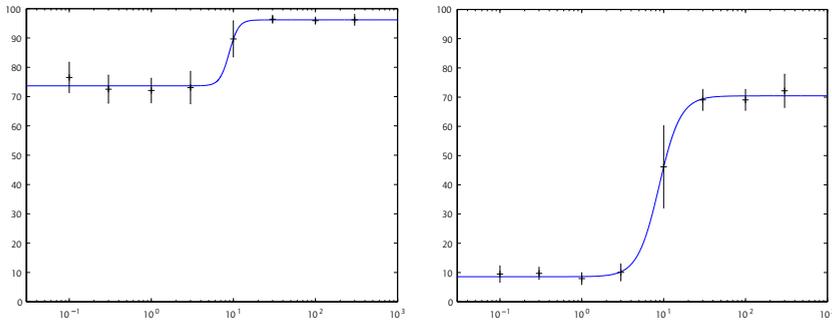


Abb. 9: Anwendung der Parameteroptimierung auf einen Beispielversuch: In der linken nichtoptimierten Dose-Response-Kurve ist der Unterschied der Reaktionen unstimulierter und stimulierter Zellen nicht sonderlich deutlich, in der rechten optimierten Kurve sind die Abweichungen weitaus prägnanter, aus [12].



und stimulierte Zellbilder mit einem Unterschied von ca. 60 % sehr viel besser zu unterscheiden.

Es haben sich mehrere mögliche Anwendungsgebiete der Parameteroptimierung aufgezeigt:

1. Das Optimierungsverfahren kann nach einer Modellierung der Wechselwirkung und groben Voreinstellung der Parameter zur Feineinstellung dieser genutzt werden.
2. Das Optimierungsverfahren eignet sich außerdem zum Einstellen der Parameter von Standard-Bildverarbeitungsroutinen ausgehend von Default-Parametern. Dies ist sogar möglich, falls die Funktionsweise der zu optimierenden Methode nicht bekannt ist, und damit ist der Algorithmus für die Parameteroptimierung einer Black-Box nutzbar.
3. Weiterhin kann die hergeleitete Methode angewendet werden, um die Parameter eines bestehenden Modells an andere aber sehr ähnliche Versuchsbedingungen anzupassen, wie z. B. andere Mikrotiterplatten oder Belichtungszeiten.

Zusammenfassend kann man sagen, dass die Parameteroptimierung mit Hilfe der Rückführung (Inverses Modell) der statistischen Auswertung ein leistungsfähiges Hilfsmittel bei der High-Content-High-Throughput-Analyse ist. Das Einstellen der Parameter von Hand kostet viel Arbeitszeit und erfordert umfassende Erfahrung bei der Analyse von Screening-Daten. Durch Verwenden der vorgestellten Parameteroptimierung kann kostbare Zeit in der Entwicklungsphase neuer Wirkstoffe eingespart werden und

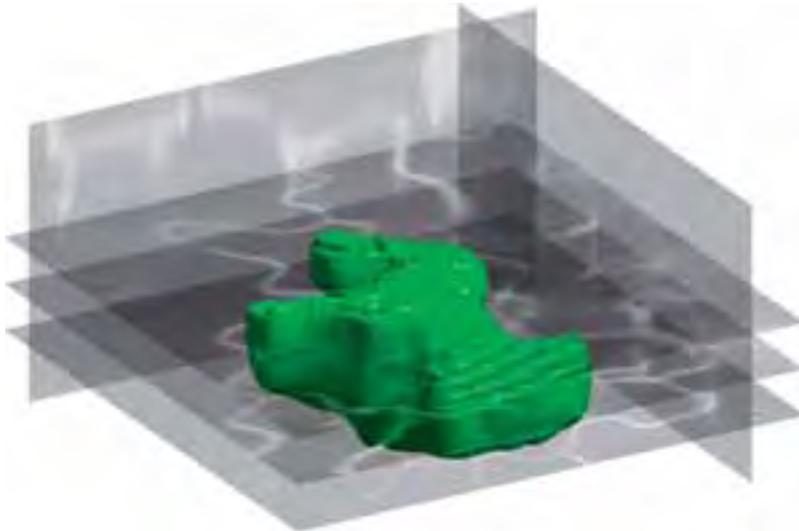
somit können Medikamente schneller auf den Markt kommen. Jüngst haben wir die Parameteroptimierung als Schutzrecht beim Europäischen Patentamt, Aktenzeichen EP08155784.5, unter dem Titel „Verfahren zur Analyse der Wirkung einer Testsubstanz auf biologische und/oder biochemische Proben“ angemeldet.

3.2 VERBESSERTER ZELLSEGMENTIERUNG

Die meisten Segmentierungsverfahren für Zellen und Zellkerne basieren auf einfachen Schwellenwertverfahren. Sehr vereinfacht geschrieben sind dies Methoden, die das Bild durch den Vergleich der Helligkeit der Pixel in Segmente unterteilen. Ein Vorteil der Schwellenwertverfahren ist unter anderem der geringe Rechenaufwand, im Allgemeinen ist die Qualität der Segmentierung dieser vergleichsweise einfachen Verfahren allerdings schlechter als bei komplexeren Verfahren. Probleme treten z. B. bei der sehr schwierigen Segmentierung der ganzen Zellen auf.

Wir haben am Zentrum für Technomathematik der Universität Bremen in Kooperation mit PerkinElmer Cellular Technologies eine Methode entwickelt und getestet, die den vergleichsweise jungen Ansatz der aktiven Konturen zur Segmentierung verwendet, vgl. [7, 12]. Eine sehr ähnliche Herangehensweise wird in [14] gewählt. Das Segmen-

Abb. 10: Dreidimensionale Zellsegmentierung mit Hilfe aktiver Konturen



tieren der Zellen geschieht bei dem Verfahren der aktiven Konturen durch Lösen der partiellen Differentialgleichung

$$\frac{\partial u}{\partial t} = g \cdot \left(\operatorname{div} \left(\frac{\nabla u}{\|\nabla u\|} \right) + \rho \right) \|\nabla u\| + \langle \nabla g, \nabla u \rangle,$$

wobei u die Bildsegmente implizit beschreibt, g eine Steuerfunktion ist und ρ einen weiteren Parameter darstellt. Die Gleichung wird motiviert durch den physikalischen Transport von Masse und wird in sehr ähnlicher Art und Weise auch in der Strömungsmechanik verwendet [2,8]. Zur Berechnung von Lösungen der Differentialgleichung existieren numerische Methoden, die die Lösung auf einem diskreten Gitter näherungsweise berechnen.

Ein weiterer Vorteil der aktiven Konturen gegenüber vielen anderen Verfahren ist, dass es leicht auf Daten höherer Dimension verallgemeinert werden kann. Der Zellanalysator Opera besitzt die Besonderheit, dass Bilder von Schnittebenen erzeugen werden können und damit auch räumlich-dreidimensionale Zellanalysen möglich sind. In Abb. 10 sind übereinanderliegende Schnitte, die mit dem Gerät aufgenommen wurden, und das Ergebnis einer dreidimensionalen Segmentierung mit der entwickelten Methode dargestellt.

4 STÄRKEN/SCHWÄCHEN-ANALYSE: DAS DILEMMA DISZIPLINÄREN DENKENS

Der Einsatz moderner mathematischer Methoden in der Medikamentenforschung und insbesondere bei der Auswertung von Screening-Experimenten zur Wirkstoffsuche steht noch am Anfang. Weit verbreitet sind nach wie vor klassische Verfahren aus Standard-Softwarepaketen, an denen die Forschungsergebnisse der letzten zehn Jahre spurlos vorbeigegangen sind. Wieso ist dies so?

Die Labore in den experimentell arbeitenden Abteilungen der Pharmaunternehmen sind bevölkert von Biologen und Chemikern, die Abteilungen für technische Entwicklung und Softwareentwicklung werden dominiert von Physikern und Informatikern. Mathematiker sind in diesem Umfeld eine äußerst seltene Spezies. Allenfalls werden Statistiker in bioinformatische Arbeitsgruppen integriert. Wenn aber erstmal der Denkapparat auf eine bestimmte Schiene gesetzt ist, wird die Lösung jeglichen Problems in dieser Richtung gesucht.

Der Bereich der Bildverarbeitung fällt aber in einen anderen, neuen Wissensbereich. Das gesunde Selbstverständnis, mit dem gerade Physiker und Informatiker das mathematische Know-How oft für sich beanspruchen, führt hier nicht zum Ziel. Es ist eben nicht damit getan, ein paar vorgegebene Formeln zu implementieren. Die Komplexität sowohl des Anwendungsfeldes als auch der zum Einsatz kommenden Methoden erfordern eine detaillierte Analyse.

Die erfolgreichen, zukünftigen Entwicklungen werden in dem beschriebenen Anwendungsfeld entscheidend von der Kooperation mit Mathematikern geprägt sein. In

einzelnen, hochspezialisierten Anwendungsszenarien hat sich der Nutzen des Einsatzes neuester mathematischer Algorithmen bereits bewiesen. Bis ihr Einsatz in der vollen Breite des Spektrums von der optimierten Auswertung der ersten Zellexperimente bis hin zur Begleitung der letztendlich erforderlichen klinischen Studien zur Routine wird, ist allerdings noch einiges zu tun.

Der Beitrag des Mathematikers liegt hierbei in zwei Bereichen. Zum einen bietet die Mathematik eine ebenso klare wie unbestechliche Sprache zur Formulierung der Zielvorgaben bei der Auswertung experimentell gewonnener Daten. Diese Auswertung erfolgt in der Regel in einer mehrstufigen Prozesskette. Auch wenn dies selbstverständlich klingt, so verstellt der Blick auf die Optimierung der Einzelschritte dieser Prozesskette oft den Blick auf das Ganze. Eine durchgehende mathematische Modellierung bietet eine klare Beschreibung, die gleichzeitig Grundlage für den Einsatz von Optimierungsstrategien für den Gesamtprozess ist. Zum anderen liefert die Mathematik einen ganzen Werkzeugkasten neuer Methoden zur Auswertung der Bilddaten, zur Modellierung einzelner Prozessschritte und zur Optimierung bzw. Parameteridentifizierung der gesamten Prozesskette.

5 VISIONEN UND HANDLUNGSEMPFEHLUNGEN

5.1 KONKRETE HERAUSFORDERUNGEN

Die Verbreitung mathematischer Modelle und Methoden in der Medikamentenentwicklung ist die Herausforderung der nächsten zehn Jahre. Mathematische Methoden können in fast allen Stufen der Medikamentenentwicklung sinnvoll eingesetzt werden. Wir wollen uns in diesem Beitrag auf den ersten Schritt, der Suche nach geeigneten Wirkstoffen, beschränken. Konkret steht bei Wirkstoffsuche die mathematische Modellierung quantitativer Qualitätskriterien für allgemeine High-Content-High-Throughput-Screening-Experimente und die Entwicklung von adaptiven Bildverarbeitungsroutinen zur Bestimmung kritischer Indikatoren in komplexen Zellkulturen im Vordergrund des Interesses.

Die Schwierigkeiten bei der Umsetzung liegen insbesondere in der ständig wachsenden Komplexität der Modellierung chemisch-biologischer Zellprozesse, die jeweils auf mathematischer Seite nachvollzogen werden müssen, und in einer kompetenten Bewertung der Vielzahl neuer mathematischer Verfahren, die zur Auswertung der experimentell gewonnenen Daten eingesetzt werden können.

Die Einführung interdisziplinärer Studiengänge bietet in diesem Bereich allerdings keine Lösung. Der Versuch, in interdisziplinären Studiengängen Biologen genügend Mathematik oder Mathematikern hinreichend Biologie und Chemie beizubringen, und somit hybride Wesen Halb-Biologe-half-Mathematiker zu erzielen, ist zum Scheitern verurteilt. Auf beiden Seiten ist die Komplexität des Wissenschaftsgebietes derart angewachsen, dass mehr als ein grundsätzliches Verständnis der anderen Seite nicht mehr

möglich ist. Entscheidend ist allerdings, dass die zukünftigen Wissenschaftler und Entwickler bereits während des Studiums einige Grundlagen der jeweils anderen Disziplin – und viel wichtiger eine Offenheit und Neugier für deren Ergebnisse – vermittelt bekommen. Dies betrifft auch die Mathematik-Studiengänge, in denen meistens die durchaus denkbaren Nebenfächer Chemie oder Biologie äußerst selten gewählt werden und in denen in Numerik- oder Modellierungs-Vorlesungen biologische Anwendungen meist auf die Diskussion von Räuber-Beute-Modellen beschränkt bleiben.

Letztendlich sind aber keine auch noch so gut ausgebildeten Einzelkämpfer gefragt, es sind vielmehr Teams von in ihrer jeweiligen Disziplin hervorragend ausgebildeten Wissenschaftlern und Entwicklern gefordert. Die Phase der Promotion ist hier der richtige Zeitpunkt für ernsthafte Kooperationen über die disziplinären Grenzen hinaus.

Die Herausforderung besteht in der Schaffung attraktiver Arbeitsbedingungen, unter denen Kooperationen von Biologen, Chemikern, Informatikern und Mathematikern auf höchstem Niveau möglich sind. Dies ist in einzelnen Forschungsinstituten bereits realisiert, die Etablierung einer breiteren, industrienahen Plattform für derartige Kooperationen fehlt jedoch.

5.2 KOOPERATION

Die Zukunft liegt in der kooperativen Umsetzung mathematisch-biologisch-technischer Ansätze zur Optimierung aller unterschiedlichen Prozessschritte in der Medikamentenentwicklung. Mathematiker sind künftig ebenso selbstverständlich in jedem Entwicklungsteam der Pharmaindustrie vertreten wie Biologen, Chemiker und Informatiker.

Das Potential ist offensichtlich, die Hemmschwelle zwischen den Disziplinen ist jedoch nach wie vor hoch. Gezielte Maßnahmen, die Kooperationen zwischen industriellen Entwicklerteams und mathematischen Forschergruppen initiieren, sind notwendig. Eine koordinierte Förderrichtlinie, die von Verbänden, Forschungsförderungsinstitutionen und einzelnen Industrieunternehmen getragen wird, würde mit überschaubarem Aufwand messbare Erfolge erzielen. Zum einen würde in gezielt ausgewählten Anwendungsfeldern der Beitrag moderner Mathematik bis hin zur Entwicklung prototypischer Analyse-Tools demonstriert werden, zum anderen würden hierbei zukünftige Know-How-Träger für die Industrie mit einem hohen Multiplikatoreffekt ausgebildet werden.

6 LITERATUR

- [1] Andreas Beuthner. Bilder aus dem Zellkern, Firmenportrait über das Unternehmen Evotec Technologies. *Fraunhofer Magazin*, 3. Quartal 2004.
- [2] Vincent Caselles, Ron Kimmel, and Guillermo Sapiro. Geodesic active contours. *International Journal of Computer Vision*, 22(1):61–79, 1997.

- [3] Evotec Technologies. Acapella product flyer. http://las.perkinelmer.com/Content/RelatedMaterials/Brochures/BRO_AcapellaSoftware.pdf.
- [4] Evotec Technologies. Opera product flyer. http://las.perkinelmer.com/Content/RelatedMaterials/Brochures/BRO_OperaHighContentScreeningSystem.pdf.
- [5] Christof Fattinger. High Throughput Screening. *Innovation – Das Magazin von Carl Zeiss*, 12:4–5, 2002.
- [6] Thouis R. Jones, Anne E. Carpenter, David M. Sabatini, and Polina Golland. Methods for high-content, high-throughput image-based cell screening. In *Proceedings of the First MICCAI Workshop on Microscopic Image Analysis with Applications in Biology*, pages 65–72, 2006.
- [7] Dirk A. Lorenz, Peter Maass, Hartwig Preckel, and Dennis Trede. Topology-preserving geodesic active contours for segmentation of high content fluorescent cellular images. In *Proceedings in Applied Mathematics and Mechanics (PAMM)*, volume 8, 2008.
- [8] Ravikanth Malladi, James A. Sethian, and Baba C. Vemuri. Shape modeling with front propagation: A level set approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(2):158–175, 1995.
- [9] Nathalie Malo, James A. Hanley, Sonia Cerquozzi, Jerry Pelletier, and Robert Naddon. Statistical practice in high-throughput screening data analysis. *Nature Biotechnology*, 24(2):167–175, 2006.
- [10] Harvey Motulsky and Arthur Christopoulos. *Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting*. Oxford University Press Inc, 2004.
- [11] Christian Ritz and Jens C. Streibig. Bioassay analysis using R. *Journal of Statistical Software*, 12(5), 2005.
- [12] Dennis Trede. Parameteroptimierung und Segmentieren mit aktiven Konturen für die High-Content-Analyse beim Hochdurchsatz-Screening. Diplomarbeit, Universität Bremen, Zentrum für Technomathematik, September 2007.
- [13] Verband forschender Arzneimittelhersteller e.V. In Labors und Kliniken: Wie entsteht ein neues Medikament? <http://www.vfa.de/de/forschung/so-funktioniert-pharmaforschung/wie-ein-medikament-entsteht/amf-laborskliniken.html>, 2008.

- [14] Pingkun Yan, Xiaobo Zhou, Mubarak Shah, and Stephen Wong. Automatic segmentation of high throughput Rnai fluorescent cellular images. *IEEE Transactions on Information Technology in Biomedicine*, 12(1):109–117, 2008.
- [15] Ji-Hu Zhang, Thomas D. Y. Chung, and Kevin R. Oldenburg. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *Journal of Biomolecular Screening*, 4(2):67–73, 1999.

> AUTORENVERZEICHNIS

Prof. Dr.-Ing. **Dirk Abel** ist Professor für Regelungstechnik an der RWTH Aachen in der Fakultät für Maschinenwesen. Er ist stellvertretender Vorsitzender der VDI/VDE-Gesellschaft für Mess- und Automatisierungstechnik (GMA). Seine Forschungsschwerpunkte liegen in regelungs-, steuerungs- und automatisierungstechnischen Aufgaben der Anwendungsgebiete Automotive, Rail, Industrieautomatisierung und Medizintechnik. Methodische Schwerpunkte sind – neben anderen – Modellbasierte Prädiktive Regelungen, auch in Verbindung mit Objektorientierter Systemmodellierung und Simulation.

Prof. Dr.-Ing. **Frank Allgöwer** ist Professor für Systemtheorie technischer Prozesse an der Universität Stuttgart. Seine Forschungsschwerpunkte sind die Entwicklung von neuen Methoden der System- und Regelungstheorie mit speziellem Schwerpunkt auf der nichtlinearen und prädiktiven Regelung; Anwendungsgebiete umfassen u. a. die verfahrenstechnische Prozessregelung, die Mechatronik, die biomedizinische Technik und die Nanotechnologie. Ein weiterer Forschungsschwerpunkt liegt auf dem Gebiet der Systembiologie. Frank Allgöwer wurde mit diversen Preisen ausgezeichnet, u. a. dem Leibniz-Preis der DFG und dem Landeslehrpreis des Landes Baden-Württemberg.

Dr.-Ing. **Matthias Baitsch** hat an der Universität Dortmund Bauingenieurwesen studiert und im Rahmen des Bochumer DFG Graduiertenkollegs „Computational Structural Dynamics“ über die Optimierung imperfektionsgefährdeter Tragwerke promoviert. Aktuelle Forschungsschwerpunkte sind die p -Version der Finite-Element-Methode sowie die Entwicklung moderner Softwarekonzepte für Simulation und Optimierung.

Dr. **Günter J. Bauer**, PerkinElmer Cellular Technologies Germany GmbH, betreut wissenschaftliche und geschäftliche Kooperationen zur Erweiterung des Einsatzspektrums bestehender Produkte und Methoden und zur Evaluierung neuer Technologien und Produktideen.

Prof. Dr. Dr. **Holger Boche** ist Professor für Mobilkommunikation an der TU Berlin und Institutsleiter der Fraunhofer-Institute HHI und MCI. Seine Forschungsschwerpunkte sind Nachrichtentechnik und Informationstheorie sowie angewandte Mathematik. Er wurde mit dem Leibniz-, Vodafone Innovations- und weiteren Preisen ausgezeichnet.

Prof. Dr. Dr. h.c. **Hans Georg Bock** ist Professor für Mathematik und Informatik an der Universität Heidelberg, Direktor des Interdisziplinären Zentrums für Wissenschaftliches Rechnen (IWR) und der Heidelberg Graduate School of Mathematical and Computa-

tional Methods for the Sciences (HGS MathComp) sowie Sprecher des Internationalen Graduiertenkollegs „Complex Processes Modeling, Simulation and Optimization“. Sein Forschungsschwerpunkt ist die Optimierung bei Differentialgleichungen und hybriden Systemen, speziell Echtzeitmethoden der Optimalen Steuerung (NMPC), Parameterschätzung und Optimale Versuchsplanung. Er engagiert sich in der Kooperation mit Südostasien, speziell Vietnam. Hans Georg Bock wurde u. a. mit dem Hausdorff-Preis und dem Microsoft Research Award ausgezeichnet.

Dr. **Ralf Borndörfer** ist stellvertretender Leiter der Abteilung Optimierung des Konrad-Zuse-Zentrums und Gesellschafter der Dres. Löbel, Borndörfer und Weider GbR. Sein Arbeitsgebiet ist die Lösung von Problemen der Verkehrsplanung mit Methoden der diskreten Optimierung. Er wurde mit dem Joachim-Tiburtius-Preis und dem Dissertationspreis der Gesellschaft für Operations Research ausgezeichnet.

Prof. Dr. **Wolfgang Dahmen** ist Professor für Mathematik an der RWTH Aachen, Mitglied der Graduiertenschule „Aachen Institute for Advanced Studies in Computational Engineering Science (AICES)“, der Nordrhein-Westfälischen Akademie der Wissenschaften, des Senats der DFG und des Panels PE1 für Mathematik des European Research Councils. Seine Forschungsinteressen im Bereich Numerik, Harmonischer Analyse und Approximationstheorie reichen von grundlegenden Untersuchungen zur Entwicklung und Analyse adaptiver Lösungskonzepte im Rahmen von Multiskalenmethoden, zu interdisziplinären Anwendungen in Verfahrenstechnik und Aerodynamik. Er wurde mit dem Leibniz-Preis der DFG ausgezeichnet.

Prof. Dr.-Ing. **Berend Denkena** ist Leiter des Instituts für Fertigungstechnik und Werkzeugmaschinen (IFW) der Universität Hannover. Nach praktischen Erfahrungen in einer Schlosserlehre studierte er Maschinenbau und Betriebswirtschaft. Er war nach seiner Promotion als Konstrukteur und später als Entwicklungsleiter in der Werkzeugmaschinenpartie der Thyssen AG in Deutschland und den USA tätig. Danach wirkte er bei der Gildemeister AG, wo er die Entwicklung und Konstruktion von Drehmaschinen führte. Seine Forschungsgebiete liegen im Bereich neuer Werkzeugmaschinen und spanender Fertigungstechnologien. Er ist Sprecher des SFB 653 „Gentelligente Bauteile im Lebenszyklus“ und Koordinator des DFG Schwerpunktprogramms 1180 „Wechselwirkungen zwischen Prozessen und Strukturen“.

Prof. Dr. Dr. h.c. **Peter Deuffhard** ist Professor für Scientific Computing an der FU Berlin, Gründer und seit 1987 Präsident des Zuse-Instituts Berlin (ZIB) sowie Mitinitiator des DFG-Forschungszentrums MATHEON. Seine mathematischen Forschungsinteressen liegen im Bereich nichtlineare Systeme, gewöhnliche und partielle Differentialgleichungen sowie Markov-Ketten, sowohl in der Numerik als auch in der mathematischen Modellierung. Er arbeitete an Problemen aus Raumfahrttechnik, Chemie- und Verfahrens-

technik, Medizintechnik und Biotechnologie. Er wurde u. a. mit der Damköhler-Medaille der Chemischen Industrie und dem ICIAM Maxwell-Preis für grundlegende Beiträge zur angewandten Mathematik ausgezeichnet.

Prof. Dr. **Moritz Diehl** ist Professor im Department für Elektrotechnik der Katholieke Universiteit Leuven und Leiter des dortigen Optimization in Engineering Center OPTEC. Hauptarbeitsgebiete sind effiziente und maßgeschneiderte Optimierungsalgorithmen für Ingenieursanwendungen, speziell für die lineare und nichtlineare modell-prädiktive Regelung (MPC) und die online Zustands- und Parameterschätzung. Er arbeitet an Anwendungen in Mechatronik, Robotik, Verfahrenstechnik, Automobiltechnik, und Energietechnik.

acatech-Mitglied Prof. Dr. **Olaf Dössel** leitet das Institut für Biomedizinische Technik an der Universität Karlsruhe (TH). Er ist Mitglied im Präsidium des VDE und stellvertretender Vorsitzender der Deutschen Gesellschaft für Biomedizinische Technik im VDE. Er ist außerdem Mitglied im Kuratorium der Physikalisch-Technischen Bundesanstalt PTB und Mitglied der Berlin-Brandenburgischen Akademie der Wissenschaften. Seine Forschungsgebiete sind die mathematische Modellierung des Herzens, das inverse Problem der Elektrokardiographie (EKG), die EKG-Signalanalyse und die numerische Feldberechnung im menschlichen Körper. Zusammen mit Prof. Dr. Wolfgang Schlegel wird er den „World Congress on Medical Physics and Biomedical Engineering 2009“ leiten.

Priv. Doz. Dr. rer. nat. **Wolfgang Dreyer** ist Leiter der Forschungsabteilung Thermodynamische Modellierung und Analysis von Phasenübergängen am Weierstraß-Institut für Angewandte Analysis und Stochastik und seit 2008 Honorarprofessor an der TU Berlin. Seine Forschungsschwerpunkte sind Mathematische Modellierung, Phasenübergänge, Thermodynamik, Statistische Mechanik und Kinetische Theorie.

Prof. Dr.-Ing. **Jörg Eberspächer** leitet den Lehrstuhl für Kommunikationsnetze an der TU München. Er war viele Jahre bei der Siemens AG in der Forschung auf dem Gebiet der Kommunikationssysteme tätig. Seine Forschungsschwerpunkte sind Architekturen und Technologien für das Next Generation Internet, Mobile Kommunikationssysteme und Netzplanung. Er ist Gastprofessor an der Tongji-Universität Schanghai, Vorsitzender mehrerer wissenschaftlicher Beiräte von Forschungsinstituten, Mitglied der Nationalen Akademie der Wissenschaften Leopoldina und Vorstandsmitglied des Münchner Kreises.

Prof. Dr.-Ing. **Wolfgang Ehlers** ist Professor für Kontinuumsmechanik an der Universität Stuttgart und leitet dort den Exzellenzcluster „Simulation Technology“. Er war Mitglied im Senat und im Hauptausschuss der DFG und Prorektor seiner Universität. Seine Arbeitsgebiete umfassen die Kontinuumsmechanik, die Materialtheorie, die Experimentelle und die Numerische Mechanik. Dabei ist er insbesondere an der Modellierung mehr-

phasiger Materialien bei Anwendungen im Bereich der Geomechanik und der Biomechanik interessiert. Für seine Habilitationsschrift wurde er mit dem Gottschalk-Diederich-Baedeker-Preis ausgezeichnet.

Dr. **Andreas Eisenblätter** ist Geschäftsführer der atesio GmbH. Er gründete atesio als Spin-Off des Zuse-Instituts Berlin zusammen mit Kollegen im Jahr 2000. atesio unterstützt Betreiber von Telekommunikationsnetzen bei der kurz-, mittel- und langfristigen Planung ihrer Netze. Seine Forschung zur Optimierung von Funknetzen wurde mit internationalen Preisen ausgezeichnet.

Prof. Dr.-Ing. **Sebastian Engell** vertritt das Fach Systemdynamik und Prozessführung in der Fakultät Bio- und Chemieingenieurwesen an der TU Dortmund. Er war Prorektor für Forschung an der Universität Dortmund. Seine Forschungsschwerpunkte sind die Steuerung, Regelung und Online-Optimierung verfahrenstechnischer Prozesse, optimale Ressourcenzuteilung und hybride (gemischt kontinuierlich-diskrete) dynamische Systeme. Er erhielt den Joseph-von-Fraunhofer-Preis und ist Fellow der International Federation of Automatic Control (IFAC).

Dr. **Jochen Garcke** ist Nachwuchsgruppenleiter am MATHEON und der TU Berlin. Seine Forschung beschäftigt sich insbesondere mit der Untersuchung von Methoden der numerischen Mathematik für Anwendungen im Data Mining und Maschinellen Lernen.

Prof. Dr. **Michael Griebel** ist Mathematikprofessor an der Universität Bonn. Er ist dort Direktor am Institut für Numerische Simulation und Sprecher des SFB 611 „Singuläre Phänomene und Skalierung in mathematischen Modellen“. Seine Forschungsschwerpunkte sind die Numerische Simulation und das Wissenschaftliche Rechnen mit den Anwendungsbereichen Strömungsmechanik, Materialwissenschaften, Quantenchemie, Finanzwissenschaften sowie insbesondere Datenanalyse und Data Mining.

acatech-Mitglied Prof. Dr. Dr. h.c. mult. **Martin Grötschel** ist Mathematikprofessor an der TU Berlin, Vizepräsident des Zuse-Instituts (ZIB), Mitglied des MATHEON-Vorstandes und Generalsekretär der International Mathematical Union. Seine mathematischen Forschungsschwerpunkte sind Optimierung, Diskrete Mathematik und Operations Research; seine Anwendungsfelder u. a. Telekommunikation, Verkehr und Produktion. Er wurde mit dem Leibniz-, dem Karl Heinz Beckurts- und weiteren Preisen ausgezeichnet.

Prof. Dr. **Lars Grüne** ist Professor für Angewandte Mathematik am Mathematischen Institut der Universität Bayreuth. Seine Forschungsinteressen liegen im Bereich der Mathematischen Kontroll- und Systemtheorie mit Schwerpunkt auf numerischen und optimierungsbasierten Methoden der Stabilitätsanalyse und Stabilisierung nichtlinearer Systeme.

Prof. Dr. **Edmund Handschin** gründete den Lehrstuhl für Energiesysteme und Energiewirtschaft an der Technischen Universität Dortmund und die Firma EUS GmbH in Dortmund, die für ihre Kunden Lösungen an der Schnittstelle zwischen Energie- und Informationstechnologie realisiert. Er leitete die EUS GmbH als geschäftsführender Gesellschafter. Er ist Mitglied der EU Technologie-Plattform „Smart Grids“ und Fellow des IEEE. Seine Forschung befasst sich mit dezentralen Energieversorgungssystemen, Informationstechnologie in der Energietechnik, statistischen Methoden der Energietechnik und Optimierungsverfahren für das Energiemanagement.

acatech-Mitglied Prof. Dr.-Ing. habil. **Dietrich Hartmann** ist Professor für Informatik an der RU Bochum. Er gehört dem Advisory Board von FOMAAS (Siegen) sowie dem Expertenrat „Zukunft Bau“ des BMVBS an; er war Alexander von Humboldt Research Fellow an der UC Berkeley, USA. Momentane Hauptforschungsschwerpunkte sind vor allem die Simulation und Optimierung von Ingenieursystemen und -prozessen sowie das sensorbasierte Monitoring und die Systemidentifikation von Tragwerken mit Methoden der Computational Intelligence, jeweils unter Berücksichtigung von Unsicherheiten.

Prof. Dr. **Dietmar Hömberg** ist Mathematikprofessor an der TU Berlin und Leiter der Forschungsgruppe „Nichtlineare Optimierung und Inverse Probleme“ am Weierstraß-Institut für Angewandte Analysis und Stochastik Berlin. Seine Forschungsschwerpunkte sind die Modellierung, Simulation und Optimalsteuerung technischer Prozesse. Anwendungen reichen von der Modellierung von Gefügemwandlungen in der Stahlproduktion über die Untersuchung konkreter Produktionsprozesse wie Schweißen, Fräsen, Härten bis zur Optimierung ganzer Prozessketten.

Ulrich Jaeger ist Geschäftsführer der Busgesellschaften der DB Stadtverkehr in Nordrhein-Westfalen sowie Leiter des Competence Centers Produktion. Er verantwortet in dieser Funktion die Verfahren zur Erstellung der Fahr-, Dienst- und Umlaufpläne bei den Busgesellschaften sowie einigen S-Bahnen der Deutschen Bahn AG. Hierbei spielt die integrierte Optimierung eine wesentliche Rolle für den wirtschaftlichen Betrieb der Verkehre bzw. bei der Erstellung von Angeboten.

Dipl.-Ing. **Moritz Kiese** ist wissenschaftlicher Mitarbeiter am Lehrstuhl für Kommunikationsnetze der TU München. Seine Forschungsgebiete sind Entwurfs- und Planungsverfahren für Backbone-Netze und Mobile Adhoc-Netze. Der Schwerpunkt liegt auf Verfahren der Linearen Optimierung. Von 2005 bis 2008 arbeitete er im BMBF-Projekt Efficient Integrated Backbone (EIBONE).

Prof. Dr.-Ing. habil. **Rudibert King** ist Professor für Mess- und Regelungstechnik an der TU Berlin, Sprecher des Sonderforschungsbereiches „Beeinflussung komplexer turbulen-

ter Scherströmungen“ und des interdisziplinären Forschungsschwerpunktes „Biokompatible Keramikschäume“. Seine Forschungsschwerpunkte liegen in der Modellbildung, Regelung und optimierungsbasierten Führung von nichtlinearen und unsicheren Systemen; die Anwendungsgebiete reichen von der Strömungsmechanik bis zur Bioverfahrenstechnik. Er wurde mit dem Dechema-Preis der Max-Buchner-Forschungstiftung ausgezeichnet.

Dr. **Jürgen Koehl** studierte Mathematik in Bonn und Paris. Er war wissenschaftlicher Mitarbeiter am Max-Planck-Institut für Mathematik und am Forschungsinstitut für Diskrete Mathematik in Bonn. Er ist Mitarbeiter der IBM Deutschland, verantwortlich für die Verbesserung der Entwicklungseffizienz des weltweiten IBM ASIC Design Prozesses und die weltweite Unterstützung der in Kooperation mit dem Forschungsinstitut für Diskrete Mathematik entwickelten BonnTools®. Er ist IBM Distinguished Engineer und Mitglied der IBM Academy of Technology.

acatech-Mitglied Prof. Dr. rer. nat. Dr. sc. techn. h. c. **Bernhard Korte** ist Direktor des Forschungsinstituts für Diskrete Mathematik der Universität Bonn. Er ist Mitglied mehrerer Wissenschaftsakademien und hat zahlreiche nationale und internationale Preise und Ehrungen erhalten. Sein Hauptarbeitsgebiet ist Diskrete Optimierung im Chip-Design.

Prof. Dr. **Günter Leugering** ist Mathematikprofessor an der Universität Erlangen-Nürnberg. Seine Forschungsgebiete umfassen die Optimierung mit Partiellen Differentialgleichungen und bei Multiskalenproblemen. Er ist im Leitungsgremium des Exzellenclusters „Engineering of Advanced Materials“ und Direktor des „Zentrums für Multiskalenmodellierung und Simulation“, er ist Sprecher eines DFG-Schwerpunktprogramms „Optimierung mit Partiellen Differentialgleichungen“ und eines Doktorandenkollegs im Elitenetzwerk Bayerns.

Dr. **Dirk Lorenz** ist wissenschaftlicher Mitarbeiter am Zentrum für Technomathematik an der Universität Bremen. Seine Forschungsschwerpunkte sind Inverse Probleme und mathematische Methoden der Bildverarbeitung und ihre Anwendungen in Ingenieurs- und Lebenswissenschaften.

Prof. Dr. **Alfred Louis** ist Professor für Angewandte Mathematik an der Universität des Saarlandes in Saarbrücken. Seine Forschungsinteressen im Bereich der Angewandten Mathematik reichen von der mathematischen Modellierung komplexer Messprozesse über die Analyse der auftretenden Integralgleichungen bis zur Entwicklung effizienter Algorithmen und deren Implementierung. Anwendungen liegen insbesondere im Bereich der Medizintechnik und des zerstörungsfreien Prüfens. Für seine intensiven Kooperationen mit französischen Partnern ist er von der Republik Frankreich mit dem Orden ‚Chevalier dans l’Ordre des Palmes Académiques‘ ausgezeichnet worden.

acatech-Mitglied Prof. Dr.-Ing. **Klaus Lucas** ist Inhaber des Lehrstuhls für Technische Thermodynamik an der RWTH Aachen. Er ist zur Zeit Vizepräsident der Berlin-Brandenburgischen Akademie der Wissenschaften. Seine aktuellen Arbeitsgebiete sind die Energiesystemanalyse, die molekulare Modellierung von Flüssigkeiten und Gasen und die Phasen- und Reaktionsgleichgewichte fluider Systeme. Er ist Berater mehrerer Unternehmen.

Prof. Dr. **Peter Maaß** ist Professor für Angewandte Mathematik und Leiter des Zentrums für Technomathematik an der Universität Bremen. Nach Professuren an der Tufts University in Boston und der Universität Potsdam ist Peter Maaß seit 1999 an der Universität Bremen tätig. Seine Forschungs- und Arbeitsgebiete sind Inverse Probleme, Wavelet-Analysis und mathematische Methoden der Signal- und Bildverarbeitung. Er hat mehrere Patente im Bereich der Bildverarbeitung und wurde 1997 mit einem bundesweiten Innovationspreis ausgezeichnet.

acatech-Mitglied Prof. Dr.-Ing. **Wolfgang Marquardt** ist Professor für Prozesstechnik an der RWTH Aachen, Sprecher der AVT – Aachener Verfahrenstechnik, Mitglied des Vorstands des Exzellenz-Clusters „Tailor-made Fuels from Biomass (TMFB)“ und der Graduiertenschule „Aachen Institute for Advanced Studies in Computational Engineering Science (AICES)“. Seine systemwissenschaftlich orientierten Forschungsinteressen liegen im Bereich der modellgestützten Methoden für Entwurf und Betrieb verfahrenstechnischer Prozesse. Er wurde mit dem Leibniz-Preis der DFG ausgezeichnet.

Prof. Dr. **Alexander Martin** ist Mathematikprofessor an der TU Darmstadt. Seine Forschungsgebiete umfassen die algorithmische diskrete Optimierung und das Operations Research. Praxiserfahrungen liegen in den Bereichen Logistik und Verkehr, Produktionsplanung sowie Energie- und Finanzmanagement vor. Er ist derzeit Vizepräsident der TU Darmstadt und wurde für die gute Kooperation zwischen Wissenschaft und Unternehmen vom Land Hessen ausgezeichnet.

Prof. Dr. **Volker Mehrmann** ist Mathematikprofessor an der TU Berlin und Sprecher des DFG Forschungszentrums MATHEON. Seine mathematischen Forschungsschwerpunkte sind Numerische Mathematik, Kontrolltheorie, Matrizen Theorie mit Anwendungen u. a. in der Strömungskontrolle, sowie der Simulation und Regelung von Mehrkörpersystemen und elektrischen Schaltungen.

Prof. Dr. **Rolf Möhring** ist Mathematikprofessor an der TU Berlin und Mitglied im DFG-Forschungszentrum MATHEON. Seine Forschungsschwerpunkte sind Kombinatorische Optimierung und industrielle Anwendungen in Produktion und Logistik. Er wurde mit dem Wissenschaftspreis der Gesellschaft für Operations Research ausgezeichnet und ist zurzeit Vizepräsident der Mathematical Programming Society.

Prof. Dr.-Ing. **Dirk Müller** leitet den Lehrstuhl für Gebäude- und Raumklimatechnik des E.ON Energy Research Centers der RWTH Aachen. Er war zuvor in der zentralen Forschung der Robert Bosch GmbH tätig. Seine Forschungsschwerpunkte liegen bei den Themen Raumluftrömungen, thermische Behaglichkeit, Luftqualität in Innenräumen und innovative Versorgungssysteme für Gebäude.

Van Vinh Nguyen, M.Sc., ist wissenschaftlicher Mitarbeiter am Lehrstuhl für Informatik an der Ruhr-Universität Bochum. Er hat dort den Masterstudiengang Computational Engineering absolviert. Aktuelle Forschungsschwerpunkte sind die agentenbasierte Strukturgestaltung und -optimierung, Computational Steering und Artificial Intelligence Technologien.

Prof. Dr.-Ing. **Wolfgang Nitsche** ist Professor für Aerodynamik am Institut für Luft- und Raumfahrt der TU Berlin. Einer seiner Forschungsschwerpunkte ist die aktive Kontrolle aerodynamischer Strömungen.

Dr. **Hartwig Preckel**, Perkin Elmer Cellular Technologies Germany GmbH, betreut kundenspezifische und Produktentwicklungsprojekte zur Etablierung neuer Instrumente und Methoden in der pharmakologischen Wirkstoffforschung. Schwerpunkt ist Software zur Automatisierung von bildbasierten zellbiologischen Analysen.

Prof. Dr.-Ing. habil. **Christian Rehtanz** ist Inhaber des Lehrstuhls für Energiesysteme und Energiewirtschaft der TU Dortmund. Er war in leitenden Positionen bei ABB in der Schweiz und in China tätig. Seine Forschungsschwerpunkte sind die Erhöhung der Netzeffizienz durch Integration neuer Technologien sowie neue schutz- und leittechnische Konzepte. Er ist Inhaber des „World Top 100 Young Innovators Award 2003“ des Massachusetts Institute of Technology (MIT).

Dr. **Sebastian Sager** ist Leiter einer Forschungs-Nachwuchsgruppe im Bereich des Wissenschaftlichen Rechnens an der Universität Heidelberg. Seine mathematischen Forschungsinteressen umfassen die Optimierung und die Optimale Steuerung von komplexen Prozessen; im Bereich der Anwendungen liegt der Schwerpunkt auf Transport und Verfahrenstechnik. Er erhielt den Dissertationspreis der Deutschen Gesellschaft für Operations Research und den Klaus-Tschira-Preis für verständliche Wissenschaft.

Prof. Dr.-Ing. habil. Dr.-Ing. E. h. **Michael Schenk** leitet das Fraunhofer-Institut für Fabrikbetrieb und -automatisierung IFF in Magdeburg und den „Lehrstuhl für Logistische Systeme“ sowie das Institut für Logistik und Materialflusstechnik (ILM) an der Otto-von-Guericke-Universität Magdeburg. Er ist Ehrendoktor der Staatlichen Technischen Universität Moskau – MADI, Vorsitzender des VDI-Landesverbandes Sachsen-Anhalt und Mitglied des Wissenschaftlichen Beirats der Bundesvereinigung Logistik e. V. (BVL).

Prof. Dr. **Rüdiger Schultz** ist Mathematikprofessor an der Universität Duisburg-Essen. Seine mathematischen Forschungsschwerpunkte sind diskrete und stochastische Optimierung, seine Anwendungsfelder u. a. Optimierung in Energiesystemen, beim Gastransport und bei der Ressourceneinsatzplanung.

Dr. **Michael Stingl** ist wissenschaftlicher Mitarbeiter am Lehrstuhl für Angewandte Mathematik der Universität Erlangen-Nürnberg. Er leitet die Arbeitsgruppe Strukturoptimierung und ist Mitglied der DFG-Forschergruppe 894. Seine aktuellen Forschungsthemen sind nichtlineare semidefinite Optimierung sowie Strukturoptimierung, insbesondere Materialoptimierung mit Anwendungen im Bereich Design von Flug- und Fahrzeugkomponenten sowie der Medizin.

Dr. **Michael Thess** studierte Mathematik an der TU Chemnitz und der Staatlichen Uni Sankt Petersburg. Seit 1998 leitet er als Vorstand die Forschung/Entwicklung des Data-Mining-Spezialisten prudsys AG.

Dipl.-Math. techn. **Dennis Trede** ist wissenschaftlicher Mitarbeiter und Promotionsstudent am Zentrum für Technomathematik der Universität Bremen. Er studierte Technomathematik an der TU Dresden und der Universität Bremen. Seine Forschungsinteressen liegen im Bereich der Bild- und Signalverarbeitung und der Lösung inverser Probleme mit Anwendungen aus den Ingenieur- und Lebenswissenschaften.

Prof. Dr. h.c. Dr.-Ing. **Eckart Uhlmann** ist Leiter des Fraunhofer-Instituts für Produktionsanlagen und Konstruktionstechnik (IPK) sowie des Fachgebiets Werkzeugmaschinen und Fertigungstechnik am Institut für Werkzeugmaschinen und Fabrikbetrieb (IWF) der TU Berlin im Produktionstechnischen Zentrum Berlin. Er war in verantwortlichen Positionen im Bereich der Forschung, Entwicklung und Anwendungstechnik in der Firmengruppe Hermes Schleifmittel GmbH & Co. Hamburg tätig.

Prof. Dr. **Jens Vygen** ist Professor für Diskrete Mathematik an der Universität Bonn. Er leitet die Arbeitsstelle „Diskrete Mathematik und Anwendungen“ der Nordrhein-Westfälischen Akademie der Wissenschaften und das Forschungsgebiet „Optimization in Large and Complex Networks“ im Hausdorff Center for Mathematics (Exzellenzcluster). Seine Forschungsgebiete sind Kombinatorische Optimierung und Chip-Design.

Dr. **Roland Wessälly** ist Geschäftsführer der atesio GmbH. Im Jahr 2000 gründete er atesio gemeinsam mit Kollegen als Spin-Off des Zuse-Instituts Berlin. atesio unterstützt Netzbetreiber bei der strategischen Vorbereitung von Investitionsentscheidungen sowie der Kosten- und Qualitätsoptimierung ihrer Netze. Seine Forschung zur Optimierung von Telekommunikationsnetzen wurde mit dem Vodafone Innovationspreis ausgezeichnet.

acatech-Mitglied Prof. Dr. Ing. habil. **Peter Wriggers** ist Professor für Mechanik an der Fakultät für Maschinenbau der Leibniz Universität Hannover. Diskretisierungs- und Simulationstechniken mit Bezug zur Materialtheorie und Kontaktmechanik stellen Schwerpunkte seiner Forschungen dar. Anwendungsgebiete sind Material- und Oberflächenentwicklung und zugehörige Simulationen im Bereich des Bauwesens, Maschinenbaus und der Biotechnologie. Er ist Mitglied in weiteren wissenschaftlichen Gesellschaften und Akademien und erhielt mehrere nationale und internationale Preise.

Dr.-Ing. **Stefan Zachow** leitet die Arbeitsgruppe Medical Planning am Zuse-Institut Berlin (ZIB) und ist kooptierter Leiter der AG Computational Medicine. Seine Arbeitsschwerpunkte sind Computergestützte Chirurgie, Visualisierung, medizinische Bild- und Geometrieverarbeitung sowie modellgestützte Therapieplanung in der Medizin. Er ist Diplom-Ingenieur, Diplom-Informatiker und absolvierte ein postgraduales Studium der Medizinphysik.

> acatech – DEUTSCHE AKADEMIE DER TECHNIKWISSENSCHAFTEN

acatech vertritt die Interessen der deutschen Technikwissenschaften im In- und Ausland in selbstbestimmter, unabhängiger und gemeinwohlorientierter Weise. Als Arbeitsakademie berät acatech Politik und Gesellschaft in technikwissenschaftlichen und technologiepolitischen Zukunftsfragen. Darüber hinaus hat es sich acatech zum Ziel gesetzt, den Wissenstransfer zwischen Wissenschaft und Wirtschaft zu erleichtern und den technikwissenschaftlichen Nachwuchs zu fördern. Zu den Mitgliedern der Akademie zählen herausragende Wissenschaftler aus Hochschulen, Forschungseinrichtungen und Unternehmen. acatech finanziert sich durch eine institutionelle Förderung von Bund und Ländern sowie durch Spenden und projektbezogene Drittmittel. Um die Akzeptanz des technischen Fortschritts in Deutschland zu fördern und das Potenzial zukunftsweisender Technologien für Wirtschaft und Gesellschaft deutlich zu machen, veranstaltet acatech Symposien, Foren, Podiumsdiskussionen und Workshops. Mit Studien, Empfehlungen und Stellungnahmen wendet sich acatech an die Öffentlichkeit. acatech besteht aus drei Organen: Die Mitglieder der Akademie sind in der Mitgliederversammlung organisiert; der Vorstand, der von den Mitgliedern der Akademie bestimmt wird, lenkt die Arbeit; ein Senat mit namhaften Persönlichkeiten aus Industrie, Wissenschaft und Politik berät acatech in Fragen der strategischen Ausrichtung und sorgt für den Austausch mit der Wirtschaft und anderen Wissenschaftsorganisationen in Deutschland. Die Geschäftsstelle von acatech befindet sich in München; zudem ist acatech mit einem Hauptstadtbüro in Berlin vertreten.

Weitere Informationen unter www.acatech.de

> acatech diskutiert

Die Reihe „acatech diskutiert“ dient der Dokumentation von Symposien, Workshops und weiteren Veranstaltungen der Deutschen Akademie der Technikwissenschaften. Darüber hinaus werden in der Reihe auch Ergebnisse aus Projektarbeiten bei acatech veröffentlicht. Die Bände dieser Reihe liegen generell in der inhaltlichen Verantwortung der jeweiligen Herausgeber und Autoren.